PhD Thesis

Paul Alexander Bernal

London School of Economics and Political Science

Department of Law

"Do deficiencies in data privacy threaten our autonomy and if so, can informational privacy rights meet this threat?"

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work.

Paul Alexander Bernal

Word count as submitted (including footnotes, excluding preface, table of contents and bibliography): 99,995

Abstract

This thesis sets out a model to examine how the internet functions. 'The symbiotic web' suggests a symbiotic relationship between corporations that have built business models dependent upon the gathering of personal data from people, and the individuals themselves who have begun to rely on apparently 'free' services (from search to email, social networking to YouTube). Having set out the model, the thesis looks at its implications: how it has contributed to many, both the positive and negative, developments on the internet in recent years, but also driven the mass gathering, use and holding of personal data. The symbiotic web is currently essentially beneficial to both businesses and individuals, but there are significant risks attached - risks associated with the accumulation of data and risks that the symbiotic relationship could become negative and parasitic, putting individuals' privacy and autonomy at risk.

The implications of this model are examined through the use of case studies: the dispute between Google and the Article 29 Working Party over data retention, Phorm's 'Webwise' behavioural targeting system, and a number of smaller case studies about data vulnerability from the HMRC data disc loss to the ACS:Law hack/leak.

The thesis suggests the development and use of specific rights designed for the internet to address the associated risks: a 'right to roam the internet with privacy', a right to monitor those who monitor us, and a 'right to delete data'. These rights would be set out as principles rather than enacted and enforced as laws, and brought into play through Murray's model of symbiotic regulation. These rights would support the positive development of the web symbiosis and encourage and shape new business models that are more supportive of individual autonomy and privacy.

Preface and Acknowledgements

For most people in what might loosely be described as the developed world, the internet can no longer be considered an optional extra, but is an intrinsic part of life. Significant aspects of life take place on the internet – interactions with government, access to news and information, banking, shopping, communication and social interaction on an ever-growing scale. Accompanying this growth in significance of the internet has been a massive growth in the gathering, using and holding of personal data, gathered both in the 'real' world and through the internet itself. The ways that this data is gathered, used and held, and the intrusions to our privacy connected to that gathering and use, have an impact that needs to be considered very carefully. It is the contention of this thesis that both the data and the processes through which it is gathered, used and held represent a threat to our autonomy.

Central to the thesis is a theoretical model of how the internet in its current, substantially commercial form functions: 'the Symbiotic Web'. This model suggests a symbiotic relationship between corporations that have built business models dependent upon the gathering of personal data from people, and the individuals themselves who have begun to rely on apparently 'free' services (from search to email, social networking to YouTube). This symbiotic relationship is currently mutually beneficial – providing individuals with powerful and enjoyable services and businesses with significant economic opportunities – but there are significant risks attached, most directly to privacy and autonomy. The thesis looks at the implications of this model – it helps to explain the growth in data and data gathering methods and the potential risks that are attached to them. It also helps to suggest the primary way forward posited in this thesis: the use of a rights-based approach, with specific rights arising in relation to our functioning on the internet.

The specific rights suggested are: a right to roam the internet with privacy, a right to monitor those who monitor us, and a right to delete personal data. They are suggested not as imposed, purely legal rights, but as more natural rights that arise from the community, from what people feel and believe are right – as well as being pragmatic rights that can help to guide and shape the business models that in turn guide and shape the way that the internet itself functions. The thesis argues that a clearer exposition of these rights could help people to both be and feel less threatened, as well as helping businesses to develop more positive and effective business models into the future.

The thesis is organised as follows. Chapter One is an extended introduction, looking at the conceptual background of autonomy and privacy, as well as the primary challenges and criticisms to an approach based on autonomy and privacy, particularly in relation to the internet. Chapter Two sets out the theoretical model of the 'Symbiotic Web'.

Chapters Three to Five detail the case studies: the rights put forward in the thesis arise from the examination of the case studies. Chapter Three looks at search engines, and in particular the interaction between Google and the Article 29 Working Party – and the idea of a right to roam the internet with privacy emerges from it. Chapter Four looks at behavioural advertising, and in particular the contentious case of Phorm – and through it looks in some depth at the crucial issue of consent, positing a new way to look at consent in an interactive internet: 'Collaborative Consent'. From this case study, the idea of a right to monitor those monitoring us emerges. Chapter Five looks at data vulnerability in its many manifestations, through a number of mini-case-studies, and from this emerges the idea of a right to delete personal data, because ultimately the only way to prevent data from being vulnerable is for that data not to exist.

Chapter Six draws together the rights posited in Chapters Three to Five, and looks in more depth and the nature and benefits of a rights-based approach. It suggests a new concept – 'Autonomy by Design' – through which these rights might begin to be satisfied. It also looks in more detail at the way in which Murray's approach of 'symbiotic regulation' applies here – and at the role of the community in the process.

Chapter Seven draws together the thesis, and looks at how the internet might function if these rights were realised. Through this examination of a potential future internet it looks at how the critiques and challenges set out at the start of the thesis may be met – and at the possibilities for more privacy-friendly future.

Elements of this thesis have inspired the following publications: *'Web 2.5: The Symbiotic Web'*: International Review of Law, Computers & Technology (Volume 24, Issue 1 March 2010), *'Collaborative Consent: harnessing the strengths of the Internet for consent in the online environment'*: International Review of Law, Computers & Technology (Volume 24, Issue 3, Nov. 2010), *'Rise and Phall: Lessons from the Phorm Saga'*: Chapter 13 in *'Computers, Privacy and Data Protection: an element of choice'*, published by Springer, February 2011, and *'A Right to Delete?'*, European Journal of Law and Technology, Vol. 2, No.2, 2011

I would like to thank my doctoral supervisors, Professor Conor Gearty and Professor Andrew Murray, for their invaluable help, advice, encouragement and support throughout my research, and my wife Corina and daughter Alice for their support in every aspect of my life and work.

My work was funded by a Doctoral Award from the Arts and Humanities Research Council, for which I am deeply grateful.

Chapter 1: Privacy, Autonomy and the Internet

"Do deficiencies in data privacy threaten our autonomy and if so, can informational privacy rights meet this threat?"

1 Introduction and research question

Data privacy has become increasingly topical over the last few years. Among the events that have brought the issue to the attention of the general public have been the disappearance in 2007 of two CDs containing personal data of 25 million people in the UK, lost in transit between Her Majesty's Revenue and Customs and the National Audit Office and the revelation that while sending its cars around most of the streets in the UK to compile its online photographic database for its Street View service Google had gathered data from a vast number of private Wi-Fi networks with neither the knowledge nor the permission of the owners of those networks.[1]

As society becomes increasingly dependent on the use of digital information and as that use becomes increasingly integrated into almost all aspects of our lives, the implications of this development are considerable. One of the most direct of these is that people can feel a loss of control. The Information Commissioner's Office reported in October 2010 that its research indicated that 'individuals increasingly feel they have lost control of their personal information'.[2] This raises many questions. Are these feelings of lost control representative of a real loss of control? If control over personal information is lost, what are the implications for real lives? Is it ever possible to regain control, or do people simply have to learn to live with that loss of control? The research question for this thesis encompasses all of these questions. It

[1] See ICO press release on their investigation into Street View: http://www.ico.gov.uk/~/media/documents/pressreleases/2010/google_inc_street_view_press_release_03112010.ashx

[2] As reported for example in ICO 2010. Response to the Ministry of Justice's Call for Evidence on the current data protection legislative framework., p2

asks about the impact upon individuals' autonomy of current and future issues surrounding data privacy, and whether a rights-based approach could be an effective way to address this impact.

The primary conclusion of this thesis is that the impact upon autonomy is real – that is, that deficiencies in data protection *do* exist, and that they *do* threaten people's autonomy, both online and in the 'real' world, in a number of different ways. First of all, autonomy is threatened directly by those who wish to make use of personal data. Some of these uses of data are for obviously malign purposes like identity theft or other criminal purposes: others for more neutral purposes such as those connected to marketing and advertising. In related ways, and using similar methods, there is potential for manipulation or interference with autonomy for political and security purposes.

Secondly, the threat arises where the use is for more subtle purposes like highly targeted advertising or other forms of persuasion – purposes that have variable and potentially dubious levels of legality and morality. Thirdly, in potentially even more far-reaching and insidious ways, there is the use through automatic profiling and 'tailoring' of information, links and even access rights based on data and behavioural targeting – something that will be examined in some depth in Chapter Two. The threats to individuals' autonomy are, it will be suggested, both real and potentially dangerous, and need to be investigated, understood and challenged.

The second conclusion is that a rights-based approach, using appropriately set out informational privacy rights, could begin to address this threat and begin to ameliorate this problem. This thesis sets out some of the rights that might do this– and attempts to demonstrate, through an examination of case studies from recent years, how and where they might be able to be successful.

The third, more tentative but ultimately more important conclusion is that a more privacy- and autonomy-friendly internet is possible. How that kind of a future might work, and what it might look and feel like, is a central part of the last chapter of this thesis. This conclusion is tentative because the situation is highly fluid and, as shall be shown in the case studies developed there, the forces arrayed against privacy and autonomy are extensive and powerful. It is important because the internet has become such a fundamental part of modern life and society.

1.1 The internet in modern life

The primary focus of this thesis is the internet. In practice, for most people in what might loosely be described as the developed world the internet can no longer be considered an optional extra, but is an intrinsic part of life in a modern, developed society. Significant aspects of life take place on the internet – interactions with government, for example, are becoming increasingly electronic, not only in terms of access to information but more directly and interactively – the completion of tax returns, access to health services,[3] interaction with local government, and much more.[4] The digital economy has already become a significant part of the economy as a whole, and this is increasing all the time. In the UK, it is predicted that by 2012 £1 in every £5 of all new commerce in this country will be online.[5] It is increasingly the case that people who are not able to access products and services online are at a significant disadvantage – for example being unable to take advantages of discounts for insurance,[6] better interest rates on savings,[7] having tighter deadlines for submission of information and so

[3] See http://www.nhsdirect.nhs.uk/ NHS direct is suggested as the first port of call for health problems in the UK.
[4] For a fuller list of governmental services available http://www.direct.gov.uk/en/index.htm
[5] See 'Digital Britain', the interim report from January 2009, downloadable from http://www.culture.gov.uk/images/publications/digital_britain_interimreportjan09.pdf
[6] Aviva insurance, for example, in October 2010, was offering a 20% discount for online applications for car insurance. See http://www.aviva.co.uk/car-insurance/
[7] Most UK banks offer 'e-savings' accounts or equivalents, only accessible online, offering better interest rates or other advantages. Examples include: http://www.natwest.com/personal/savings/g1/instant-access/e-savings.ashx ,

forth,[8] Moreover, there are some very useful services that are only available online – price comparison sites for insurance and other financial services are two such examples.[9] Shopping has been revolutionised, from specialised online services like Amazon and auction sites like eBay to the online versions of existing supermarkets, allowing ordering online and delivery to your home.[10]

All this is without considering the most direct, 'traditional' uses of the internet, as an unparalleled source of information – for educational or recreational purposes, as an increasingly important news source,[11] or simply to discover practical information such as the location and opening hours of shops, events and so forth.

Perhaps even more important is not the extent to which a capacity to use the internet is now required, but the reality of how much it is used in practice. The numerous sites and services noted above are only a small part of what has become a significant element of life. There are many others that have become part of the social fabric for a large section of society. Social networking sites are just one example – they cannot be said to be either practically necessary or economically advantageous, but they are used, extensively and increasingly, and not just by young people.[12] The same can be

---

http://www.barclays.co.uk/Savings/Instantaccess/esavingsReward/P1242557963964, http://www.firstdirect.com/savings/everyday-esaver-overview.shtml

[8] UK tax returns submitted by paper, for example, are required to be submitted by 31st October each year, while online submissions are allowed until 31st January the following year. See http://www.hmrc.gov.uk/sa/dead-pen.htm

[9] E.g. http://www.gocompare.com/, http://www.confused.com/, http://www.comparethemarket.com/

[10] Amazon is www.amazon.com or www.amazon.co.uk, ebay www.ebay.com, and see for example http://www.sainsburys.co.uk/home or http://www.tesco.com/ for online stores of supermarkets

[11] In the 2008 US election, for example, the Internet was one of the most important sources of news for voters, particularly for young people. Pew Internet Research reported that '42% of those ages 18 to 29 say they regularly learn about the campaign from the internet, the highest percentage for any news source'. See http://people-press.org/report/384/internets-broader-role-in-campaign-2008

[12] Facebook alone has more than 750 million active users worldwide. See http://www.facebook.com/press/info.php?statistics

said of a whole range of other services, from message boards and blogs to media services such as YouTube.[13]

Further to this, the internet is no longer something that can (or is) only be accessed through computers. More and more devices can and do use or provide a connection to the internet, from smartphones such as the iPhone or the RIM's Blackberry, tablet devices like the iPad and its rivals to Blu-ray players, TV receivers and game machines like the X-Box 360, the Sony Playstation and the various Nintendo consoles. This trend appears likely to increase, and increase rapidly, as the advantages of using internet connections become more apparent.

The ultimate implication of this is that living without using the internet places people at significant disadvantages in many different ways, including social, cultural, democratic and financial ways. The concept of a 'digital divide', or more accurately 'digital divides'[14] between those who have the skills and opportunities to take advantage of digital services and those who don't has been discussed since the 1990s – see for example the work of Norris[15] and Mossberger.[16] The nature of the relevant divides has changed considerably over the last decade as the role that the internet plays in society has become more significant, as outlined above, and access to it has become the norm rather than the exception. The disadvantages to those who do not have internet access are continuing to grow both in scale and breadth – which is one of the reasons that there are increasing calls to consider access to the internet a 'right'.

---

[13] www.youtube.com

[14] Divides between rich and poor nations, between the rich and the poor within nations, between the better and worse educated, between the urban and the rural, divides based on gender, disability, race and more – there are many possible reasons for what might be termed digital disadvantage. Mossberger also identifies different aspects of the divides – what she terms the 'access divide', the 'skills divide', the 'economic opportunity divide' and the 'democratic divide', paralleling some of the discussion in this chapter. See MOSSBERGER, K., TOLBERT, C. J. & STANSBURY, M. 2003. *Virtual inequality : beyond the digital divide,* Washington, D.C., Georgetown University Press. particularly p9

[15] See for example NORRIS, P. 2001. *Digital divide : civic engagement, information poverty, and the Internet worldwide,* Cambridge, Cambridge University Press.

[16] In for example MOSSBERGER, K., TOLBERT, C. J. & STANSBURY, M. 2003. *Virtual inequality : beyond the digital divide,* Washington, D.C., Georgetown University Press.

The idea of internet access as a basic human right has been put forward by many, and according to a large survey by the BBC World Service, nearly 80% of people around the world believe that it should be.[17] In Estonia,[18] France,[19] and Greece,[20] internet access has already been made a constitutional right, while in Finland this right has become legally enforceable.[21] The EU Telecoms Reform Package agreed in 2009 supports high-speed access for 'all citizens' throughout the EU.[22]

In the UK, surveys suggest the same. In 2009, a survey for the Communications Consumer Panel showed that '84 per cent of people agreed that it should be possible for everyone in the UK to have broadband at home, regardless of where they live. Many people already see broadband as essential and even more believe that soon it will be essential for everyone'. As Communications Consumer Panel Chair Anna Bradley put it:

> "The tipping point will be when broadband does not just provide an advantage to people who have it, but disadvantages people who do not. Interestingly some people already feel disadvantaged: those who live in not-spots and those who have school-age children but do not have broadband at home."[23]

Whilst the idea of internet access as a fundamental human right is one that is highly debatable, the nature and scale of the discussion, and the reality of the use of the internet in practice does leads to a significant conclusion: that if

[17] http://news.bbc.co.uk/1/shared/bsp/hi/pdfs/08_03_10_BBC_internet_poll.pdf The survey included more than 27,000 people in 26 countries,
[18] See http://news.bbc.co.uk/1/hi/world/europe/3603943.stm
[19] See e.g. http://www.dailymail.co.uk/news/worldnews/article-1192359/Internet-access-fundamental-human-right-rules-French-court.html?ITO=1490
[20] Article 5A, paragraph 2 of the Constitution of Greece states that "All persons are entitled to participate in the Information Society. Facilitation of access to electronically handled information, as well as of the production, exchange and diffusion thereof constitutes an obligation of the State". See for example http://www.unhcr.org/refworld/docid/4c52794f2.html
[21] See for example http://www.bbc.co.uk/news/10461048 Finland not only made internet access a legal right, but specified a minimum speed of access of 1Mbps.
[22] See http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/09/491
[23] http://www.communicationsconsumerpanel.org.uk/smartweb/news-releases/soon-it-will-be-essential-for-everyone-to-have-broadband

autonomy is considered in an appropriately broad sense, as shall be discussed in section 2 below, it should now be understood as including the opportunity to act freely on the internet.

1.2 Data and the internet

The internet also matters because it is the focus of efforts to gather, analyse, use and store personal data – and the internet offers hitherto unheard of opportunities to perform this gathering, analysis, use and storage of data.[24] The case studies in Chapters Three to Five reveal just some of the ways in which this is already happening, and give at least some idea of how this could develop into the future.

The nature of the internet makes manipulation of the lives of individuals through the use of personal data particularly easy. The ways in which this can work are analysed throughout the thesis – and specifically in Chapter Two, where a model describing the current functions of the internet is set out, and its implications explored, including some of the direct and indirect ways that individuals' autonomy can be threatened. This model, the Symbiotic Web, suggests that there is a symbiotic relationship between the individuals who use the internet and are reliant on 'free' sites and services, and the businesses that provide those services and which have built business models dependent on their ability to gather and process personal data from those individuals. The Symbiotic Web is a central theme to this thesis, and is discussed briefly in 1.5 below, before the full model is set out in Chapter Two.

It is becoming increasingly difficult to separate 'online' and 'offline' data. In effect, data is commixed as the internet becomes more and more integrated into 'real' life. To take one example, one of the largest types of data gathered in the 'real' world is that gathered by supermarkets for their loyalty schemes,

[24] Each of Cate's four principles for data growth, set out in CATE, F. H. 1997. *Privacy in the Information Age,* Washington, D.C., Brookings Institution Press., pp13-16, applies directly to the internet. His fourth principle in particular refers to the impact of computer networks.

such as the Tesco Clubcard and the Nectar service operated by Sainsbury's, BP and others. Though initially these data are gathered and used in relation to 'real world' shopping, they now include the shopping done online – and they are held in a way that they can be accessed online, and used online. The data themselves have become online data.

Even data held by corporations or government departments that are held on 'private' computers or networks are also becoming part of the 'online' world, as those networks are using 'public' infrastructure or running on 'virtual private networks' on the internet, with the same computers being used to gather, hold and access the data that are also used to access the internet. Separation and isolation of computers from the internet is increasingly uncommon and likely to become more so. Added to that, data gathered offline may be (and is likely to increasingly be) integrated and aggregated with other data, much of which is gathered online, and the results are then stored and used online – this integration and aggregation is discussed in more depth in Chapters Two and Four.

The concepts of an 'internet of things' and 'augmented reality' take the integration between the online and offline worlds further steps forward. The 'internet of things' refers to the way that more and more 'real' objects have an online 'presence' through chips (and particularly RFID chips) built-in to them, allowing them to be mapped, tracked, inventoried and so forth,[25] while 'augmented reality' refers to the use of digital information to supplement 'real' information – for example in heads-up displays in aeroplanes and cars providing assistance for pilots and drivers, or in mapping applications for smartphone. The use of augmented reality in smartphones in particular – taking advantage of the geo-location systems built into such phones – is already relatively widespread.[26] With the increasing prevalence of

[25] The term 'internet of things' may have been coined by Kevin Ashton in 1999, though it is now of common usage. See Ashton's article in RFID news in 2009 'That 'internet of things' thing', accessible at http://www.rfidjournal.com/article/view/4986

[26] In May 2011 there were 180 augmented reality apps available on the iTunes App Store for example.

smartphones, augmented reality might be expected to become more common.

Finally, the internet introduces new levels of vulnerability and new ways in which private data, once it has been gathered, stolen or otherwise acquired, appropriately or inappropriately, may be loosed upon the world. The most graphic examples of this involve leaks like those performed by Wikileaks. This is a trend that is likely to continue – to grow – and one that demands attention.

As a consequence of all these factors, the primary focus of this thesis is upon the internet, though some of the examples and case studies used relate to events and scenarios that are 'offline', such as the HMRC data loss referred to above and detailed in Chapter Five. The implications of this thesis have a wider application, and relate quite directly to offline as well as online policies, laws and practices concerning the gathering, use and holding of data.

1.3 Underlying questions and a paradigm shift

One implication of the focus on the internet is that it raises one of the underlying questions that will be addressed throughout this thesis: how 'public' is the internet? Should the internet, or some significant part of it, be considered a 'public space', and if it should, what does that imply? If the answer to this question is 'yes', as is one of the ultimate contentions of this thesis, then the implications are considerable. Many things would need to be considered, not just in terms of the rights of individuals as they browse the web or use internet-based services, but in terms of the obligations of those providing or hosting websites or offering internet-based services. Those rights and the corresponding obligations are elaborated throughout this thesis, and brought together in Chapters Six and Seven.

How 'public' the internet should be considered is a complex question, and one that cannot easily be answered using what might be termed 'old style' rules. It brings up a lot of issues: what is cyberspace, and what is the internet? Is it simply a collection of connected private spaces, each owned and governed by the people who run the websites concerned? In practice, the vast majority of the internet is owned and run privately – so should the web be considered something effectively private, with browsers having to follow whatever rules the web owner sets, particularly in terms of privacy? Or is it a public space, and governed by public rules, public norms and so forth – with people having an expectation that they should have certain rights, and that those rights will be respected as they browse the internet?

The implication of the suggestion that the internet is now an intrinsic part of modern life is that it should, in certain ways, be considered public, and that people who use it should be able to rely on their rights being respected. This is already true to an extent in terms of commerce – commercial law including contract law applies to commercial transactions that take place over the internet – and issues like defamation, pornography and so forth. Though there are complications, jurisdictional issues and so forth, the principles in all these areas are clear. Despite the declarations of independence of cyberspace from Barlow onwards,[27] law has been applied to online life, with varying degrees of success, in many different ways.

This leads to the conclusion that we must consider the internet to be to a significant extent a public space. This thesis suggests that this is taken several logical steps further – and that rights be applied to the internet as a public space as a consequence. The rights suggested in this thesis should be viewed in this light. If we as people have the *need* to use the internet, and the

[27] Barlow's famous 'Declaration of Independence for Cyberspace', found at https://projects.eff.org/~barlow/Declaration-Final.html , was made in 1996, but there have been similar claims made subsequently over the years, right up to the claims by the hacker group 'Anonymous' in 2010. See http://www.youtube.com/watch?v=gbqC8BnvVHQ

right to use the internet, we should have appropriate protections and rights *when* we use the internet.

This brings up the question of which parts of the internet should be considered public and which private – and hence what kinds of rights (and in particular what degree of privacy) someone using those parts can reasonably expect. This is a question that this thesis investigates in Chapters Six and Seven – but the principle answer, this thesis suggests, is that the default position, the assumption, should be that everywhere on the internet should be considered public unless there is a compelling reason to the contrary.

A second underlying question that arises as a result of the focus of this thesis is that concerning the personal data themselves: to what extent are personal data 'ours'? And, behind that question is the question of what actually counts as 'personal' data. Opinion, law and practice produce a wide variety of potential answers to both of these questions. In places like the U.S. few forms of data are considered personal enough that an individual has any rights over them at all, while the data protection regimes in Europe effectively consider any data that can be directly linked to an individual as 'personal'. The issue of what rights an individual has concerning data held 'about' them is another of the central themes of this thesis – and one of the conclusions drawn is that more rights are needed in order to give individuals more control, and hence more autonomy.

If the answers to these underlying questions are as suggested in this thesis, what is required is a paradigm shift in attitudes to privacy on the internet, and data privacy in general. In a private place, individuals control their own 'privacy settings', while in a public place individuals do not, and hence require legal protection – protection through privacy rights.[28] The default position needs to shift from one where privacy is the exception to one where

[28] Cases such as *Campbell v Mirror Group Newspapers Ltd [2004] UKHL 22*, *Von Hannover v Germany [2004] ECHR 294* and *Mosley v News Group Newspapers [2008] EWHC 1777 (QB)* have centred around what expectations of privacy are appropriate in private or public spaces

privacy is the general rule. Surveillance on the net should not be assumed to be acceptable, and neither should the gathering, processing or holding of personal data. At present, unless an objection is made, it appears that surveillance can and does happen, without the knowledge or consent of the individual, and that data can be and are gathered, processed and held, similarly without the knowledge or consent of the individual. The opposite needs to become the case – those who want to monitor people and those who desire to gather, use or hold data about people should need to justify that monitoring, that data gathering use or holding. If they cannot justify it, or if their justification is inadequate or inappropriate, they should not be able to perform that monitoring or data gathering, and they should not be able to use or hold that data. The informational privacy rights suggested in this thesis are designed to support and enable that paradigm shift.

1.4 Approach and methodology: autonomy as the prime concern

This thesis takes an essentially liberal perspective. What is meant by autonomy in the context of this thesis is examined in depth below. – The approach is drawn primarily from Raz's conception of autonomy, describing an autonomous person as one who '…is a (part) author of his own life.'[29] It is an approach that sees autonomy as a 'constituent element of the good life'.[30] The rights that are suggested in this thesis flow directly from this idea of autonomy – they arise from autonomy and if brought into play they support, protect and help preserve autonomy.

Though the issue of privacy is central to this thesis, it is privacy as a protector of autonomy rather than privacy *per se* that is of prime concern. As already noted, it is particularly true that in the digital world privacy is crucial to protect autonomy. As Nissenbaum puts it:

[29] RAZ, J. 1986. *The Morality of Freedom,* Oxford, Clarendon., p369
[30] Ibid. p408

"Widespread surveillance and the aggregation and analysis of information enhance the range of influence that powerful actors, such as government agencies, marketers, and potential employees, can have in shaping people's choices and actions."[31]

Nissenbaum's analysis categorises the relationship between privacy and autonomy in the digital context in three ways. Firstly, that privacy can itself be considered an aspect of autonomy: autonomy over one's personal information. Secondly, that as privacy frees us from the 'stultifying effects of scrutiny and approbation (or disapprobation), it contributes to an environment that supports the 'development and exercise of autonomy and freedom in thought and action.'[32]This can be looked on as a converse to the panopticon effect: if we don't feel ourselves to be under the constant risk of observation we will feel more able to think and act freely. Thirdly, and most directly for the purposes of this thesis, that without privacy our ability both to make effective choices and crucially to follow them through can be curtailed.[33] The nature of the manipulations possible, as shall be set out throughout this thesis can be both in terms of the choices suggested or offered and the information provided in order to aid in making those choices.

1.5 Privacy per se is not the central concern

The existence and nature of any 'right to privacy' is a subject that is much discussed, and as the digital world becomes more significant it is likely to be discussed even more, but it is not of key concern here. The conclusions and suggestions of this thesis would have a significant effect on privacy in many ways as well as through their providing more autonomy for individuals – but these can be considered as side effects or peripheral benefits rather than main intentions.

[31] NISSENBAUM, H. F. 2010. *Privacy in context : technology, policy, and the integrity of social life,* Stanford, Calif., Stanford Law Books. p83
[32] Ibid. p82
[33] Ibid. p82-83

Privacy and autonomy go hand in hand in protecting and supporting many of what are currently considered 'human rights'. Most directly, such rights are often called 'civil liberties' – freedom of association, freedom of expression, freedom of assembly, freedom of religion and so forth – but they also embrace other important rights including social, cultural and economic rights. The last of these is one that demonstrates some of the most insidious problems on the internet – as shall be set out in Chapter Two. Without appropriate privacy it can be possible for people to be economically disadvantaged and economically discriminated against, with services made unavailable for some or priced differently for others based on private and personal information about those individuals.

The approach taken here is rights-based: what is meant by a rights-based approach, how it differs from current practice, whether such an approach would be effective, and precisely what rights should be used is another central theme to this thesis. The rights themselves are described in more depth, and the reasons behind the approach taken here are developed in Chapter Six. The suggestion of the thesis is first of all that the rights identified in Chapter Six – a right to roam the internet with privacy, a right to monitor those monitoring us, and a right to delete personal data – *are* rights in a philosophical sense, secondly that that they represent the real wishes, desires and understanding of the people concerned – they reflect what people consider to be their rights – and thirdly that they could, from a practical perspective, have a significant and positive impact upon people's autonomy.

1.6 Symbiotic Regulation and the Symbiotic Web

The thesis suggests that the implementation of the suggested rights could be effected through the application of Murray's model of symbiotic regulation.[34] Symbiotic regulation starts by understanding the significant and active role

[34] As set out in chapter 8 of MURRAY, A. D. 2006. *The Regulation of Cyberspace: Control in the Online Environment,* Milton Park, Abingdon, UK ; New York, NY, Routledge-Cavendish.

played by the community in responding to regulatory actions. Further, it suggests that by mapping the relationships between the regulated parties, a desired regulatory result may be brought about through the harnessing of those relationships. The theory of symbiotic regulation is particularly suited to the digital environment – and as shall be shown, applies directly to the case studies used in this thesis. Understanding the role that the community can play – and has already played – in ways that regulation functions in the online world is of crucial importance: the symbiotic regulation model clarifies that role as well as suggesting effective routes for future regulatory approaches.

Building on this, the thesis sets out an underlying theory of how the current, substantially commercial state of the internet functions: the Symbiotic Web. This theory, which is set out in depth in Chapter Two, underpins the analysis not only of the case studies in the thesis, but the conclusions that are drawn from it. Essentially, it suggests that a form of symbiosis is developing on the web. Individuals and commercial enterprises are mutually dependent: enterprises have built business models reliant on a currency of personal data, while individuals depend on free access to many services, from search engines, email systems and social networking sites to media services such as YouTube. These 'free' services use personal data as their way of generating revenues – through targeted advertising, profile building, and the direct sale of personal data.

The symbiosis is essentially benign – it lies behind many positive developments on the internet. It benefits both the users and the businesses that provide services through the internet. Nevertheless, there are significant risks associated with this symbiotic nature that need to be addressed – which is to a great extent what this thesis is about. These risks are examined in depth in Chapter Two. In essence, the problems lie in the way that the symbiosis can drive an even greater tendency to gather personal data than before, and then pressurise those who hold data to find ways to use it, ways that are potentially not merely non-consensual but harmful. Further, and

most significantly from the perspective of autonomy, techniques are developing for the targeting and 'tailoring' both content and links to individuals, often without their knowledge, understanding or consent. This in turn can lead to control over (and potentially automatic selection of) the choices made for individuals, material displayed for individuals and so forth, and potentially the opening up of opportunities for such pernicious practices as price discrimination and even racial, religious or gender-based access to particular websites.

The terms symbiotic regulation and Symbiotic Web are not coincidentally similar: rather, they both reflect the complexity and nature of the internet, the intertwined relationships between individuals, businesses, governments and others, between the hardware, software and services and so forth. It should be noted, however, that the current situation is not something that is separate and unrelated from what has gone on before, but is a modern (and current) illustration of long-standing tensions that have always existed: those between privacy, security and economics on the one hand and between individuals, governments and businesses on the other.

1.7 Themes, case studies and related rights

This thesis primarily concerns the mainstream: the ordinary lives of ordinary individuals as they function in the modern world. The issues of privacy and autonomy concerning those outside the mainstream will be touched upon throughout, and are discussed in Section 2 below, but the mainstream remains the focus.

There is a particular emphasis on business and commercial activities – commercial data gathering and commercial surveillance and so forth. In the current stage of the internet, as described in the model of the Symbiotic Web, most developments both in technology and in its use are driven primarily by business. More data are gathered and in more different ways by businesses – and this trend is accelerating. Governments, as the case studies reveal, have a

tendency to 'piggy back' on commercial data gathering techniques, and to try to find ways to bend business methods to their advantage, with varying degrees of success.

Accordingly, the case studies have been chosen to reflect mainstream activities, and are primarily business related: data gathering by search engines (and in particular Google) in Chapter Three, behavioural advertising, a current favourite on the internet, in Chapter Four, and in Chapter Five data vulnerability, from the use by governments of data stolen from banks in Switzerland and Lichtenstein to the HMRC disc loss, Wikileaks and the Google Street View Wi-Fi data gathering scandal. These case studies show the central principles and underlying issues of this thesis in action. The nature of search engines exemplify the contentiousness of the question of whether the internet is 'public' or 'private', behavioural advertising hits at the heart of the question of whether personal data are 'ours', while the vulnerabilities of that personal data played out in Chapter Five emphasises how significant the ownership and control of that data can be – and hence the importance of asking and answering these questions.

These case studies also loosely follow what might be described as the 'data cycle', starting with the data gathering process in Chapter Three, the data utilisation process in Chapter Four, and the holding of data in Chapter Five. The suggested solutions – the suggested *rights* – emerge from that analysis, and following that pattern. First of all, emerging from the analysis of data gathering, is a proposed *right to roam the internet with privacy*. This is perhaps the most radical of the rights suggested, and the most all-encompassing. The idea of such a right is closely connected with the paradigm shift suggested in 1.3 above, and with the underlying question of the extent to which the internet can be regarded as a public space.

Secondly, and emerging from the examination of behavioural advertising, and the way that data can be used during the browsing process, it is suggested that there be a *right to monitor the monitors*. Underlying that right is the

related and newly introduced concept of 'Collaborative Consent'. Consent is another recurrent theme in this thesis, something that although of crucial importance, particularly for those concerned with autonomy, has in general been given only superficial treatment in the context of the internet. Consent is touched upon throughout the thesis and given particular attention in Chapter Four.

Thirdly, arising through the analysis of data vulnerability and concerning the data holding process, there is a *right to delete personal data*. This touches directly upon the second of the underlying questions discussed in 1.2 above: the extent to which personal data should be considered 'ours'. Again, this is a topic that is touched upon throughout the thesis – it is examined in depth in Chapter Five.

How these rights work together, and combine to form the basis of a comprehensive right to informational privacy is set out in Chapter Six, together with a practical approach towards its implementation, '*Autonomy by Design'*. This then leads to an examination, in Chapter Seven, of the kind of impact these rights could have on an internet in the future – and how these rights might address the potential challenges and criticisms that face them.

2 Autonomy

Autonomy is the underlying theme of this thesis – so it is important to establish both what is meant by autonomy and why it matters. The starting point is to look at what is meant by autonomy, first from a philosophical perspective, and then from a practical perspective in the society that exists today. What is more, as noted above, on the assumption that underpins this thesis one function of rights is to protect autonomy – and where autonomy is threatened, rights may be required to protect it. In this case, as shall be argued in this thesis, what are required are rights of informational privacy.

2.1 A broad definition of autonomy

The starting point for autonomy in the context of this thesis is the work of Joseph Raz. In *The Morality of Freedom* he sets out his conception of autonomy:

"The autonomous person is a (part) author of his own life. The ideal of personal autonomy is the vision of people controlling, to some degree, their own destiny, fashioning it through successive decisions throughout their lives."[35]

For the purposes of this thesis, this can be considered a definition of autonomy. This form of autonomy follows Raz in learning from the more conventional liberal conception of autonomy as the capacity for people to 'decide upon, to revise, and rationally to pursue a conception of the good',[36] but the idea of being an 'author' of your own life can be seen as broader and less rigid than that more commonly proposed. It also makes the concept of autonomy effectively of value in its own right. As Raz puts it, "Autonomy is a constituent part of a good life."[37]That is, according to this view, without autonomy a good life is not possible: this view underlies the thesis as a whole.

This thesis also follows Raz in terms of the degree to which it is appropriate for governments to intervene in order to protect autonomy – that indeed is the ultimate purpose of the rights suggested. In these terms, there is a clear role for government in implementing and supporting those rights. In the words of Raz:

[35] RAZ, J. 1986. *The Morality of Freedom,* Oxford, Clarendon. p369
[36] See for example Rawls, in C.F. RAWLS, J. & FREEMAN, S. R. 1999. *Collected papers,* Cambridge, Mass., Harvard University Press., p365 or FABRE, C. 2000. A Philosophical Argument for a Bill of Rights. *British Journal of Political Science,* 30**,** 77-98. p77
[37] RAZ, J. 1986. *The Morality of Freedom,* Oxford, Clarendon. p408

"The doctrine of autonomy-based freedom is not inimical to political authority. On the contrary, it looks to governments to take positive action to enhance the freedom of their subjects."[40]

The autonomy that is being considered here must allow choice, it must allow changes to be made continually, and it must allow choices to come into action in reality as well as in theory. To be autonomous, therefore, meaningful choices have to be present and one needs to be given the opportunity to make those choices, appropriately informed and free from coercion, restraint or excessive or undue influence. This means not only that choices must be available, but that these choices must be meaningful – contrast the dystopian vision of Orwell's 1984 where there are almost no choices at all with that of Huxley's Brave New World in which there are a plethora of choices, all of them effectively meaningless.

Accordingly, for autonomy to function it is necessary to ensure that choice exists and that there is an opportunity to exercise choice – and, often more to the point in our modern, seemingly choice-filled society, that these opportunities are appropriately informed and free from coercion, restraint and excessive or undue influence. As will be shown through the case studies, in many current situations those who would control people, who would restrict people's autonomy, are often far too subtle to use obvious coercion or restraint – but their ability to influence and persuade is remarkable and appears to be growing, and their methods of limiting and controlling the choices available to others are becoming ever more sophisticated. Freedom from manipulation is as important in this context as freedom from coercion. As Raz puts it:

"Manipulating people, for example, interferes with their autonomy, and does so in much the same way and to the same degree, as coercing them. Resort to

[40] Ibid. p427

manipulation should be subject to the same condition as resort to coercion."[41]

When looked at in the context of the digital world, this matches closely with Nissembaum's description of the world of 'pervasive monitoring, data aggregation, unconstrained publication, profiling and segmentation':

"…the manipulation that deprives us of autonomy is more subtle than the world in which lifestyle choices are punished and explicitly blocked."[42]

Protecting people from this kind of manipulation is one of the prime functions of the rights set out in this thesis.

It is important to differentiate between capacity and opportunity. Capacity is essentially a biological question, clearly present in most adult humans, clearly absent in inanimate objects, and debatable at the margins – children, animals, the mentally ill, people in persistent vegetative states and so forth.[43] These marginal cases are not central to the issue here – as already mentioned this thesis principally concerns the mainstream, namely mentally functional adults in ordinary society, generally those with the capacity to use computers and the internet. Those on the margins may still have a need for some kinds of special protection in relation to privacy and personal data precisely because they either do not or might not pass the *capacity* test for autonomy. They may benefit from some of the kinds of right envisaged in this thesis – the personal data of many of the persons within many of these categories require specifically tailored protection, and the impact of the introduction of the kinds of rights envisaged in this thesis might well provide or enable that

[41] Ibid. p420
[42] NISSENBAUM, H. F. 2010. *Privacy in context : technology, policy, and the integrity of social life,* Stanford, Calif., Stanford Law Books. p83. The world that Nissenbaum describes draws from the work of Gandy, in GANDY, O. H. 1993. *The panoptic sort : a political economy of personal information,* Boulder, Colo, Westview.
[43] For work in some of these areas see for example JACKSON, E. 2001. *Regulating Reproduction: Law, Technology and Autonomy,* Oxford, Hart. and COCHRANE, A. 2007. Animal Rights and Animal Experiments: An Interest-Based Approach. *Res Publica,* 13**,** 26.,

kind of protection. An examination of this is, however, beyond the scope of this thesis.

What is central to this thesis is not capacity but the opportunity aspect of autonomy – and one of the suppositions of this thesis is that if someone possesses the *capacity* then they should be allowed the *opportunity* for autonomy. For the purposes of this thesis, as noted above, the concept of autonomy is understood broadly, beyond the purely individualistic and rationalistic form, in two specific ways in particular, both of which emerge from the theories of Raz:

1) Firstly, it is taken to include a 'freedom to be irrational'. Autonomy as it is commonly discussed can appear cold, based only on 'rational decision-making',[44] when people in real life often make their decisions emotionally, or based on tastes or instincts rather than values or logic. Raz's broad definition of an autonomous person as 'author' of their own life allows for that freedom. Conversely, again following Raz as noted above, autonomy should mean freedom not only from coercion but also from undue influence, unfair or excessive persuasion (which may also be based on emotions and tastes rather than values or logic) and so forth.[45] This kind of manipulation can interfere with autonomy in another qualitatively different way, ultimately resulting in alienation. As Raz puts it:

   "A person who feels driven by forces which he disowns but cannot control, who hates or detests the desires which motivate him or the aims that he is pursuing, does not lead an autonomous life. His life is not his own. He is thoroughly alienated from it."

[44] C.F. Rawls' description noted earlier, '…and rationally to pursue a conception of the good', in RAWLS, J. & FREEMAN, S. R. 1999. *Collected papers,* Cambridge, Mass., Harvard University Press., p365

[45] See for example in the context of advertising CRISP, R. 1987. Persuasive Advertising, Autonomy, and the Creation of Desire. *Journal of Business Eithics,* 6**,** 413-418.

2) Secondly, it considers people not only as individuals, but also in a social context – autonomy should include the opportunity to function fully in all the key aspects of society, without undue restriction, discrimination and the like. This might best be described as a form of 'social freedom'. An autonomous life will often – indeed usually – involve lots of contact and engagement with others, in a social context. As society develops in new ways, autonomy 'expands' to include these new aspects of society – and in particular, as will be discussed in depth below, in the digital age this means that autonomy must include the opportunity to function fully in the online world, as well as in the 'offline world.'

Autonomy set out in this way, incorporating these two aspects, has advantages in this context. Firstly, it begins to address the feminist and communitarian critiques of autonomy and privacy, which will be discussed later in this chapter, allowing autonomy to become something stronger and more rounded as a result. Secondly it takes account of the nature of the real problems found in this field – for many of the ways in which autonomy is threatened in the internet are connected with the ways that people are able to function in a social context; the methods used are often based on persuasion and emotion. In particular, practices are arising though advertising techniques that by their nature have a tendency to use psychological rather than coercive tactics. This is a theme that runs through many different parts of this thesis, but in particular in Chapter Two, which sets out the model of the Symbiotic Web, and Chapter Four, which looks at what is known as 'behavioural targeting' and is the most significant current manifestation of this kind of technique.

The internet, in its current manifestation, is a social medium. It provides a basis for communication in many forms, from email and online chat to VoIP telephony and video conferencing. It allows collaborative work on a hitherto unimaginable scale: Wikipedia is the best-known example, but there are vast numbers of others. It allows sharing of creative work via services like

YouTube and through systems like MySpace. Perhaps most significant of all is the social networking service – Facebook,[46] the most popular, was estimated to have more than 750 million active users in August 2011.[47] Some of the most valued data held about individuals is what might be described as 'social data' – who people know, who they communicate with and how. All these together mean that something which impacts upon individuals' behaviour on the internet invariably has a social as well as an individual impact – and that manipulation of an individual's behaviour often happens in a social context, from pressure to join (and then to participate in particular ways in) a social network like Facebook onwards. Autonomy, in the current context, must take this into account – hence the broadening of the definition for the purposes of this thesis.

This extended form of autonomy follows Raz beyond the more commonly used conception, but it is consistent with the essential meaning of the word ('self-government'[48]) and arguably more so than the 'purer' form, which in many ways bears less resemblance to the way that people both behave and wish to behave in the real world. Few people wish to live separately from society, and to follow purely rational thought – the ways that people 'govern' themselves include emotion (there are those who eat 'unhealthy' food, follow 'hopeless' sports teams, smoke or drink 'to excess' and so forth) and operate in a social context (hermits and recluses exist, but are very much a rarity) so 'self-government' both does and should include both those expansions.

2.2 Legal Philosophy

In legal philosophy there are two main theories about rights, the 'will theory' and the 'interests' theory. It is not the purpose of this piece to enter into the debate as to which school is better – rather, it is suggested that adherents of

[46] www.facebook.com
[47] http://www.facebook.com/press/info.php?statistics
[48] Dictionary definitions of autonomy centre around this concept. Merriam-Webster's Dictionary of Law, for example defines autonomy as 'the quality or state of being self-governing; especially: the right of self-government', (see http://dictionary.reference.com/browse/autonomy)

both would support the central importance of autonomy, and of its close connection to human rights. Whichever of the theories is followed, whether or not a person believes in natural rights or is a positivist, or if a simply pragmatic approach is followed, from a liberal perspective autonomy remains central, and rights can and do exist to support it.

Taking the will theory first, the situation is relatively straightforward – the importance of autonomy is essentially a presupposition of the whole theory, since it rests on the importance of people's choices, and autonomy is a requirement for valid, meaningful choices to exist. It may be expressed in a number of different ways – Hart, for example, talks about all rights as being effectively reducible to the 'equal right of all men to be free'[49] – but it ultimately amounts to substantially the same thing.[50]

For those who follow the interests theory the situation is a little more complex. It can be argued that there is a right to autonomy, since there is a clear interest in autonomy for the vast majority of humanity. There are potential objections to this contention, principally concerning the broadness of the concept of autonomy. Raz for example considers that the suggestion that there is a right to autonomy could fail since it would need to impose too many duties on too many people, but in practice he takes autonomy even more seriously, considering it a 'moral ideal' and joining the Will Theorists in considering it a suitable basis for rights rather than a right in itself.

For the purposes of this thesis it is sufficient to say that from both legal philosophical perspectives autonomy can be seen as a basis for rights, and one of the strongest and most important of possible bases, at least for the main body of humanity under consideration here.

---

[49] See HART, H. L. A. 1955. Are There Any Natural Rights? *Philosophical Review,* 64, 175-191.

[50] Another Will Theorist, Gewirth, expressed it more directly: 'All the human rights, those of well-being as well as of freedom, have as their aim that each person have rational autonomy in the sense of being a self-controlling, self-developing agent who can relate to other persons on a basis of mutual respect and cooperation, in contrast to being a dependent, passive recipient of the agency of others' – from GEWIRTH, A. 1982. *Human Rights: Essays on Justification and Applications,* London, University of Chicago Press.

2.3 Historical, natural rights and positivist perspectives

Given this, it is not surprising that autonomy and its related concepts have historically taken a central place in human rights documents, albeit often not explicitly referred to under this label. Thomas Paine, for example, in The Rights of Man, considered 'natural rights' to include 'all those rights *of acting as an individual* for his own comfort and happiness'[51] – i.e. all those rights which support an individual's autonomy. Not only believers and supporters of 'natural rights' took autonomy seriously – Bentham, one of the staunchest and most effective opponents of the very idea of natural rights, considered that one of the functions of the law was to allow individuals to form and pursue their own conception of well-being,[52] something close to the definitions used earlier in this chapter. In a manner that draws parallels with the current disagreements between Will and Interest theorists, Bentham and Paine disagreed deeply about the existence and effectiveness of 'natural rights', but they both thought autonomy of the most fundamental importance. For JS Mill, the significance of autonomy was even clearer: indeed, autonomy could be said to be the most important aspect of his political philosophy.[53]

More recently, the approach adopted in the plethora of human rights conventions and declarations that have emerged after the Second World War has followed a similar pattern, supporting autonomy as an underlying concept upon which to base detailed, substantive rights. In the Universal Declaration of Human Rights, for example, articles refer to human beings being 'born free' and 'endowed with reason and conscience',[54] and that everyone has a 'right to liberty'[55] in a general and undefined way before going on to talk about specifics.

---

[51] PAINE, T. & BURKE, E. 1791. *Rights of Man: Being an Answer to Mr. Burke's Attack on the French Revolution,* Dublin, [s.n.]., p34
[52] See for example KELLY, P. J. 1990. *Utilitarianism and Distributive Justice: Jeremy Bentham and the Civil Law,* Oxford, Clarendon., particularly pp102-103
[53] See MILL, J. S. & HIMMELFARB, G. 1982. *On Liberty,* Harmondsworth, Penguin., as well as much of the subsequent study of Mill's philosophy.
[54] UDHR Article 1
[55] UDHR Article 3

Conversely, the European Convention for the Protection of Human Rights and Fundamental Freedoms makes no direct mention of anything directly resembling autonomy, but includes specific rights that are closely connected to it – for example, respect for family and private life,[56] freedom of thought, conscience and religion[57], expression[58], and assembly and association.[59] It is into this spectrum of rights that the proposed rights governing informational privacy is intended to fit. Indeed, as will be shown, the rights proposed help support, amongst other things the rights of individuals to online privacy, online freedom of expression, online assembly and association and so forth.

Taking this approach a step further, one way to look at this issue is to consider autonomy as an aspect of freedom. Indeed, it is difficult to frame a conception of freedom that does not include some degree of autonomy. As a consequence, if rights supportive of freedom are posited, then so must these rights also be supportive of autonomy – without autonomy, people are not really free.

Further, there is little doubting the power of human rights in today's world, whether or not someone believes human rights spring from the most solid of foundations (religious or otherwise), are often born at least in part, and unashamedly from emotion and sentimentality (as Rorty appears to suggest),[60] or do not really exist at all (as MacIntyre contends).[61] They appear to be, as Klug puts it, an idea whose time has come.[62] On its own this is a pragmatic reason for using the rights-based approach, provided that the importance of autonomy is accepted as within it.

---

[56] CPHRFF Article 8
[57] CPHRFF Article 9
[58] CPHRFF Article 10
[59] CPHRFF Article 11
[60] In RORTY, R. 1993. Human Rights, Rationality and Sentimentality. *In:* SHUTE, S. & HURLEY, S. L. (eds.) *On Human Rights: Oxford Amnesty Lectures.* Oxford: BasicBooks.
[61] In MACINTYRE, A. 1981. *After Virtue: A Study in Moral Theory,* London, Duckworth., p67, MacIntyre famously suggests that 'There are no [human] rights, and belief in them is one with belief in unicorns and witches'.
[62] KLUG, F. 2000. *Values for a Godless Age: The Story of the UK's New Bill of Rights,* London, Penguin.

This thesis does not suggest that such pragmatism should be relied upon, or that any particular school of legal philosophy be followed. It is suggesting that whether it is looked at from a philosophical, historical, positivist, instinctive or pragmatic perspective, for the vast majority of humanity – and certainly those under consideration here – autonomy is an appropriate basis for establishing human rights. It may be the most appropriate basis possible if, as in this thesis a liberal, democratic stance is taken.

2.4 Limitations to Autonomy

This form of autonomy, extended as discussed above, is not absolute and is compromised, limited and restricted in many ways. Keeping the balance between the needs and wishes of individuals and the requirements of governments and communities (and indeed the needs and wishes of other individuals) means that individual freedom (and therefore individual autonomy) cannot be unrestricted. The key question is which restrictions are to be considered acceptable – or even necessary – and which of them are to be rejected. Essentially, a restriction to autonomy can be viewed as acceptable, appropriate or necessary if the benefits that arise as a result of that restriction are of sufficient importance, such as when the restrictions protect the rights of others or protect the nation or community – the criminal law is perhaps the most obvious example of this.[63] As noted above, the 'Razian' understanding of autonomy used in this thesis also allows for – indeed requires – intervention by governments in support of autonomy.

Part of the contention of this thesis is that many of the threats to autonomy that result from deficiencies in informational privacy are of sufficient moral and practical importance to be unacceptable, and that therefore they are threats which require to be addressed. This is because:

- They are often covert, and based on information asymmetry

[63] For example, see Raz's work on the relationship between authority and autonomy, particularly in chapters 2-4 of RAZ, J. 1986. *The Morality of Freedom,* Oxford, Clarendon.,

- They have a real and increasing effect on both real lives and online activities
- They can be discriminatory and reinforce existing and pernicious imbalances of power
- They tend to be open-ended in scope, and to be increasing in significance as people's reliance on data and in particular online activity for societal functions increases

And crucially from the perspective of the justifiability of limitations, the compensating 'goods' (most directly financial benefits to business and, more debatably, security benefits to governments) are insufficient to make the loss of autonomy acceptable on any analysis however generous of the breadth of such limitations. Taking this a step further, it will be suggested in the thesis that many of these 'goods' are not even as 'good' as they appear to be. Specifically, Chapter Two will set out the underlying theory, the Symbiotic Web, concerning the mutual dependence of businesses and individuals in the current state of the internet – a theory which ultimately suggests that unless businesses find ways to be more open, direct and cooperative with their customers and potential customers, the economic benefits they enjoy from the internet will be short-lived.

What is more, as will be shown, the threats to autonomy posed by current practices compromise key human rights – most directly 'civil liberties' such as freedoms of assembly, association, thought and religion, but also other rights, including social, cultural and economic rights. Those direct threats also need to be taken seriously, for they impact not only upon our 'online' lives, but also upon the real lives of increasing numbers of people throughout the world.

3 Autonomy, rights, challenges and criticisms

As discussed above, this thesis takes an essentially liberal, democratic, rights-based approach. What is being suggested here in terms of rights (including

the related and supporting concepts of Collaborative Consent and Autonomy by Design) are intended to support the idea of giving individuals more autonomy both in their lives online and in the 'real' world – but neither autonomy itself nor the idea of privacy, and in particular informational privacy on the internet, are ideas that are universally accepted as having central (let alone fundamental) importance. There are criticisms both of the concept of autonomy and of its specific application here – the most important of them are set out in this section, together with a preliminary indication of how this thesis intends to answer them. It should be noted that whilst these critiques fuel approaches to exceptions to autonomy-based rights that are very broad, they do not undermine rights talk in today's society. Rather they make this talk more porous and vulnerable to sanctioned exception – and hence less effective in real terms.

The primary challenges and critiques can be broadly divided into five: two essentially pragmatic challenges – the security and economic challenges – based on specific issues concerning the internet, and three theoretical critiques – the communitarian, feminist and transparency critiques – which, though of general application, have a particular relevance to the field under scrutiny too. These challenges and critiques are related and overlapping, both from a theoretical and a practical perspective. The security and economic challenges, in particular, can both be viewed from a communitarian perspective – maintaining security and prioritising business success over individual privacy and autonomy can both be seen as communitarian and even utilitarian goals. All five critiques and challenges question priorities and motives – and all five have strong arguments and powerful supporters.

3.1 The Security Challenge

The essence of the security challenge is that it suggests that excessive rights to privacy tend to protect the criminal and terrorist – and, ultimately, that autonomy is less important than security. This is played out in a number of ways.

First of all, the internet can be seen as a 'tool for terrorists', for example to spread their 'hate speech' – the 2010 case of Roshonara Choudhry, the young woman who stabbed her MP, Stephen Timms at his surgery in East London is a prime example. According to reports, she had been 'radicalised' by watching video 'sermons' by Anwar al-Awlaki, a radical Muslim cleric on YouTube.[64] Moreover, many of the tools that have made the internet such a success – near-instant, worldwide communications, the apparent opportunity for anonymity, relative cheapness and so forth – make it ideally suited for small, mobile and geographically spread organisations to function.

The general approach to security matters, particularly evident in the concept of 'data retention', which plays a central part in Chapter Three, is to make surveillance as general as possible, and gather as much data as possible, with the idea that by sifting through all that data it will be possible to find what is needed to catch terrorists and other threats to security. Surveillance should, under these terms, be universal – there should be no exceptions. However, it will be suggested in this thesis that the idea of general surveillance and universal data gathering is not something that should be accepted [65]– and that it is not the only possible approach. Specific solutions to some of the particular problems in security that do not impinge so directly on general privacy will be suggested, particularly in Chapters Three and Seven. It is difficult, however, to make an entirely convincing case for privacy in the security field, as the security forces are not, for understandable reasons, willing to provide complete information either as to their methods of countering terrorism or as to how successful particular tactics have proven to be. Privacy advocates, as a consequence, cannot point to evidence to suggest let alone prove that the idea of general data retention or universal surveillance is ineffective or inefficient in the field of counter-terrorism. It is equally true that those in favour of such tactics cannot prove that they are

[64] See for example http://www.bbc.co.uk/news/uk-11686764

[65] This suggestion is supported by the ruling in *Liberty v UK [2008] 48 EHRR 1*, where the European Court of Human Rights held that the broadness and lack of clarity of the rights of government agencies to intercept communications set out by the *Interception of Communications Act 1985* led to a breach of Article 8 of the ECHR.

effective – but they can argue that the risks involved are sufficient to warrant the tactics.

However, as Gearty puts it, "In taming counter-terrorism law human rights has the chance to renew its soul"[66]: addressing the security challenge over data and surveillance on the internet is a prime example of what is required. The presumption of innocence, making surveillance, data collection and its equivalents the exception – and an exception that must be justified – rather than the rule, the paradigm shift suggested in 1.3 above, is a key element of that. There are parallels here with the transparency critiques below and in particular with responses to the suggestion 'if you've got nothing to hide, you've got nothing to fear' so often deployed by 'security advocates' in favour of what amounts to general internet surveillance. As Solove argues,[67] this 'reason' has many flaws, some fatal when looked at closely with anything resembling a liberal standpoint: this is something dealt with in more detail below when looking at transparency critiques, but it remains an issue that needs to be addressed.

3.2 The Economic Challenge

The essence of the economic challenge is that excessive support for privacy reduces business opportunities and the niceties of individual autonomy are less important than a thriving economy. From another angle it can be argued that it may be impossible to overcome the power and momentum of business in this field.

It is a central contention of this thesis that the economic benefits of abuses of privacy and autonomy are short term. For long term and sustainable economic benefits, a positive relationship between the needs and desires of individuals must be held in balance – and that, ultimately, means more

[66] The sixth point of his 'Manifesto' which makes up the central part of the collaborative web project 'The Rights' Future', at http://therightsfuture.com/manifesto/
[67] Particularly in his article SOLOVE, D. J. 2007. "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review,* 44.

respect for the privacy and autonomy of those individuals. The model of the Symbiotic Web, as mentioned above and as detailed in Chapter Two – and as demonstrated though the case studies in the thesis – suggests that where these needs are not respected, businesses are likely to suffer or even fail.

The economic challenge remains a crucial challenge to the whole concept of privacy. Businesses are the key drivers in the development of both the technology and the reality of the internet, and particularly in the current climate the idea of persuading businesses to change their methods for a perceived and potentially highly debatable long term future benefit is a very difficult one to put over. Moreover, businesses are unlikely to be persuaded by any kind of a moral or philosophical argument – they will respond only to the demands of law, of finance, or of their customers. That is what underlies the suggestion in this thesis of an approach using symbiotic regulation – effectively, of working through precisely those mechanisms, the relationships between businesses and their customers, their competitors and so forth. It is crucial to understand, however, that the rights envisaged in this thesis are not intended to override the freedoms that businesses require in order to thrive. Rather, the competing rights and freedoms must be kept in balance – it is through that balance that the beneficial aspects of the symbiosis set out in chapter 2 is maintained and supported. The way in which this could work is discussed in more depth in Chapter Six.

3.3 The Communitarian Critique

The communitarian critique suggests that privacy and individual autonomy prioritize the individual over the community and in some ways misunderstand the essentially social nature of humanity. In the communitarian critiques, the suggestion is that individual autonomy is given inappropriate priority over the needs of the community as a whole.[68] Privacy,

[68] See for example Charles Taylor, who describes individualism as the first 'source of worry', in relation to the 'malaises of modernity'. Taylor, recalling Alexis de Tocqueville,

according to this account, provides excessive benefits to individuals and often individuals with little need or right to such benefits. The way that privacy law has often focussed on cases of celebrities, politicians and sports stars is just part of how this plays out.

Communitarian critiques are addressed partly through the breadth of the definition of autonomy being used, as noted above, and by considering rights to both autonomy and privacy only in balance with the needs of society. It is a delicate balance, and one that needs constant monitoring and rebalancing. However, there is another factor that comes into play here – a consideration of how privacy, anonymity and autonomy can play key parts in protecting the functioning of communities on the internet. Indeed, the ability of individuals to fulfil their full role in a community is one of the things that the rights proposed are specifically designed to protect, and a fracturing of online communities is one of the specific dangers discussed in this thesis – a risk inherent in the current system as it is developing, and as set out in the model of the Symbiotic Web, as discussed above and in more detail in Chapter Three.

The communitarian critique has, in some forms, a close relationship with the security challenge – certainly the Chinese government would make use of both arguments in favour of their approach to internet surveillance and control. They do what they do, they say, both to ensure security and to support the coherence and stability of their community.[69] From a political standpoint, once more it is a question of where the balance is placed – and what the overall result of policies produces.

---

suggests that 'the dark side of individualism is a centring on the self, which both flattens and narrows our lives, makes them poorer in meaning, and less concerned with others or society.' See TAYLOR, C. 1992. *The ethics of authenticity,* Cambridge, MA, Harvard University Press., Chapter 1. Other communitarian critics of individualistic autonomy include MacIntyre and Sandel

[69] See for example the Chinese government's white paper 'The Internet in China' (available online at http://www.china.org.cn/government/whitepaper/node_7093508.htm). In it, the Chinese government suggests that provisions forbidding amongst other things 'damaging state honor and interests; instigating ethnic hatred or discrimination and jeopardizing ethnic unity; jeopardizing state religious policy, *propagating heretical or superstitious ideas; spreading rumors, disrupting social order and stability*;' must be followed.

The communitarian critique of privacy and autonomy is related to a broader challenge on communitarian grounds to the whole concept of human rights. That challenge was met, at least to an extent, by the inclusion of economic, social and cultural rights into human rights as a whole, most notably through the International Covenant on Economic, Social and Cultural Rights.[70] In a similar manner, the challenge is met here through the expansion of the concept of autonomy to include what is described in 2.2 above as social freedom.

3.4 Feminist Critiques

The essence of feminist critiques is that protection for privacy and autonomy is often a force for patriarchy and conservatism, and that common conceptions of autonomy are generally a reflection of masculine values. Some aspects of the feminist critiques follow somewhat similar lines to those used in the communitarian critique: the whole idea of autonomy is one that is based on male concepts of independence, rather than the more social and in particular family-based ideals that can sometimes be seen as female.[71] Privacy can be used to protect the dominant position that men have in society – to allow them to maintain their power.

Taking this a step further, privacy can be used as a way of protecting those who abuse their dominant positions, by preventing their abuses from becoming known. The way in which parts of the Catholic Church have at times tried to protect child-abusing priests is an extreme example, more mundane examples might include men using privacy to hide their extra-marital affairs from their wives, or finding ways to conceal their assets from

[70] International Covenant on Economic, Social and Cultural Rights, 1966, downloadable from http://www2.ohchr.org/english/law/cescr.htm

[71] See for example ALLEN, A. L. 2003. *Why Privacy Isn't Everything: Feminist Reflections on Personal Accountability,* Lanham, Maryland, Rowman & Littlefield. And ALLEN, A. L. & MACK, E. 1991. How Privacy Got Its Gender. *Northern Illinois Law Review,* 10**,** 441-471. A full examination of Feminist perspectives on privacy is beyond the scope of this thesis, but the strength and importance of their arguments needs to be acknowledged.

their wives to avoid them being taken into account for divorce settlements. In all these cases, it is about privacy being used to protect the power of those who have it, against just claims that are being brought - the furore over the Ryan Giggs 'super-injunction' in 2011 is perhaps the most dramatic recent example.[72]

Feminist critiques are addressed in this thesis again partly through defining autonomy to include 'social freedom' as noted above and partly by emphasizing the positive role that privacy rights (particularly on the internet) can play in protecting precisely those aspects of private life that the feminist critique considers important. The kinds of privacy and anonymity that the internet allows can be an enabler and a liberator for women – allowing women and girls the right to roam the net without their gender being revealed, for example, reducing the risks of discrimination of many different kinds. Moreover, similar methods to the suggestions for alleviating security issues that will be suggested in Chapters Three and Seven could help solve some of the particular problems in this area – catching abusers, child pornographers, stalkers and so forth in the private sphere in the same ways that you can catch terrorists.

The most important and effective answers to both the communitarian and feminist critiques, however, lie in an examination of what the internet would look like, and how it would function, from the perspectives of both women and communities, if the rights suggested in this thesis were brought into action. That kind of an examination is a central part of Chapter Seven – and it is the contention of this thesis that a future internet in which these rights were understood and respected would be significantly better from both of those perspectives.

3.5 Transparency critiques and challenges

[72] The case involved was CTB v News Group Newspapers [2011] EWHC 1232 QB.

Transparency critiques revolve around the idea that privacy is an outdated value or is to all intents and purposes unenforceable in today's technological society, and that what is needed is an adjustment to a new way of living.[73] In many ways they are more complex than the critiques and challenges that have been discussed so far – it could be argued that they are more challenges than critiques, a kind of 'technology challenge' or 'information society challenge', something said to have been made inevitable as a result of the developments of technology. Do individuals need to give up privacy in order to function in the new, online world? Whether it is called a critique or a challenge, meeting it is central to this thesis: ultimately this boils down to the question of whether it is possible to have privacy in this new world. This thesis will suggests that not only is it possible, but it is crucial that a way is found to ensure that it takes place.

One example of just such a risk is to be found in looking at the future of the web itself. Web 3.0, the semantic web, as it is envisaged by Berners-Lee and others has the potential to be a great benefit to autonomy, increasing freedom and choice, giving individuals more power to control their own lives[74] – but it is an essentially technological development, and technology is value neutral. What governs the future of the web is not just the technology but *all* aspects of the web from laws and government actions, business models and commercial practices to community attitudes and social interactions.[75] As the current development of the web symbiosis is beginning to show and as will be discussed in Chapter Two, the potential of the technology could be turned into precisely the opposite of the laudable and liberating visions of Berners-Lee and others, something that controls us more rather than providing us with more freedom. It is the contention of this thesis

[73] See for example BRIN, D. 1998. *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?,* Reading, Mass., Addison-Wesley.
[74] See BERNERS-LEE, T. & FISCHETTI, M. 2000. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor,* New York, HarperCollins Publishers., particularly Chapter 12
[75] The recognition of this is inherent in the work of cyber-regulation theorists from Lessig to Murray – see for example LESSIG, L. 2006. *Code: Version 2.0,* New York, Basic Books. and MURRAY, A. D. 2006. *The Regulation of Cyberspace: Control in the Online Environment,* Milton Park, Abingdon, UK ; New York, NY, Routledge-Cavendish.

that privacy in general, and informational privacy in particular, is the key to ensuring that this does not happen.

There are three principle variants of the transparency critique:

1) That the struggle for privacy is already lost;
2) That the struggle for privacy is outdated; and
3) That the struggle for privacy is 'wrong', and that we should in fact 'embrace' transparency and make lack of privacy a virtue to be enjoyed.

Scott McNealy, then CEO of Sun Microsystems, expressed the first of these directly when he told reporters in 1999 "You have zero privacy anyway, get over it."[76] Mark Zuckerberg, co-founder and CEO of Facebook, is perhaps the best known proponent of the second version – essentially his argument has been that given that more than half a billion people use Facebook and put up some of their most personal information there, that means that people are no longer really interested in privacy.[77]

The third variation is more complex – versions of it have been around since David Brin's 1998 work 'The Transparent Society'.[78] More recently, another angle of it has emerged in Bell and Gemmell's 2009 book 'Total Recall: How the E-Memory Revolution will change everything',[79] though it has also faced its antithesis in Mayer-Schönberger's 'Delete: the virtue of forgetting in the digital age'.[80] This last critique has a close relationship to the old ideas 'If you've done nothing wrong, you've got nothing to fear', or 'If you've got nothing to hide you've got nothing to fear', referred to above in 3.2,

[76] Quoted for example in Wired, at http://www.wired.com/politics/law/news/1999/01/17538
[77] See for example Chris Matyszczyk's blog on CNET, at http://news.cnet.com/8301-17852_3-10431741-71.html. Zuckerberg has been making related statements to different elements of the media for much of 2010.
[78] BRIN, D. 1998. *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?,* Reading, Mass., Addison-Wesley.
[79] BELL, C. G. & GEMMELL, J. 2009. *Total recall : how the E-memory revolution will change everything,* New York, N.Y., Dutton.
[80] MAYER-SCHÖNBERGER, V. 2009. *Delete : the virtue of forgetting in the digital age,* Princeton, N.J., Princeton University Press.

concerning the security challenge. The idea that lack of privacy should be embraced as something positive and transformative to an extent presupposes the truth of the 'nothing to hide/nothing to fear' argument. The converse is also true, at least to a certain extent. That is, if the 'nothing to hide/nothing to fear' argument is false or flawed, then embracing transparency would need to be done extremely carefully.

Solove, particularly in his 2007 piece "'I've got nothing to hide' and Other Misunderstandings of Privacy'[81] challenges that argument convincingly, demonstrating amongst other things how the idea is based on a very limited understanding of what privacy really means. The arguments made by Solove are detailed and compelling, looking at the problem from many different angles. One of particular relevance here is the quote that Solove makes from Judge Richard Posner, that in Posner's view (presented as representative of part of the 'nothing to hide' school) privacy involves a person's 'right to conceal discreditable facts about himself'.[82] In the context of the internet, and particularly when the concepts of profiling, targeting and data aggregation are considered, it is not only the discreditable facts that are at issue. Indeed, as will be shown throughout this thesis, it is often the least obvious, the most apparently mundane or inconsequential of information that makes the difference.

Solove's arguments are powerful, but even so the transparency critique is a strong one in all of its forms, and needs to be carefully addressed. If McNealy is right, this whole discussion is pointless. If Zuckerberg is right, it is an argument that is becoming more and more irrelevant day by day. If Brin, Bell & Gemmell and others are right, then what is being suggested in this thesis is retrograde and regressive. This thesis can and does answer all three forms of the transparency critique. The case studies will show that people do still care about privacy, and that they are quite capable of winning battles in the area

[81] See SOLOVE, D. J. 2007. "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review,* 44.
[82] Quoted by Solove, in Ibid. originally in POSNER, R. A. 1998. *Economic analysis of law,* Boston, Aspen Law & Business.

of privacy when the situation is appropriate, when they have the right support, and when, using the terms set out in Murray's symbiotic regulation, they find ways to use the regulatory matrix in an appropriate and effective way.

4 Autonomy on the Internet – and rights to protect it

The research question for this thesis asks whether deficiencies in data privacy threaten our autonomy and if so, can informational privacy rights meet this threat. In order to answer this question, the following has been set out so far.

Firstly, that autonomy is of central importance from a liberal perspective, both philosophically and practically – so the research question is of importance. The concept of autonomy has been examined – and, for the purposes of this thesis, extended to include both irrational and emotional wishes and freedom in a social context, to reflect a more nuanced appreciation of the way that people live and wish to live. It is suggested that autonomy is therefore a suitable basis for rights – that rights established to support or protect autonomy are rights in a real sense.

Secondly, that the internet is now intrinsic to life in a modern society. The role of the internet in this society has been looked at, and some of the most important issues surrounding its role have been set out. In particular, two key questions about that role have been asked: how 'public' is the internet, and to what extent is personal data 'ours'? More specifically, what rights do people have – and should people have – over that personal data? The answers to those questions underpin the conclusions of this thesis, and will emerge through the case studies in the chapters that follow.

Taking these two questions together, the logical corollary is that since the use of the internet is now intrinsic to life in a modern society, autonomy needs to be supported and protected in online life. Specifically, that what is needed is

protection for autonomy on the internet. That then leads to the main thrust of this thesis. Do deficiencies in data privacy threaten that autonomy? If they do, can they be protected by rights – and if so, what rights?

The rights suggested in this thesis are intended to fulfil this role: they are designed with autonomy in mind. The *right to roam with privacy* is intended to protect the individual from the kind of manipulation discussed by both Raz[83] and Nissenbaum[84] – for if the identity and key characteristics of the individual browsing are not visible or 'useable' by others as the individual browses, the opportunities for manipulation are much reduced. Moreover, it can begin to help with what was described earlier as the converse of the panopticon effect: if we don't feel ourselves to be under the constant risk of observation, we will feel more able to act freely.[85]

The *right to monitor the monitors* is also a right grounded in autonomy – for by monitoring those who monitor us we are able to get a better understanding of how and why information is being gathered, so we can make choices that are better informed and more likely to match with our intentions and desires.

The *right to delete,* as well as providing another choice and opportunity to exercise that choice – benefits to autonomy in their own right – has specific relevance to the deletion of the kind of data used to profile, and indeed the profiles themselves. As for the right to roam with privacy, this is particularly important in relation to the way in which we may be subject to manipulation and to control. Moreover, as shall be discussed in depth in Chapters 5 and 6, the idea behind the right to delete is to encourage the development of different kinds of business models, models less dependent on the gathering and holding of personal data, and of the systematic use of that data.

[83] See note 41 above
[84] See note 42 above
[85] See Section 1.4 above

That, indeed, is the key point of all of the rights, and perhaps the key point of the thesis itself. The intention behind the rights is to aid in the development of an internet that is more 'privacy-friendly' and 'autonomy-friendly'. An internet where individuals, businesses and governments function in a better, more consensual balance – where business models are leaner, more efficient and less deceptive. Where businesses 'rights' to pursue their profits are in a better balance with individuals' needs for autonomy and rights to privacy.

How that might be possible, and what that future internet might look like, is discussed throughout this thesis, but in particular in Chapter 7. The starting point, however, is to get a better understanding of how the current, substantially commercial online world functions. This is what is now explored in Chapter Two, where the model of the Symbiotic Web is set out in detail.

Chapter 2: The Symbiotic Web

1 Autonomy and the Internet

In Chapter One, three key ideas were set out. These were firstly, that the concept of autonomy is important – even central – to the Western, liberal view of life; secondly, that this concept of autonomy includes an opportunity to fully and freely participate in society; and thirdly, that in the digital age, to function 'online' is not an optional extra but an intrinsic part of participating in society. The consequence of these three is that if the Western, liberal approach is taken, the opportunity for individuals to function fully and freely online is of central importance. What is more, privacy, another concept important in the Western liberal approach, plays a key part in protecting this autonomy – so protecting privacy as well as directly protecting autonomy is of great significance.

On the surface, the web appears to be a place where there is more freedom to move and act than in real life – international borders can be crossed, often without the surfer even knowing that they are being crossed, censorship appears much more limited, and one's identity is seemingly protected. The famous cartoon in the *New Yorker* that proclaimed that 'on the internet nobody knows you're a dog' reflects the common perception of such anonymity.[1] The reality is very different – as described below, and more extensively in Chapter Four, not only is it possible to know that you're a dog, but to know what breed you are, the names of your doggy friends, which cats you chased yesterday, and what kind of dog-food you prefer.

Added to this, as Lessig,[2] and subsequently Murray[3] have shown, the ability of those who create the web-pages that people visit – the code writers – to

[1] The cartoon by Peter Steiner, was published in 1993. In 2000, the New York Times published a piece entitled 'Cartoon Catches the Spirit of the Internet'. See http://www.nytimes.com/2000/12/14/technology/14DOGG.html?pagewanted=1&ei=5070&en=f0518aafeccf36fd&ex=1183089600

[2] This is the subject of LESSIG, L. 2006. *Code: Version 2.0,* New York, Basic Books.

control people's actions and opportunities, makes it an environment in which people are potentially subject to much less freedom and less autonomy than in the real world. This means that it is of great importance that what it means to function fully and freely online is understood, as well as what it is that stops this from being possible now – and just as importantly what could stop it from being possible in the future. To do this, there needs to be a better understanding of how the internet functions. In particular, the implications of current developments in the new, substantially commercial form of the World Wide Web need to be understood, as those developments are already having a direct impact on privacy and autonomy.

2 The Symbiotic Web

This chapter sets out a model – the Symbiotic Web – that helps with understanding these developments and their impact. It explains many of the current trends – not least the way in which the gathering of data has accelerated and expanded far beyond the conceptions of most commentators of even just a few years ago. This model forms a backdrop for much of the rest of this thesis, framing the intellectual context for the substantive case studies that make up the core of Chapters Three to Five, helping to explain their significance and their implications. Through the understanding that the concept of the Symbiotic Web provides, it becomes possible to see the potential ways forward taking shape in a more logical, connected form. The solutions proposed in this thesis and set out in Chapter One – rights-based solutions, balanced with economic and security interests, enacted through what Andrew Murray describes as 'symbiotic regulation'[4] – arise through the understanding achieved by use of this idea of the Symbiotic Web.

The symbiosis being described here is essentially benign – it lies behind many of the most positive developments on the internet and has produced a

[3] MURRAY, A. D. 2006. *The Regulation of Cyberspace: Control in the Online Environment,* Milton Park, Abingdon, UK ; New York, NY, Routledge-Cavendish.
[4] Ibid., Chapter 8

massive expansion in attractive and productive products and services available on the internet. It is a symbiosis that benefits both the individuals who use the internet and the businesses that provide services through the internet. Nevertheless, there are risks associated with its symbiotic nature, and there is a danger that it could develop into something malign, twisting the mutually beneficial symbiosis into a harmful parasitism, and producing a fractured web, manipulating and controlling those who use it. Berners-Lee and others have set out visions of the future in which the internet becomes more personalised, and users have more and more control.[5] What the malign version of the Symbiotic Web suggests is precisely the opposite: that control could be being taken out of the hands of the users, choices being made for them, rather than by them, and not necessarily for their benefit, but rather for the benefit of those wielding that control.

What is more, the symbiosis that is currently manifesting itself in the web is one that is in some ways duplicated in our 'offline' world. The gathering and use of data in these ways is not limited to online activities, but is happening to a greater or lesser extent throughout our society, from supermarket loyalty cards to transport payment systems like London's Oyster cards.[6] Moreover, as already mentioned in Chapter One the borders between the online and offline worlds are becoming more and more blurred – as demonstrated forcefully by Google's 'streetview' system.[7] It is not just online businesses that are becoming dependent on the use of personal data, and not just people who spend significant time online whose data are being gathered and used. The issue is becoming more pervasive all the time – and the need for solutions is becoming ever more important. The solutions offered in this thesis point us in a direction that will help solve – or at least ameliorate – the problems identified, problems that go beyond the virtual into the 'real' world.

---

[5] As described in BERNERS-LEE, T. & FISCHETTI, M. 2000. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor,* New York, HarperCollins Publishers.
[6] https://oyster.tfl.gov.uk/oyster/entry.do
[7] http://maps.google.com/help/maps/streetview/

2.1 What is the Symbiotic Web?

Symbiosis is where two different kinds of organism exist together in a state of mutual dependence. It can be mutually beneficial – examples occur in nature such as the birds that live on rhinos, eating the insects that infest the rhino's skins, feeding the birds and cleaning the rhino – but it can also be detrimental to one of the parties, like the tapeworms that grow inside people's stomachs, leading to sickness and even (on occasion) death.

A form of symbiosis is developing on the web. Individuals and commercial enterprises are mutually dependent: enterprises have built business models reliant on a currency of personal data, while individuals depend on 'free' access to many services, from search engines to price comparison services, social networking sites, media services such as YouTube and so forth – many of the services which, as discussed in Chapter One, now form an intrinsic part of modern life. These 'free' services use personal data, obtained through various overt and covert means, as their way of generating revenues – through targeted advertising, profile building, and the direct sale of personal data amongst other things. Even many of those services which are not free have moved towards this kind of symbiotic state, gathering personal data as part of their process in exchange for personal information – offering discounts for buying online or 'personalised services' such as, for example, the iTunes 'Genius Sidebar', which selects music for each user based on a profile they've build up which is in turn rooted in that user's musical taste.[8]

What we as individuals are doing is sacrificing one kind of freedom – 'liber' freedom, our privacy and autonomy – for another, 'gratis' freedom, receiving services and convenience without having to pay for them in the pecuniary sense. As the adage goes, there's no such thing as a free lunch – effectively we are paying for these services through the surrender of our private information, and ultimately, as will be outlined below, through giving up part of our autonomy. Conversely, the enterprises are sacrificing the opportunity

[8] Introduced as part of iTunes 8. See http://www.apple.com/itunes/features/#genius

to make money for a less tangible form of reward – information about their potential customers that may or may not be able to be transformed effectively into financial rewards at a later date. So far, for many businesses (most clearly Google and Facebook and other similar though less well-known services) these rewards have been substantial. Whether they continue to be so is another matter – but business models and ways of operating have been built on the assumption that they will be, models that can only effectively function if the gathering and use of personal data continues unabated. Indeed, as more businesses shift into this way of being, this gathering and utilisation of data can only be expected to increase.

The implications of this symbiosis are significant. It helps to explain many of the most important things that are happening in the field. It also explains why so much personal data are gathered. Further, it can help us to understand what kinds of data are being gathered, and by whom, and the principal purposes to which such information is being put commercially. Understanding the nature of this symbiosis can also provide good indications as to the ways in which this data may be used in the future – as well as why companies are less than eager to be open about either the data gathering or its purposes. It can also help us to understand the threats to our privacy and autonomy that are arising – and the further threats that might arise in the future – as a result of the ways that data are being gathered and used, and will be gathered and used in the future.

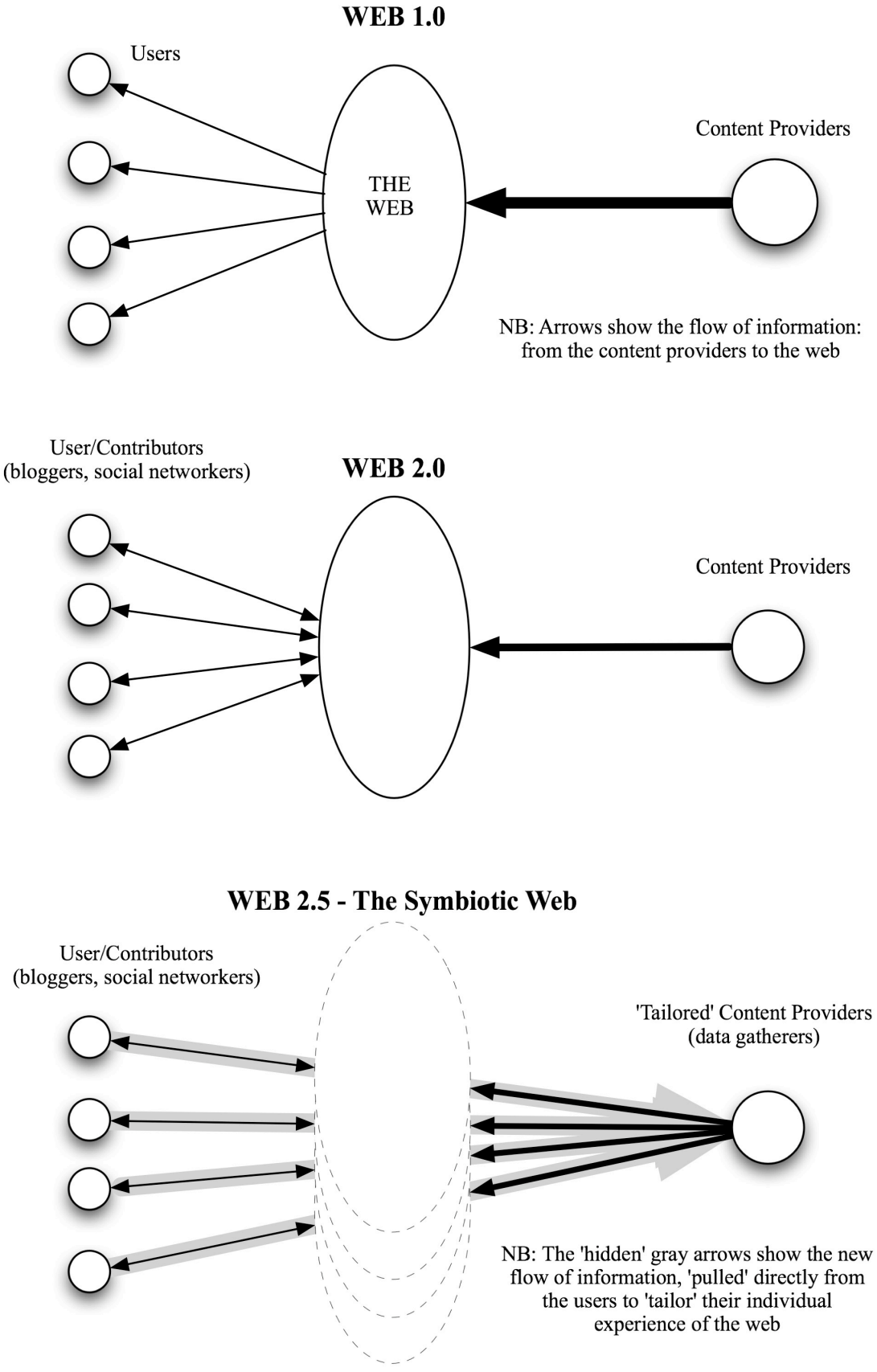
As will be detailed below, these threats are complex and manifold, but they all centre around the ways in which data can be used to persuade or manipulate people into doing what the data-user rather than the data-subject chooses, whether this means buying products or services, seeing or reading information, visiting particular websites, supplying certain information, or anything else. The implications for autonomy are clear, but it must be remembered that much of this kind of thing can be and is beneficial – links and information which is well chosen, targeted advertising which is appropriate and useful and so forth. The economic model that provides free

services to users is the one that most appreciate. It is at present, in general at least, a mutually beneficial form of symbiosis. What is important is to understand the risks associated with it, and the ways in which it can become something negative – and what can be done to prevent this from taking place.

2.2 Web 2.5: the evolution of the Symbiotic Web

It is appropriate to begin by enquiring into how this symbiosis has come about. Consider the functional evolution of the web from the beginning – in particular, look at the development of what is commonly referred to as 'Web 2.0' (see Figure 1 below). The Symbiotic Web can be considered a form of 'Web 2.5', a further development of Web 2.0.

Figure 1: Web 2.5



In its first form (which can be labelled 'Web 1.0'), the web was for almost all intents and purposes an 'information bank'. 'Content providers' put information up onto the web, while 'users' accessed and downloaded that information. The flow of information was effectively one-way: from the content providers to the users.

The much-discussed 'Web 2.0' is characterised largely by a transformation in the 'users'. Rather than simply accessing information provided for them, users began to supply information themselves. This data was provided through a wide range of 'Web 2.0' sites or applications – through web-logs ('blogs'), 'wikis' (collaborative websites such as the web-encyclopaedia wikipedia.com), social networking sites like Facebook or MySpace, media sites like YouTube, and expansions in old fashioned message-boards and equivalents. In Web 2.0, information flows into the web not only from the content providers, but also from the users themselves.

In the shift from Web 1.0 to Web 2.0, information started to flow both ways for the users. The shift from Web 2.0 to the Symbiotic Web, Web 2.5, is characterised by a converse transformation for the erstwhile 'content providers'. Not only do they provide information, but they extract it from the users as well, using a wide variety of techniques – from monitoring their activities on line to persuading users to volunteer as much personal information as possible. This personal information is in turn used by the content providers to 'tailor' the information they provide to individuals. The supremely successful business models of first Google,[9] and then Facebook[10] and the other social networking sites began this process. Google's use of search terms to target advertising demonstrated the potential that access to personal information can provide. As the Google business model developed, and other businesses emulated their success, this tailoring has expanded to individualise not just advertising but the content of websites and the

[9] www.google.com

[10] Facebook is worth billions of dollars: Goldman Sachs' investment in Facebook in early 2011 gave it an estimated value of $50 billion. See example http://www.guardian.co.uk/technology/2011/jan/03/facebook-value-50bn-goldman-sachs-investment

suggestions given (and often the options provided) as to where to go next. For search engines, this means that not only the 'sponsored links' and advertisements that appear around search results can be tailored to the searcher, but that the results the searcher gets when they search for a particular term could be different, or in a different order, than those that another person might get if they search for precisely the same term. Given that this is the way that most people navigate the internet, this has a huge impact on what sites people become aware of and as a result actually 'choose' to visit.

This has many implications, but the most direct is that it 'fractures' the web, making it potentially different for each and every individual user – and different in ways that are controlled not by the user but by the content providers. This fracturing is very significant – and though it may appear just to be a side effect of the symbiotic collection of data, it has a direct impact on autonomy. Moreover, it is also a clear demonstration that the Symbiotic Web is not simply about the gathering of data but about its use. Indeed, it must be remembered that the data are gathered in order that they be used, and that the fracturing is one of the most important results of this use.

This shift of control from the user to the content provider is one of the things that distinguishes the Symbiotic Web from most of the ideas presented as Web 3.0. There is a good deal of confusion as to what is meant by Web 3.0 – and a great deal of uncertainty about what the future of the World Wide Web is likely to bring. Nova Spivack, for example, one of the key commentators in this field, now defines it simply as the 'third decade of the web', meaning 2010-2020.[11] From most perspectives, though, the fundamental change from Web 2.0 to Web 3.0 has been that it is expected to put more power into the hands of the individual, allowing the individual to find what he or she wants or needs using 'intelligent agents' to scour the internet – building on Berners-

[11] See his blog on http://novaspivack.typepad.com/nova_spivacks_weblog/2007/10/web-30----the-a.html , which he titles 'Web 3.0: The Best Official Definition Imaginable'.

Lee's concept of the 'semantic web'.[12] As this thesis shows, the future of the web seems both less unclear and less 'liberating' than these concepts suggest. Despite the appearance of the individual taking more control, the reality could be the opposite, with as we have earlier seen individuals having more of their choices made for them, and control taken out of their hands.

2.3 The emergence of the Symbiotic Web

The Symbiotic Web is already taking shape, using existing technologies such as cookies rather than requiring the development of new, intelligent software that may be years or even decades away from practical existence. Moreover, as will be discussed below, these moves towards the Symbiotic Web are driven by financial imperatives rather than by the somewhat vague technological speculation that appears to underlie the suggestions being made about the development of Web 3.0. That in itself seems likely, in the current climate, to make the future implied by the model of the Symbiotic Web a more probable outcome than the visions of even people as eminent as Berners-Lee, making it all the more important that the situation is better understood and where necessary appropriate interventions are made.

Even leaving aside the assumptions that many are making about the future development of the web, the emergence of the symbiosis described above was not an expected one – indeed it has come about directly against many of the predictions that both academics and lawmakers have been making in terms of the development of 'free' services. What many were predicting was the development and use of e-money, particularly for micro-payments. Miller, for example, writing about online payments in Edwards and Waelde's 'Law and the Internet: a Framework for Electronic Commerce' in 2000, begins his chapter with an sketch of how he saw a couple's online lives in the near

[12] See BERNERS-LEE, T. & FISCHETTI, M. 2000. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor,* New York, HarperCollins Publishers., particularly pp169-170. Spivack, in his blog referred to above, says that '*While Web 3.0 is not synonymous with the Semantic Web (there will be several other important technology shifts in that period), it will be largely characterized by semantics in general*'.

future – and one of the keys was a number of different, small scale payments for services on-line.[13]

Indeed, the European Union brought in its 'E-money Directive'[14] in 2000, anticipating the need for regulation as this kind of system developed and expanded. This has not yet materialised – the EU's evaluation of the E-money Directive, published in 2006,[15] concluded that there was 'widespread agreement among stakeholders that the e-money market has developed more slowly than expected.'[16] As the Symbiotic Web has come into action, an effective if hidden currency of personal information has meant that there has been little need for the kind of e-money system that is needed for these kinds of small payments.[17] That may change – and the 2000 Directive has been superseded by a 2009 Directive[18] which has already led to renewed micropayment speculation and the offering of new systems by some of the big players in the internet world.[19]The success or otherwise of these systems has yet to become clear.

[13] In Chapter Four of EDWARDS, L. & WAELDE, C. 2000. *Law and the Internet : a framework for electronic commerce,* Oxford, Hart., pp55-56
[14] E-Money Directive (2000/46/EC), available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0046:EN:HTML
[15] Available online at http://ec.europa.eu/internal_market/bank/docs/e-money/evaluation_en.pdf
[16] E-money evaluation report, p2.
[17] On the other hand, Rupert Murdoch's has suggested that the current "free delivery" system is a "malfunctioning model" and that he intends to charge for access to the Sunday Times online. See http://www.guardian.co.uk/media/2009/may/07/rupert-murdoch-charging-websites and http://www.guardian.co.uk/media/2009/jun/03/sunday-times-website . This, however, seems to be very much the exception: similar comments about charging for websites have been made in the past without result. The rapid decline of FriendsReunited, which continued to charge despite the growth of rival social networking sites like Facebook, suggests that, for mainstream consumers at least, the 'free delivery' model is unlikely to be easily overcome.
[18] Directive 2009/110/EC, which came into force in April 2011is available online at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:267:0007:0017:EN:PDF
[19] Systems such as Google One Pass (http://www.google.com/landing/onepass/) and Apple's subscription system (see http://www.apple.com/pr/library/2011/02/15Apple-Launches-Subscriptions-on-the-App-Store.html)

3 The make-up of the benign symbiosis

To get a better understanding of the Symbiotic Web, it is important to understand how it is made up – the nature of the services upon which individuals have become dependent, and the kinds of data gathered through them, and how it is used. Perhaps the most important element of the symbiosis is the search engine. The Google business model played a central role in the development of the Symbiotic Web, a role that is continuing and even growing, as the business models of Google and the other search engines become increasingly sophisticated. At a basic level their part in the symbiosis is clear – they offer an excellent service to internet users, and the offer is apparently for free. This makes search engines of fundamental importance, and a key part of the benign symbiosis. The services provided by search engines – and by Google in particular – are extremely good, and that though the Google business model is dependent on targeted advertising it would not work if Google search was not sufficiently good to command such a huge user base. In comparison to what had come before it, Google provided (and continues to provide) something that makes the internet far easier to use and far more attractive to the user.

Google provides these services effectively in exchange for gathering vast amounts of data – search data. In their search logs, search engines record not just the terms that are searched for and the links that are followed as a result of such a search, but the time and location at which the search was made and other details. These data can give a very detailed picture of the searcher's interests and habits, browsing style and so forth – and this can be extremely significant not only in profiling and targeting (which will be discussed in depth in Chapter Four) but also in working out how best to ensure that the searcher reads and follows particular links. What is more important, because search engines (and again, Google in particular) are used by such huge numbers of people, they are able to analyse patterns and behaviour on an unprecedented scale, and use it to hone their profiling. Google and others use this most directly for their targeted advertising – and as much of this

advertising is paid for on a 'pay per click' basis, it is in the search engine's interest to convince users to click on the advertisements. Google has been extremely adept at this – which is one of the reasons that it has grown to be one of the most successful corporations in the world. Search engines and search data will be looked at in more depth in Chapter Three.

The second group of key elements of the symbiosis to consider are the communications providers. Communications was one of the primary initial uses of the internet, and it remains one of the most important. It has become a key part of the benign symbiosis, as most of the communications services are provided to the user for free. This includes the large-scale web-based email services such as Google-mail, Microsoft's Hotmail, Yahoo Mail and most instant messaging systems and many chatroom services – services such as MSN Messenger, AIM,[20] GoogleTalk,[21] Yahoo Messenger,[22] Blackberry Messenger (BBM),[23] as well as internet telephony services like Skype.[24] The forms and quality of these services have been improving all the time, incorporating things like video conferencing which in the past was an extremely expensive, premium service, and yet they are still in general provided for free – a prime example of the essentially benign nature of the symbiosis.

The exchange that is taking place is that when people communicate over the internet, significant amounts of data are gathered and held about that communication – precisely what is recorded and held can vary significantly. Email providers keep records of those that a user sends and receives messages to and from; internet telephony providers keep records of a user's calls and so forth. As well as this 'traffic data' some providers will keep records of the contents of the actual communications themselves. Google and Yahoo mail are perhaps the best examples of this – the contents of an email can be used to generate targeted advertising in the same way that search

[20] www.aim.com
[21] http://www.google.com/talk/
[22] http://messenger.yahoo.com/
[23] http://us.blackberry.com/apps-software/blackberrymessenger/
[24] http://www.skype.com

engines do with searches. If an email contains particular keywords, the advertisements presented around that email will be tailored accordingly – if a user is writing about planning a holiday, for example, they might find advertisements for online travel agencies.

Social Networking services form the third key – and rapidly growing – part of the symbiosis. They provide a carefully built package of communications tools (including email, instant messaging and so forth), networking tools, games and other forms of entertainment, in a user-friendly form. The services they provide would in the past have only been available in highly expensive 'group-ware' that was effectively only accessible to big business. It is now available to anyone, and for free – and is a prime example of the essentially benign nature of the Symbiotic Web.

Services like Facebook, MySpace[25] and Bebo[26] are considered central to the 'Web 2.0 phenomena'. When their nature is looked at in detail it might be more appropriate to call them 'Web 2.5 applications', for the data-gathering side of their business is in many ways just as significant as the ostensible 'social networking' function – particularly for Facebook, the most popular of the systems with around 750 million users worldwide.[27] These numbers in themselves are enough to make social networking sites worthy of serious consideration. Social networking sites are 'free' to the user, yet worth billions of dollars to the owners through their ability to advertise and through the accumulated value of the data that their users supply [28]– which makes them perfect examples of the kind of symbiosis that characterises the Symbiotic Web.

---

[25] www.myspace.com
[26] www.bebo.com
[27] http://www.facebook.com/press/info.php?statistics
[28] As noted above, in January 2011, an investment by Goldman Sachs estimated the overall value of Facebook to be around $50 billion. See http://www.guardian.co.uk/technology/2011/jan/03/facebook-value-50bn-goldman-sachs-investment. A proposed IPO might double this to $100 billion. See http://www.guardian.co.uk/technology/pda/2011/jun/14/facebook-ipo-shares-likes-100bn

What social networking sites do, effectively, is ask their users to profile themselves. Users put in biographical data, educational data, information about their careers and their tastes in everything from music and food to religion, politics and relationships. In Facebook, for example, some of the most common 'applications' are questionnaires and quizzes, all seemingly for amusement, but in reality allowing Facebook and its advertisers to put together more and more detailed information about the user. Further to that, Facebook knows who his or her friends are, so can link this data to such friends, giving another dimension to the possibilities – the simplest examples include telling the user what their friends' favourite books and movies are, or informing the user that one of their friends has just started playing a particular game online. The profiles generated from all these forms of 'social data' are currently used primarily for targeted advertising – and also to help expand the service, providing more scope for further advertising, more users, and ultimately to make the company itself more valuable, at least in part because of the value of the enormous amount of data that it owns.

Internet Service Providers ('ISPs') make up the fourth category of key elements of the Symbiotic Web. Though they do not, in the UK at least, generally provide their services for free, prices for internet access have dropped dramatically, and sometimes internet access is 'bundled' with other services in a way that can be presented as free.[29] It may be that they are also being used as 'loss leaders' – perhaps in part because providers realise that the potential benefits from the data that may be gathered outweigh the relatively small costs involved in providing the service. They provide this service, with more bandwidth and fewer problems than before, and at a vastly reduced cost – and though the general improvements in technology over the last few years and the increasingly competitive market for internet service provision have been substantially responsible for these

[29] Examples include Sky TV bundling free broadband access with their basic satellite TV access (see http://www.selectdigital.co.uk/sky/sky-broadband/?tracker=7140e&gclid=CKic4Puk3pgCFUIw3godIVWBeA), and broadband bundled with mobile phone services from Orange (http://www.orange.co.uk/time/broadbandstarter/?cd_source=Aurora&cid=189&sid=72&pid=0&mid=131&acid=TENGU-1046758)

improvements, the beneficial symbiosis may well have also played a role in this service improvement.

The most significant data type gathered by ISPs is 'clickstream data' – the record of the clicks made when browsing the web. Just as for search data, clickstream data is not simply a record of what clicks are made but when, where from, and so forth. The potential economic benefits derivable from clickstream data have yet to be exploited as fully as Google and others have exploited search data, but the potential is clear – business models like Phorm, which is discussed below and in depth in Chapter Four, are just the starting point. It should be noted, too, that many other kinds of data can be (and often are) collected whilst clickstream data are gathered – information such as the type of computer being used, the kind of browsing software being used, the location of the user and so forth.

Commercial websites such as Amazon and eBay are some of the most successful and attractive sites on the internet – and make up the fifth key element of the Symbiotic Web. Though they do not provide their services for free, they do provide them at discounts to the prices that would be paid if their products were acquired in the offline world, and therefore provide a level of convenience that would not previously have been possible.

The direct shopping sites like Amazon gather data of two distinct kinds: transaction data relating to goods and services that have been bought or bid for, and 'interest' data relating to goods and services that have been looked at or researched on their sites. Both are useful in determining possible future sales – and the latter can include detailed 'clickstream data' (which is discussed in more depth below) including details like the timing between clicks and so forth, which can be used for profiling and to predict behaviour. Van den Poel and Buckinx, for example, found that detailed clickstream data was the best indicator of future online-purchasing behaviour[30]. In particular,

[30] VAN DEN POEL, D. & BUCKINX, W. 2005. Predicting online-purchasing behaviour. *European Journal of Operational Research,* 166**,** 557-575.

they found that it was a significantly better indicator than the actual purchases made – the information that users might reasonably believe was used by online stores[31].

These two types of data have very different characteristics when considered from the perspective of privacy and autonomy – it is difficult to imagine that shoppers really understand that their browsing is being monitored and recorded as closely as their actual transactions. It is reasonable to expect the real transactions to be recorded and used for marketing – but quite possibly unreasonable for the rest of the browsing to be taken into account in the same way. Whether it is reasonable for either of these types of data to be sold on to third parties for aggregation or other commercial use (as in the Beacon system discussed below, and elsewhere) is another question entirely, one that relates very closely to the complex issue of consent, an important idea the consideration of which is a concern running through this thesis.

To summarise: the five most significant individual elements of the Symbiotic Web have been described above: the search engines, the communications providers, the social networking services, the commercial websites, and the ISPs. However, to a certain extent the symbiosis covers the majority of the web. The pure 'information providers' generally supply their information for free. There are all kinds of other free services available, from 'geographical' services like Google Maps,[32] Google Earth,[33] Google Street View[34] and the various street finder systems[35] to the recreational services like YouTube[36] and its equivalents. New kinds of services are evolving all the time – and a large proportion of them are free, built on business models using data gathering and targeted advertising.

---

[31] Amazon, for example, explains its recommendations by saying 'recommended because you purchased…'
[32] http://maps.google.com/, http://maps.google.co.uk/ etc.
[33] http://earth.google.com/, http://earth.google.co.uk/ etc
[34] http://maps.google.com/help/maps/streetview/
[35] There are a number of such services for many locations, including www.streetfinder.co.uk/ and www.streetmap.co.uk/
[36] www.youtube.com

The data gathered by these services vary enormously. First of all, controllers of websites can gather the clickstream data that relate to their own sites – when people arrive at their site, they can gather information such as where they have come to the site from and all the clicks once they arrive, including where they go to next. Then there are the more specific data related to the service provided – for geographical sites, the places people are looking at; for media sites like YouTube the tastes people have in music and video; for sports sites what sports and teams they follow; and so forth. As will be discussed in Chapter Four, all of this data can help in building up profiles, in targeting advertisements and so forth. More types of data and more methods of extraction are being developed all the time.

4 The risks of a malign symbiosis

As has been discussed, the Symbiotic Web is currently an essentially positive thing, providing benefits for individuals, for business, and potentially for society as a whole. Nevertheless, there are significant risks associated with the symbiosis. A new understanding of the potential value of personal information has developed and with it whole new markets – not just different ways to use this information but markets in the buying, selling and aggregation of these data. The competitive drives that accompany these developments will invariably impact on both privacy and autonomy.

The starting point is the understanding that personal information has a commercial value. This has led organisations to gather more and more data, not just for specific current or planned uses, but speculatively, based on an assumption that new uses and new values will be found for these data. And not only are more data gathered, but there is pressure to find new and different ways to gather the data – some of these are detailed elsewhere in this thesis. As these data are gathered, the organisations are looking for more and more ways to use the information they have – if a business has an asset, it will want to get as much commercial value from it as it can. The more competitive the market, the more attempts there will be to squeeze the

maximum value out of the data. New businesses are developing for aggregation of data and profile generation – not only to make money from the existence of the new data, but also to find even more effective ways of using such data for other businesses.[37]

4.1 Beacon and Phorm

The Facebook 'Beacon' affair and the 'Phorm' business concept are just two examples of these phenomena – demonstrating some of the possibilities that are being explored by businesses to exploit not only the nature of the data that is being gathered but showing also the potential that a computer network can provide for such exploitation. Both are examined in Chapter Four: a case study of Phorm forms the centrepiece of that chapter.

Beacon was an advertising system developed by Facebook, a method by which Facebook exploits the value of its social data. Beacon was intended to allow Facebook to share personal data with a number of online retail 'partners' – receiving the data gathered by those retailers in exchange. The Beacon system was originally intended to be to all intents and purposes a covert system, and an 'opt out' system – all Facebook users were intended to be included unless they found out about it and specifically asked not to be included. That in itself raises a lot of issues – the recurrent issue of consent in particular, which will be examined in depth in Chapter Four – but it also demonstrates how these kinds of commercial alliances can be formed. Members of the alliance would quite naturally wish to be mutually supportive – and hence do their best to support each other to the detriment or exclusion of competitors. This is just normal business practice – but people tend to see social networking sites as 'neutral', and even use them as a route to navigate the web. If similar systems were extended onto search engines it is easy to see how conflicts of interest might result in unfair or misleading search results – and consequent manipulation of how people

[37] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray. for details not only of the data aggregators but some of the more imaginative ways in which this data is beginning to be used.

navigate the web. The ongoing EC investigation into the functioning of Google, which originated from cases such as the 'Foundem' case, is exactly on this point, Foundem's complaint suggesting that Google manipulates search results to favour sites which particularly benefit it.[38] The issue with Facebook may be less intrusive or significant – but an indication of how things might move in the future.

The reality behind Beacon was discovered before it came into action, and the privacy issues surrounding it raised such a furore that Facebook was forced to change it significantly before it started, making it an opt-in rather than opt-out system, amongst other things. This is in itself is revealing, in two ways. Firstly, it demonstrates why companies might keep the real reasons for their data policies and practices effectively secret from most of their users – for when users find out what is going on, they often object, and object strongly. Secondly, it suggests some of the possible ways to change things – firstly by raising awareness of practices so there are more objections; secondly by making it harder for companies to keep such practices secret; and thirdly by making it harder for companies to use practices which do not require real express, informed consent, of an 'opt in' rather than 'opt out' form[39]. As shall be detailed in Chapter Four, Beacon was eventually abandoned in September 2009, in part at least as a result of the exposure of its privacy-intrusive nature.

Phorm's 'Webwise' was another example of how these kinds of alliances can form, and the impact they can have. As is discussed in depth in Chapter Four, Phorm's Webwise system would have allowed some of the UK's biggest ISPs to analyse individuals' browsing behaviour. This would allow potential advertisers to target users, by 'intercepting' the clicks made as a user surfs

[38] Google is being investigated under competition law after a complaint by UK price comparison site *Foundem*, French legal search engine *ejustice.fr*, and Microsoft's *Ciao*. See for example http://www.telegraph.co.uk/technology/google/7301299/Google-under-investigation-for-alleged-breach-of-EU-competition-rules.html

[39] The Facebook Beacon story is discussed Chapter 4 – and to see a summary of the reasons for the big changes to the system from Facebook's own perspective, see http://blog.facebook.com/blog.php?post=7584397130

the web, analysing them and building up a database of users' entire internet activity.[40] Effectively, Phorm intended to harness the value of clickstream data. Phorm raised a plethora of legal, ethical and commercial issues, as it monitors a user's entire online activities, and as such was challenged as a possible breach of wiretapping regulations, as a breach of data protection legislation and as another step towards a surveillance society. It was also being seen by some as an interference with other websites' commercial interests – it could, for example, gather all the search data entered into the Google search page before Google themselves gather it. The issues raised by Phorm are complex and concerning in many ways – and are examined in depth in Chapter Four. Phorm's importance lies not just in its proposed functions but in what it implies about where the symbiotic nature of the web might cause it to go, and how, as noted above, the competitive drives that underpin the Symbiotic Web will manifest themselves in more and more imaginative and potentially risky ways of using – and exploiting – the data that are being gathered.

4.2 Tailoring and Balkanisation

Another key set of risks arise through the process of 'tailoring' of web pages for individuals, which as discussed above is one of the fundamental features of the Symbiotic Web. As businesses learn more about their customers – and are able to derive more information through new profiling and data aggregation businesses – they are able to 'tailor' services even more. This tailoring can include such potentially pernicious practices as price or service discrimination. If a business can learn enough about a customer to know how much they might be willing to pay for something – which becomes more and more possible as they gather more data about that customer – then they can set prices (and display those prices on their web pages) individually for that customer, prices that might be either higher or lower than those offered to others.

[40] See http://www.phorm.com/ and for a look at the negative side of Phorm, see http://www.badphorm.co.uk/page.php?2. Phorm is examined in detail in Chapter 4.

With the development of the Symbiotic Web, companies can potentially learn much more about their customers than ever before – and not just the kind of information that the customer wants them to know. Information such as social class, salaries earned, home ownership, purchasing history from other businesses and so forth can be available through data aggregation,[41] or via commercial alliances such as those already forming through Beacon and Phorm. Profiling techniques, together with the increase in available information, can allow companies to predict with more and more accuracy not just what their customers might be persuaded to buy, but how much they might be willing to pay for it – and this in itself has its problem. The idea of 'price discrimination' might just seem like good business practice – offering better prices to regular customers and so forth – but it has a downside as well. Whilst lowering prices for regular customers to reward their loyalty is both common and appropriate it is also possible that tailoring might result in higher prices for others – or even for those who are viewed as more likely to be willing to pay more, would not be a good thing, at least not from the customer's perspective. All of this becomes possible in the Symbiotic Web – and as the competition between businesses grows, and as the availability of data and the technical and technological capabilities for processing it become better and more available, the drives to use it become stronger.

As noted above, tailoring can apply not only to content, but to links provided – most pertinently in search results[42] – which ultimately results in the personalisation of the web experience and could be seen as a 'fracturing' of

[41] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray., particularly pp33-34 for an examination of the predictive use of data. As Ayres puts it, *'…data mining can let business emulate a kind of aggregate omniscience. Indeed, because of Super Crunching, firms sometimes may be able to make more accurate predictions about how you'll behave than you could ever make yourself*.'

[42] From December 2009, Google has 'personalised' all search results, unless the searcher actively opts out (see http://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html) .See the discussion by search engine blogger Danny Sullivan http://searchengineland.com/google-now-personalizes-everyones-search-results-31195

the web.[43] This brings with it a further set of risks. The internet that a user is 'exposed' to is becoming one that is controlled for them in ways that ensure that a user only sees things that people think that they will like – they know the users' tastes, who their friends are, what kind of work they do, the kind of music they like and movies they watch, and present to them only those things that they think these individual users will be interested in. Personalised news pages will cover the topics the news providers 'know' the user cares about, possibly only from the news sources that they 'know' the user trusts. The products and services offered for the user to buy will be only those that match the profile of the user that sellers have built up– from the point of view of the seller this makes perfect sense, since these are the products the user is most likely to buy. The events, TV shows and movies that the user is told about are similarly chosen to suit what is known about them – again, something that makes perfect sense to the providers. When the user searches for something, the search results, too, are chosen with what the search engine knows about the user, what the user likes and what the user is interested in.

The result may be something instantly attractive to the user – and something comfortable and unthreatening. One somewhat extreme way to look at it would be to consider a social networking site like Facebook – where a member really does only see their friends (and only their friends see them) and all the applications suggested to the member and advertisements presented to them are tailored to what Facebook knows about them. The Symbiotic Web could end up producing an internet that functions like one, big social networking site, where people only see things that the providers think they will like, and never see things which providers don't think they'll like – and indeed are specifically excluded from places where controllers do not like their profile. Whilst this might be attractive in one way – we'll like almost everything we see, and never know what else we are missing – it is vastly less positive and stimulating than the current version of the internet. It

[43] Some aspects of this tailoring, and how it is starting to emerge in reality on the internet, are discussed in PARISER, E. 2011. *The filter bubble : what the Internet is hiding from you,* London, Viking. This book was published too late for full consideration in this thesis.

is a kind of 'sanitised' internet, where the chances of coming across something surprising and really new are limited.

One particularly pernicious version of this kind of thing is a phenomenon that can be described as 'back-door Balkanisation', to extend Sunstein's metaphor from Republic.com.[44] Sunstein discussed how the internet could have a tendency to polarize opinion and create niches with narrow and potentially extreme political views or interests. Whilst Sunstein writes largely about a phenomenon that takes place through the choices made by the individuals, what could potentially happen through the Symbiotic Web would be without the knowledge or understanding of the user, let alone through any kind of conscious or even subconscious choice – Balkanisation through the back door. Effectively, if through their profile a user is deemed to hold a particular political, religious or ideological stance, this kind of system could drive that user into a more extreme version of that stance, with dangers not only for the individual but also for society as a whole. The fact that it happens automatically would make it even more pernicious, and potentially even more dangerous than the phenomenon described by Sunstein. Sunstein's theories have been much criticised[45] – but as a significant part of that criticism rests on the rights of the individual to make his or her own choices, the back-door Balkanisation that accompanies the Symbiotic Web is something quite different: at very least it needs to be considered seriously.

Taking this a step further, there is the potential for individual service providers and web providers to make conscious, pernicious choices in particular ways – checking profiles of users before deciding what kind of information to provide. Nightmare visions such as 'whites-only websites', which check visitors' profiles to determine whether they should be allowed to see certain content, will be both a technical and practical possibility in the

[44] See SUNSTEIN, C. R. 2007. *Republic.com 2.0,* Princeton, Princeton University Press.
[45] Perhaps his strongest critic is Eugene Volokh, of the Volokh Conspiracy (http://www.volokh.com/), but there has also been active criticism in print, such as that by Dan Hunter in HUNTER, D. 2001. Philippic.com. *California Law Review,* 90**,** 70.

near future. It should be remember that this kind of profiling can be 'inclusive' – only allowing access to particular content for an 'approved' kind of person – or 'exclusive' – allowing access for everyone except a 'banned' type. The potential for misuse of this kind of profiling is significant.

4.3 Risks associated with particular data types

There are also risks associated with particular data types being gathered. With communications data, for example, where data such as the content of communications are held, even if the primary use is simply for targeting advertising and commercial profile building, there is the potential for misuse. Wherever and however data are held such material can be vulnerable, so it is not just what the communications provider might do with the data that is of concern, but what more malign use others might make of it. Security and privacy of communications are key human rights, particularly in times and places of political oppression. The well-publicised examples of dissidents being imprisoned in China as a result of information provided by Yahoo as to their communications are just one of the ways that this sort of thing can cause problems.[46] There are similar possibilities with other kinds of data – most notably the data gathered and stored by social networking sites.

It should also be remembered that whilst there are particular issues concerning each of the data types discussed above, the overall effect is greater than the sum of the individual parts. Google, for example, combines the information gathered by search with that gathered on gmail, through Google Maps, Google StreetView, Google Earth and so forth. The opportunities for profiling, for aggregation and for other forms of research multiply as more data become available.[47]

Much of the concern expressed so far has related to the risk of the symbiosis becoming unbalanced, as commercial forces drive organisations to find more

[46] See e.g http://en.rsf.org/china-yahoo-settles-lawsuit-by-families-14-11-2007,24240.html
[47] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray.

ways to gather data, and more ways to use that data, and try harder to control the individuals about whom the data have been gathered. The range of ways in which this control can be exercised has not yet been explored beyond a surface level – all that has been seen to date has been regular advertising techniques and persuasion and through the presentation of links and so forth. Powerful as these methods are, there are deeper possibilities – and as Lessig has suggested, code-writers wield enormous power in the internet, and through the design of the architecture, new ways to monitor and control individuals can and will be developed.[48]

4.4 The burgeoning market in data

Perhaps the most significant potentially malign result of the Symbiotic Web, however, is simply the burgeoning market in data[49] – one about which users are largely ignorant. Businesses are becoming acutely aware of the value of gathering data, but at the same time evidence suggests that when customers are aware that their data are being gathered for such purposes, they don't like it – the reactions to the Facebook Beacon affair and the controversy over the emergence of Phorm are two pieces of evidence to support this. Though in some cases it is made clear that data are being gathered, it is often far from clear and, even when it is clear, the true uses to which the data are being put are rarely revealed. Individuals are generally kept in the dark – and this in itself leads to a loss of autonomy.

The very existence of this massive quantity of data represents a significant risk – digital information, wherever it is and however it is stored, is vulnerable, whether from hacking, inadvertent or inappropriate selling or giving away of data, hardware and software failure, hardware theft or loss, administrative or security failures and so forth. Once the data have been 'lost', the potential for criminal misuse is huge – already crimes like identity

[48] See LESSIG, L. 2006. *Code: Version 2.0,* New York, Basic Books.
[49] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray., particularly pp134-138, for an examination of the growing market in data

theft and other forms of financial fraud are a significant problem.[50] As new kinds of information become available – particularly profiling and equivalent information – the potential for better-targeted and more pernicious identity-related crimes increases dramatically. Data vulnerability and the potential consequences of such vulnerability will be examined in Chapter Five. Furthermore, as noted earlier, the existence of the data makes it tempting for those who have access to it to find new uses for it – uses that are not necessarily in character or proportional to the reasons for which the data were gathered in the first place. This 'function creep' has been particularly evident in recent years in relation to data gathered for anti-terrorism purposes – notable examples include the use of the Regulation of Investigatory Powers Act (RIPA) 2000,[51] which was presented as a means to tackle terrorism and other serious crimes, to deal with dog fouling[52] and to spy on a couple to determine whether they were using a false address to get their child into a local school.[53] Function creep may come into play for commercial reasons and in commercial contexts even more often that it does for security or law-enforcement purposes, and is a risk whenever data is held, so the more data is being held, the greater the risk.

5 Regulating the Symbiotic web

If these risks are not addressed, many of the best features of the existing internet may risk being lost. The idea of a common knowledge base, a place where people can speak and act freely – a system that can support dissidents and the oppressed, encourage community and global interaction in a positive way: all these things are under threat, as well as the privacy and autonomy of individuals. The internet can – and to a large extent currently does – represent a great opportunity for users to grow as people, expanding their

[50] In the UK, for example, the Attorney General estimated in 2010 that the annual cost of identity theft was approximately £2.7 billion. See http://www.attorneygeneral.gov.uk/nfa/whatarewesaying/newsrelease/pages/identity-fraud-costs-27billion.aspx .
[51] Available online at http://www.legislation.gov.uk/ukpga/2000/23/contents
[52] See for example http://www.telegraph.co.uk/news/uknews/1584808/Council-spy-cases-hit-1000-a-month.html
[53] See for example http://news.bbc.co.uk/1/hi/england/dorset/7341179.stm

horizons, their knowledge, and their breadth of experience. Any throttling of that opportunity is in itself a restriction of autonomy. For all these reasons, it is important that everything is done to ensure that the positive benefits and the essentially benign nature of the web symbiosis is maintained, and such risks as there are are addressed appropriately. A key part of meeting this challenge is regulation – and this leads to consideration of the kind of role governments and other regulators can play in supporting the positive nature of the symbiosis. This will form a significant part of the thesis as a whole, and in particular in Chapters Six and Seven.

Before this can be addressed, the role that governments are already playing in the evolution of the Symbiotic Web and the complex combination of issues that affect their ability to provide appropriate regulation must be understood. The development of the Symbiotic Web has been essentially driven by commercial forces, and primarily for commercial interests, but governments have already had an important role in shaping it, and that role could become more significant in the future. The symbiosis, in its benign form, is in the interests of government. It can help provide happy and satisfied citizens, thriving businesses and technological innovation – and, almost as a side product, substantial amounts of data that can potentially be extremely useful for governments. In broad-brush terms government likes to have more information – not for sinister reasons, but because information can help it to do its job better. The more information it has about its citizens, the more accurately and appropriately government can design and implement policies to support them and to satisfy their needs. This general need covers almost every aspect of government, from housing and employment policies to taxation and health – not just the more contentious areas such as crime prevention and security. As a consequence, it is in the general interest of most governments to firstly support the symbiosis and secondly do whatever is necessary to ensure that it remains benign – for many of the malign possibilities for the symbiosis, as outlined below, are not at all in the interests of government.

5.1 Data protection and data retention law

The most significant law that already exists in this area is data protection law – which will be discussed in detail in Chapter Three. Data protection, as a concept, arose in Europe and was enacted through the Data Protection Directive[54]. It is designed to protect the rights of individuals with respect to the 'processing of personal data',[55] and sets out eight 'data protection principles': that data must be processed fairly and lawfully; that the purpose for data gathering and processing must be specified and not extended; that only relevant data may be processed; that such data be kept accurate and up to date; that the data be kept only for the period required for the specified purpose for which the date were gathered; that data subjects' rights be maintained; that data be sufficiently technologically and organisationally protected against misuse, damage or loss; and that such data not be transferred out of the European Union unless adequate protection in the country to which the data are transferred can be assured. These principles are strong, and could potentially form the basis of what would be an excellent protection of individual's privacy. Indeed, they have already played a significant part in moderating the actions of corporations and governments in terms of the gathering and processing of personal data, and hence in shaping the development of the Symbiotic Web. Data protection law, however, suffers from many problems, including its inconsistent implementation through national law and its often-inadequate enforcement. Governments have also experienced difficulty in maintaining the balance between the interests of privacy, economics and security.

Indeed, it is arising from security interests that the second significant legal moves in the field of data have arisen – data retention laws, which developed as part of counter-terrorism. Data retention as a concept is more widespread than data protection – indeed, data retention is embraced enthusiastically by

[54] Directive 95/46/EC
[55] Directive 95/46/EC, Article 1.1

governments all over the globe, and in particular in the United States.[56] In Europe it has been enacted through the Data Retention Directive,[57] which essentially requires ISPs, email service providers, internet telephony providers and the equivalent, to retain sufficient data to be able to identify where people have been on the internet and with whom they have communicated, and to make that data available to the authorities when an appropriately authorised request is made. In data retention, as in other areas such as combating illegal online activities such as illegal downloading of music and child pornography, the approach appears to be to get commercial enterprises to do much of the work on behalf of the authorities – at the very least, to gather and hold the relevant data, but sometimes to go beyond that and detect and report suspicious activities, or even actually police them, though many of these plans end up coming to nothing.[58] The authorities are trying to take advantage of the new kinds and volumes of data being gathered for commercial use, and using it for their own, very different purposes. And as further new kinds of data appear, the authorities are continually assessing whether they can use that data – with the expansion of social networking sites, for example, the UK government is considering expanding the terms of data retention to include social networking data as well as the more conventional communications data that it originally covered.[59]

Data retention is playing its part in shaping the Symbiotic Web, though in a very different way to data protection. Indeed, there is a tension between data protection and data retention – one effectively asks for less data to be held, and for less time, while the other asks for more data to be held, and for longer. This tension, reflecting the common tensions between privacy rights and counter-terrorism, is examined in detail in Chapter Three. Google, for

[56] Data retention is one of the subjects of Chapter 3, which includes a brief discussion on worldwide implementation
[57] Directive 2006/24/EC
[58] Examples include the UK Government's plan to get ISPs to detect and ban users who illegally download music and video files. This plan, like many other schemes, had to be abandoned as getting the ISPs to do all the work turned out to be fraught with legal and technical complications. See for example http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/music/article5586761.ece, from January 2009.
[59] See for example http://news.zdnet.co.uk/security/0,1000000189,39629479,00.htm

example, has been using the principles of data retention as one of its arguments against having to comply fully with data protection law, and indeed has been hinting to the authorities that they should consider extending data retention law to cover search data – which would effectively allow Google to avoid having to comply with data protection requirements, and hence hold data for longer periods and in more extensive forms than data protection would otherwise allow. Though this argument was roundly rebuffed, data retention as a concept is still in play, and indeed its extension and more rigorous implementation is on the agenda. Just as for the Data Protection Directive, there are problems and inconsistencies in implementation and enforcement – and to add to these, data retention has also been the subject of legal challenges, both to the Directive itself and to its implementation in a number of member states.[60]

Much of the problems with both data protection and data retention arise because governments (and other regulators) are not simple, monolithic organisations, and also because where the internet is concerned, the situation is complex and multi-layered. In the UK for example, though UK law applies and the UK government has jurisdiction, much of the law is to an extent driven by Europe through EU Directives – most directly the Data Protection and Data Retention directives discussed above but also the directives on ePrivacy,[61] eCommerce,[62] eMoney[63] and others. There are good reasons why the internet should be regulated on as global a level as possible, not least the global nature of the internet itself – but it still raises issues and causes conflicts both of interest and of politics. Even more problematically, attempts to find global solutions are difficult since there are differences in approach to many issues between Europe and the United States, and as so much of the internet, particularly from a commercial perspective, effectively arises from the United States, the US approach to regulation has a major

[60] These challenges are examined in Chapter 3.
[61] Directive 2002/58/EC
[62] Directive 2000/31/EC
[63] Initially Directive 2000/46/EC, repealed and replaced by Directive 2009/110/EC.

impact. One crucial difference is the effective absence of data protection legislation in the US.

Moreover, there are conflicts even within individual countries, as the three way dynamic of competing interests of autonomy, security and economics discussed in Chapter One are represented not only by the 'external' relationships between individuals, government and business but also by the 'internal' relationships between government agencies that represent those interests. In the UK, for example, in terms of privacy and autonomy this means the Information Commissioner's Office,[64] whose task amongst other things is to ensure that the UK complies with data protection law. It has found itself in conflict with different arms of the government on numerous occasions – in recent years this has included voicing regular concerns over the government's ID cards plans,[65] over proposals for DNA databases,[66] and over aspects of the government's implementation of the Data Retention Directive.[67] The extent to which these concerns have had an effect is limited, as the ICO is a body with a limited budget and whose enforcement powers still remain limited, though they have been strengthened in the last few years.[68]

5.2 The balancing act of regulation

In essence, governments have a balancing act to perform in relation to personal data – they must protect the privacy and autonomy of individuals, promote and support their businesses in achieving economic success, and at the same time ensure security for their citizens, Maintaining that balance is

[64] See http://www.ico.gov.uk/ - the role of the ICO will be discussed in depth in later chapters
[65] See http://www.ico.gov.uk/news/current_topics/identity_cards.aspx which describes the current position, and notes the numerous times since 2002 that the ICO has voiced its concerns.
[66] See http://www.ico.gov.uk/upload/documents/pressreleases/2007/ico_statement_dna_database.pdf
[67] See http://www.ico.gov.uk/upload/documents/pressreleases/2008/ico_statement_comms_data_bill.pdf and the in depth discussion in Chapter 3
[68] A more detailed examination of the ICO is contained in Chapter 5

difficult – and unless the issues involved are appreciated and understood it is easy for privacy and autonomy to be squeezed by the competing economic and security interests. Indeed, the Data Protection Directive declares in its preamble that 'the establishment and functioning of an internal market' requires 'that personal data should be able to flow freely from one Member State to another' – in other words, that our economic success relies on the ability of personal data to flow freely.[69]

Governments also work with the data gatherers in many other ways – examples include the UK government considering turning to Tesco for data on the UK's migrant population.[70] The commercial data gatherers have encouraged this – presenting good examples of how their data can be used in positive ways, to assist both the governments and the people. A dramatic example of this is Google Flu Trends, which uses search data to analyse outbreaks of flu state-by-state in the US, and claims to be able to do so up to two weeks earlier than 'traditional systems'.[71]

In these cases, governments are effectively piggybacking onto the Symbiotic Web, using not only the data that are gathered but the ways that such data are processed and used, for their own purposes. This emphasises the reasons that governments are supportive of the symbiosis. When one considers how the profiling and targeting that is performed by commercial enterprises, and examined in detail in Chapter Four, is likely to be extremely similar to the profiling and targeting that government agencies might wish to do in their struggles against crime and terrorism, or indeed against disease, the reasons become even clearer. That, of itself, is something that suggests a need to be wary. When the interests of governments and businesses coincide so closely, it is crucial that the needs of individuals do not suffer.

---

[69] From paragraph (3) of the Preamble to The Data Retention Directive, Directive 95/46/EC. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

[70] See for example http://blogs.ft.com/westminster/2008/04/can-tesco-help-the-government-count-how-many-migrants-are-in-the-uk/

[71] See http://www.google.org/flutrends/

As these discussions of data protection and data retention have demonstrated, regulators have difficulty dealing even with the current uses of the web. Dealing with the new kinds of uses that have arisen as a result of the Symbiotic Web can put them really out of their depth – and so regulators will find it very difficult to control the potentially malign consequences of the symbiosis. Data protection law, which could in principle provide some help in this area, has proven far from effective – in terms of enforcement, in particular. Little is done to deal with the contentious issue of consent, which in terms of autonomy is perhaps the single most important issue. Maintaining the balance between individuals and businesses has proved very difficult.

The UK government's response to the Phorm business idea that forms the central case study for Chapter Four has demonstrated some of these problems, and illustrates the difficulties that arise as a result of the multi-layered legal structure necessarily involved here – and some of the problems around the implementation and enforcement of data protection. In response to questions about Phorm's compliance with data protection law, the Department for Business, Enterprise and Regulatory Reform (BERR) effectively said that there was no problem, and that Phorm complied fully with UK Data Protection Law[72]. BERR was in a position of having to satisfy the demands of business and appears to have taken a somewhat unbalanced position, weighing the interests of business above those of individuals. Certainly the European Commission believes so, and has taken legal action against the UK as a result, essentially saying that if Phorm's business practices comply with UK data protection law, then UK data protection law has not effectively translated the EC Data Protection Directive into domestic law. The European Commission does not have the same balancing act to perform as BERR, at least not in relation to a specifically UK-based business initiative, so is freer to look at the issue purely from the perspective of

[72] BERR did not make the letter in which they confirmed the legality of Phorm public, but told the press that 'After conducting its enquiries with Phorm the UK authorities consider that Phorm's products are capable of being operated in a lawful, appropriate and transparent fashion'. See for example http://news.bbc.co.uk/1/hi/technology/7619297.stm

individual privacy – but as will be detailed in Chapter Four, what impact the EC's action will have is far from clear.

6 Managing the Symbiosis

So what, if anything, can be done? How can regulators manage this symbiosis, and make sure that it evolves in positive rather than negative ways? What role would be appropriate for governments and other regulators to play, and how should they play it? What place is there for other forms of regulation or control – in Lessig's terms, for markets, for norms, and for code? Ultimately, these are the questions that this thesis is intending to answer – and the answer is given in terms of human rights, for it is the contention of the thesis that a right-based approach is not only the most appropriate but is likely to be the most effective one so far as tackling these issues is concerned.

One of the functions of human rights is to provide protection for individuals against more powerful forces – most directly those of governments, but increasingly also those of business. That, in this context, is particularly important – the forces behind the Symbiotic Web are powerful and potentially dangerous, and individuals need to be protected. The current methods, not in any real sense based on rights, do not seem to be providing an adequate solution – a rights-based solution could be far more effective. There are a number of possible ways forward, each of which suggests a different kind and level of involvement by governments and other regulators.

The first option would be to try to break the dependence – to make the use of personal information in this kind of way impossible through stronger, better-enforced laws. Current data protection laws should theoretically be a good start – in particular, the principles of data minimisation, of using data only for a set, lawful process and not allowing further processing, and the need for express, informed consent before data is gathered or processed, could potentially provide a great deal of protection . In reality they appear to fail to live up to their promise, whether through weakness of implementation or

through poorly resourced enforcement. If sufficiently strengthened and properly enforced, they could make a significant impact. This, however, could effectively mean the end of the Symbiotic Web, as many of the business methods that have driven its development would become effectively illegal. Not only might this mean giving up all the positive aspects of what is a benign symbiosis, but it would mean having to come into conflict with very powerful business interests, which would be difficult to say the least.

The second, and converse, approach would be to try to change the paradigm and 'give up' on privacy to a great extent. It may be an option to accept the trade in personal data, encourage personalisation, and deal with the consequences by penalising excessive or inappropriate use and encouraging understanding in the general public. Though it might appear a purely pragmatic solution, it might end up with something beneficial, as suggested by writers such as David Brin.[73] However as will be demonstrated particularly by the case studies in Chapters Four and Five, there are too many negative consequences for this to be a practical possibility.

The third approach would be to do very little, and allow markets and norms to redress the balance. There are some signs so far that this kind of approach might work – the increased newsworthiness of privacy breaches, discussed in Chapter Five, suggests that things might be changing. The intervention of lawmakers might be producing even better results. Google has reduced the time that it holds onto individualised server logs of search data from an unlimited time first to 18 months, then to 9 months, to a great extent because of the pressure exerted by the Article 29 Data Protection Working Party, the EC body responsible for oversight, advice and expert opinion concerning data protection,[74] as will be detailed in Chapter Three.

The fourth approach, and the one advocated by this thesis, is to weaken the dependence, to loosen the symbiosis and strengthen the rights of the

[73] Most notably in BRIN, D. 1998. *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?*, Reading, Mass., Addison-Wesley.
[74] See http://ec.europa.eu/justice/policies/privacy/workinggroup/index_en.htm

individuals – particularly in terms of consent and rights to be informed. This would alter the balance, but still allow the mutually beneficial symbiosis. It could be brought about through a combination of legal, technological and other measures. The Google/EU case study in Chapter Three suggests that this may be the best approach – effectively, the Article 29 Working Party used the pressure of law based on individual rights, while both Google and the Working Party attempted to get the public on their side, to use norms to deliver the result they wanted.

6.1 Symbiotic Regulation for the Symbiotic Web

The starting point is to have adequate, well-expressed, coherent, consistent and up-to-date rights. These rights, set out in detail in Chapter Six, are expressed in terms designed with the Symbiotic Web in mind, using Murray's regulatory model which he calls 'symbiotic regulation'[75] – it is not simply a coincidence of labels that both use the concept of symbiosis, though their origins are different, the term Symbiotic Web being one developed by this author alone. Symbiotic regulation starts with a recognition that the relationships in the web are complex, and that the best way to achieve regulatory results is to work with those existing relationships rather than seek to impose something authoritarian upon them.

The first step to achieving a regulatory result, as outlined by Murray, is to map out the relationships involved. This is precisely the purpose of the Symbiotic Web – to be a model of the relationships between individuals, businesses and others in the Web as it currently exists. Setting out that map is what has been started in this chapter, and will be continued in the substantive chapters that follow.

The next steps are to establishing the appropriate response, from a regulatory perspective. There are already some clues as to how this might

[75] See MURRAY, A. D. 2006. *The Regulation of Cyberspace: Control in the Online Environment,* Milton Park, Abingdon, UK ; New York, NY, Routledge-Cavendish.

work. A key part of the symbiotic regulation approach is to work with existing relationships, 'tweaking' them to produce results – and it is clear that existing relationships are, in some cases, already producing these results. The Article 29 Working Party's relationship with Google has existed for some time – it is the Working Party's 'tweaking' that seems to have made Google change policy. Similarly, the existing relationship between Facebook and its users was the key to halting the Beacon system in its tracks – and has had a number of other effects on Facebook, most recently in convincing it to reverse a change in policy about the deletion of data when users delete their profiles[76]. These examples also highlight another feature of symbiotic regulation – the development and support of an active, positive community of users, able to participate in the regulatory process rather than be merely passive subjects of it.

6.2 New Business Models

Just as the movement towards the Symbiotic Web was driven by a new business model – the Google/Facebook personal data/targeted advertising model – the movement to moderate it and keep it benign will in all likelihood require the same. If a new business model, not dependent on the gathering and use of personal data – or using it in less intrusive, less manipulative way – could be developed, that would surely lead us in a more positive direction. The Google vs EC case study in Chapter Three provides some evidence that this may already be happening. Google's relatively bloodless acceptance of a reduction of data retention periods might be because they are developing a different model for their business – but pressure from lawmakers, the computer and hacker communities, and most importantly from users, could make this happen faster. What is more, as Google and others become more aware of their place in the symbiosis, and of the way in which maintaining the benign nature of the symbiosis benefits them as well as the individuals,

[76] The policy change covered new terms and conditions, but the key terms seemed to be about the deletion (or rather non-deletion) of some data even when profiles were deleted. See for example http://news.bbc.co.uk/1/hi/technology/7896309.stm

they can become drivers towards positive solutions, rather than seeking to delay or block them.

One way in which this might already be happening is that, as revealed in the case study in Chapter Three, Google believes in setting global standards for their global business. For example, Google sets global deletion periods for their search logs, though outside Europe there appears no legal need for data deletion at all, as data protection rules do not apply. As noted above, global solutions make much more sense for the internet – and perhaps global businesses like Google can help drive governments into developing those global solutions – another example of symbiotic regulation in action.

This is where a rights-based approach can be particularly effective. Rights can work not only as practical ways to bring about change – the Google story is just one example – but as global standards to aspire to, and principles that can guide future actions. The existence of a right can help to establish in people's minds that the subject of the right is of importance – here, to establish the idea that our privacy and autonomy on the internet is important. If commercial actors, in particular, wish to be seen to do 'the right thing' – Google famously states that 'You can make money without doing evil'[77] – then we need to ensure that they know what 'the right thing' is. Rights-based approaches fit perfectly with the symbiotic regulation model suggested above, working directly with existing relationships principally by changing the perspective from which those relationships are viewed – and as noted in Chapter One, those rights are not seen as 'absolute', as 'trumps' which overrule other laws or interests, but as part of the balancing act, as part of what Murray calls the 'regulatory matrix'[78]. Moreover, the existence and expression of these rights can support, inform and empower the key communities that are at risk of being 'squeezed' – the communities of individuals whose autonomy and privacy are under threat.

[77] Item 6 on Google's statement of corporate 'philosophy', accessible at http://www.google.com/corporate/tenthings.html

[78] See MURRAY, A. D. 2006. *The Regulation of Cyberspace: Control in the Online Environment,* Milton Park, Abingdon, UK ; New York, NY, Routledge-Cavendish., Chapter 8

There are rights that can help address the issues arising through the Symbiotic Web directly. A right to roam the internet (and use its principle tools) with privacy; a right to require proper consent before data is gathered and used (though this is nominally in place through Data Protection law, the reality is very different); a right to monitor those who are monitoring us; a right for people to have data held about them deleted, even if that data is accurate. These and other related rights, together with ways to enact them, form the core of Chapter Six and ultimately offer the best answers to the challenges of the Symbiotic Web.

Chapter 3 - Data Protection, Data Retention and Internet Searching

1 Competing interests over data

There are competing interests at play when considering the gathering and holding of personal data – these include governments' interests in security, the business interest in economic success, and the private individual's interests in privacy[1]. It is clear that in this field the interests of business and government in economics and security respectively coincide to a great extent. Both benefit from having access to as much information as possible on individuals, and being able to use that information however they want – even when this conflicts with the needs or wishes of those concerned for their privacy, and ultimately therefore their autonomy.

Two of the key concepts in the regulation of this area arise directly from the protection of these interests. Data protection is intended to protect private individuals' interests in privacy, while data retention is designed to protect governments' interests in security – originally, specifically their interest in preventing terrorism and the spread of organised criminal activity. The respective origins of the two systems are discussed in section 2, after which - in section 3 – we discuss the role of search engines in the current internet, and how data protection and data retention apply to their activities and the data they gather. Commercial interests do not possess specific legal protection but in practice, as the case study in section 4 will show, they use whatever legislation is in place to attempt to gain advantage. In this case, this means trying to use data retention to 'trump' data protection.

[1] Governments are interested in more than just security, and individuals in more than just privacy – however, in this context, this is the most significant dynamic.

1.1 Search Data

Certain forms of data are of particular significance when looking at how data are gathered and stored, especially when considered from the perspective of autonomy. Search data is perhaps the most important of all, for a number of reasons. The first of these is related to its nature. As well as including some of the most private, even intimate things, the most mundane and ordinary things we search for can be perfect information for profiling. Secondly, this importance lies in the role searching plays in the internet – it is the primary method not only of finding things but also of navigating the web. This has direct implications when looking at our ability to find our 'own' way around the web, rather than having ways chosen for us – as will be discussed, search engines have a direct role to play in what was introduced in Chapter Two as the Symbiotic Web. These first two factors are discussed in section 3, which looks at the role of searching and search engines.

The third factor is that of choice – a crucial factor from the perspective of autonomy. Searching and search data are important because there is little choice involved as to whether to use a search engine at all – almost all search engines work similarly, gathering data and using it for their own purposes[2]. As searching is a fundamental tool of the internet, this makes it almost impossible to use the internet at all without data being gathered.

Fourthly, and perhaps even more importantly from the perspective of autonomy, there is the issue of consent. Consent is one of the keys to data protection – if you can get express, informed consent from someone to use their data, some of the key aspects of data protection law are effectively bypassed.[3] The argument that search engines get any kind of consent at all, let alone express, informed consent, is highly contentious given that it is not

[2] The only significant exception is the 'AskEraser' offered by Ask.com, discussed in section 6.

[3] Certain key data protection principles such as data minimisation, subject access rights and protection against function creep remain in place even with consent, but if broad enough terms are set when the consent is given restrictions on the use of data are much reduced.

made clear when you make a search that data are being gathered at all. Other kinds of data such as the 'social' data gathered by social networking sites do at least have a login procedure and terms and conditions to be agreed to, providing at least a surface level of consent. How Google's particular methods of obtaining what it considers to be consent for the use of search data fit with the terms of data protection will be looked at in section 2 of this chapter, and consent itself will be examined in more depth in Chapter Four.

These are just some of the reasons that search data has become the first real battlefield where the potential conflict between data protection and data retention is being fought – in the ongoing dispute between Google and the EU Article 29 Working Party (see below), a case study of which forms section 4 of this chapter. It is a critical conflict not least because it could be the first of many such confrontations as the struggle over data retention spreads into different areas and different kinds of data. The next in significance might be clickstream data,[4] but the battle over that has yet to be fought – most of the arguments that apply to search data apply similarly to clickstream data.[5]

The final sections of the chapter look at why we should be concerned about all this, and at possible resolutions to the problem and implications for the rest of the thesis. The most radical solution suggested arises directly from the case study, though its implications go far beyond it. It is a positive answer to the question of whether we have, or should have, the right to navigate the internet and to use the main tools of the internet without having to have our data gathered and our activities not only scrutinized but the results of that scrutiny held in case it is of some use either to business or to governments. In this context, in the conclusion to this chapter a right is proposed, which is intended to form part of the overall right to informational privacy, namely a 'right to roam with privacy'. This would be a right to roam the internet without having data gathered and recorded, and a key point of this right is

[4] Clickstream data records links followed and websites visited (and related data) and is gathered generally by ISPs.

[5] The issues of the nature of the data (private/intimate and mundanely revealing) of the crucial role in the use of the Internet, of consent, and of choice (all ISPs gather clickstream data) are relevant in similar ways.

that the default position is that data are not gathered for storage and use unless an express choice has been made, for example by logging in to a premium, specified service. Further, that search engines (and, in the case of clickstream data, ISPs) are effectively required to offer their basic services in an alternative, 'privacy-friendly' form.

It is a radical suggestion, and one that would require a significant change in the approaches of search engines, ISPs and other web-providers – and in governments' approaches to security. It is, however, something that could ultimately be of benefit to all. It is also a part of the process of changing the paradigm of internet use so that privacy is the default, and that surveillance and data gathering are the exception – either opted into or brought into play when needed for security or other similarly important reasons.

The place of the internet in modern society was discussed in Chapter One, suggesting in particular that the internet can no longer be considered as an 'optional extra', but is now a crucial and intrinsic part of our lives. One of the consequences of that is that our rights in our 'real' lives need to be extended and expanded into rights online – and if the internet is to be considered a 'public' space, then the right to roam is part of that extension and expansion.

2 Data Protection and Data Retention

2.1 The Data Protection Directive

The Data Protection Directive (Directive 95/46/EC)[6] is a key pillar in European privacy regulation. The principle idea is for citizens and residents of the EU to have some degree of control over what information is stored about them, and what is done to that information, and that they should be assured that the data is processed, stored and used only for appropriate and

[6] Directive 95/46/EC, downloadable from http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf

legitimate purposes[7]. It is a powerful piece of legislation and sometimes a point of pride for European privacy advocates, particularly in comparison with the situation in the US[8]. There are problems with it, not least the inconsistent and sometimes weak implementation across the states of the European Union as well as weaknesses of enforcement. Nonetheless, by world standards it is a strong and comprehensive piece of legislation.[9]

The Directive arose directly from the felt need, derived from the Convention for the Protection of Human Rights and Fundamental Freedoms[10] ('ECHR') to provide some of the protection required under Article 8, the right to respect for private life[11]. It is a human rights document, and should be viewed as such – this seems sometimes to be forgotten, and the Data Protection Directive looked at more as a technical document, or a 'trade' document regulating businesses rather than protecting individual rights.

2.1.1 Objectives, Scope, Terms and Definitions

Article 1, setting out the objectives of the Directive, states: 'In accordance with this Directive, Member States shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data.'[12]

---

[7] See 'Data Protection in the European Union', http://ec.europa.eu/justice_home/fsj/privacy/guide/index_en.htm

[8] As Cate puts it: 'protection for information privacy in the United States is disjointed, inconsistent, and limited by conflicting interests' CATE, F. H. 1997. *Privacy in the Information Age,* Washington, D.C., Brookings Institution Press., p98

[9] The situation in the US has already been noted. In Asia, the concept of data privacy is still to be fully accepted (see e.g. http://www.caslon.com.au/privacyguide6.htm)

[10] Downloadable from http://www.echr.coe.int/NR/rdonlyres/D5CC24A7-DC13-4318-B457-5C9014916D7A/0/EnglishAnglais.pdf

[11] Article 8 states that
'(1) Everyone has the right to respect for his private and family life, his home and his correspondence.
(2) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.'

[12] Data Protection Directive, Article 1

The scope of the Directive is set out in Article 3, which states that it applies to 'the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system.'[13] Effectively, this means that it covers all computer systems – and prevents any possible avoidance techniques involving transferring computer systems onto paper records or their equivalents.

Article 3 allows exceptions for operations concerning public security, defence, state security and the criminal law, and for processing 'by a natural person in the course of a purely personal or household activity'. The scope is broad, potentially covering for example all commercial, financial, medical and recreational systems as well as the vast majority of government records, and most of the data on the internet.

The definitions used in the Directive, set out in Article 2, are where the legislation's strengths and weaknesses start to appear. 'Personal Data' are defined as 'any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity'[14]. This is a broad, strong definition. The definition of what constitutes an 'identifiable' person is similarly strong, and could cover all kinds of profiling as well as the more obvious direct definitions – but is (and perhaps inevitably) somewhat vague and subject to argument.[15]

---

[13] ibid, Article 3
[14] Data Protection Directive, Article 2(a)
[15] The Information Commissioner's Office in the UK produces a 'specialist guide' as to what constitutes personal data, downloadable from http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/personal_data_flowchart_v1_with_preface001.pdf. The guide was produced in response to a narrower than expected interpretation of the term in a Court of Appeals case (Durant vs FSA, 8/12/2003). The fact that it was needed reflects the problems in the language of the UK's implementation of the Directive, the Data Protection Act 1998 (available online at http://www.opsi.gov.uk/Acts/Acts1998/ukpga_19980029_en_1) which to this extent at least follows the Data Protection Directive closely.

'The data subject's consent' is defined as 'any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed.'[16] Again, it can be argued that this is a strong definition – but terms like 'freely given' and 'informed' are hard to pin down. This makes the role of the Article 29 Data Protection Working Party (see below) in interpreting and advising on the implementation of the Directive critical.

2.1.2 Processing of Data

Article 7 of the Directive details the criteria for making data processing legitimate. The first and perhaps most important is when 'the data subject has unambiguously given his consent' – an issue that is fundamental if the impact on autonomy is to be considered. Other basic categories include compliance with a legal obligation, protecting the vital interests of the data subject and performance of a task carried out in the public interest.[17]

There are two specific categories relevant to business use:

- '…for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract';[18] or
- '…for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject which require protection under Article 1 (1)[19]

Essentially, this means that either a contract must be entered into with consent to the processing expressly given, or the data processing must not be of sufficient importance to the individual concerned to warrant protection.

[16] Data Protection Directive, Article 2(h)
[17] Data Protection Directive, Articles 7(a), (c), (d) and (e)
[18] Data Protection Directive, Article 7(b)
[19] Data Protection Directive, Article 7(f)

All of these possibilities might be relevant here – not just in terms of search data, but other types such as clickstream data or the social data gathered by social networking services. Is a contract entered into when a term is put into a search engine or a link followed to a particular website? Is the fact that someone is searching for a particular term or looking at a particular website of significant importance to require that protection? Search engines or ISPs could put forward arguments under all three of these: that a contract has been entered into, that search data or clickstream data is not really personal data, or indeed that even if it is personal data that it is not, of itself, of sufficient significance to require protection.

Google has not attempted to use these arguments, as the case study below will show – though the Working Party in its Opinion 148 (see section 4 below) to a certain extent anticipated that it would if the conflict went far enough. The contractual argument appears the strongest, in spite of the Working Party's objections, but has other implications, not least bringing into play the question of explicit, informed, freely given consent and potentially of unfair contract terms.[20]

Google's 'terms of service'[21] set out what purports to be a contractual arrangement, stating that you must agree to the terms and may not use Google's services unless you accept these terms. You can accept the terms by either:

> "(A) clicking to accept or agree to the Terms, where this option is made available to you by Google in the user interface for any Service; or
>
> (B) by actually using the Services. In this case, you understand and agree that Google will treat your use of the Services as acceptance of the Terms from that point onwards."

[20] E.g. in the UK under the Unfair Contract Terms Act 1977.
[21] See http://www.google.co.uk/accounts/TOS

In accepting the terms, you accept Google's privacy policy,[22] which gives a number of categories of possible use of your personal data. The first and most significant of these is "Providing our products and services to users, including the display of customized content and advertising", which gives Google a great degree of freedom of action, and as the case study will show, the Working Party does not consider sufficiently specific to meet the requirements of data protection.

But do the Terms of Service constitute consent? Google's two cases (A) and (B) differ significantly. In (A), the user must click to accept, whilst in (B) their acceptance is assumed on the basis of use – here, this means that by choosing to search for something, you are deemed to have accepted the contract and all its terms. 'Boilerplate' click-wrap contracts are commonplace in online business, and have often been found to be legitimate and enforceable.[23] That validity has sometimes depended on the specific 'clicking' of acceptance,[24] but in *Cairo v Crossmedia*[25] a 'browse-wrap' contract similar to Google's case (B) was upheld – though the case is contentious[26] and there are significant differences, not least the fact that website in question had a notice on every page reminding user that *'by continuing past this page and/or using this site, you agree to abide by the Terms of Use for this site,'*[27] while Google's search page does not even mention their Terms of Service.

When considering the Directive's requirement for express, informed and freely given consent, it seems much more clear cut. If you access the services (as most users do) without logging on or clicking to accept, your consent cannot be express, and if you are not required to look at either the terms of

[22] See http://www.google.co.uk/privacypolicy.html
[23] Examples include *Hotmail Corp. v. Van$ Money Pie Inc.*, 1998 WL 388389, *6 (N.D. Cal. 1998) and *Ticketmaster L.L.C. v. RMG Tech., Inc.*, 507 F.Supp.2d 1096, 1102-1103 (C.D. Cal., 2007)
[24] E.g. *Specht v. Netscape Communications Corp.*, 150 F. Supp. 2d 585 (S.D.N.Y.2001),
[25] *Cairo, Inc. v. Crossmedia Services*, Inc., 2005 WL 756610 (N.D. Cal. Apr. 1, 2005)
[26] E.g. see http://blog.ericgoldman.org/archives/2005/04/cairo_v_crossme.htm
[27] See commentary on this case at http://www.internetlibrary.com/cases/lib_case389.cfm

service or the privacy policy (links to neither of which can be found on the search page[28]) your consent cannot be assumed to be informed.

As will be seen in the case study, the Working Party has come to a similar conclusion – that anonymous users (i.e. those who have not logged on) cannot be deemed to have consented to the use of their data.

Considering consent even in case (A) gives rise to further questions. Firstly, when is consent truly informed? Is 'legal' consent really sufficient in any broader sense? Reading documents that are legally satisfactory but almost meaningless to a layperson, or scrolling down a long document without even reading it and then clicking 'OK' does not mean that the 'consenter' is informed in a meaningful way. For many purposes this would not matter – a *caveat emptor* approach is often appropriate – but if privacy and autonomy are of the fundamental importance suggested throughout this thesis, then where they are concerned something stronger should be required. The issue of consent will be examined in depth in Chapter Four, and possible solutions to the problem suggested.

Secondly, if the use of search engines is a crucial tool in the use of the internet, then can any such consent be 'freely given'? As discussed in Chapter Two, internet use is becoming an essential part of our lives in modern society, and hence so is the use of search – so we have little choice but to use it. The lack of consent of anonymous users and the question of the 'freely given' nature of any consent even for logged-in users – could be solved by the requirement of search engines to provide two levels of service, an anonymous one without data gathering and a 'premium' service with full log-in and properly informed and freely given consent. This forms part of the possible solutions that will be suggested in the conclusion to this chapter.

[28] This in itself is subject to challenge under Californian law – see http://news.bbc.co.uk/1/hi/technology/7434558.stm and http://blogs.zdnet.com/Howlett/?p=402

2.1.3 Data quality requirements

Article 6 of the Directive sets out data quality requirements.

1) Data must be accurate and kept up to date, and 'every reasonable step' must be take to ensure it stays that way
2) Data must be collected only for 'specified, explicit and legitimate purposes', and not 'further processed in a way incompatible with these purposes'. This protects against 'function creep', an important risk for personal data.
3) Data must be 'adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.' The 'not excessive' part is crucial – this calls for 'data minimisation'.
4) Data must be 'kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed.' Effectively, this calls for anonymisation of data as soon as possible.

These latter three protections (against function creep, for data minimisation, and for minimum retention periods) are particularly important in the context of the internet.

2.2 Implementation and the Article 29 Working Party

Interpretation and advice on the implementation of the Data Protection Directive has been placed in the hands of the 'Working Party on the Protection of Individuals', otherwise known as the Article 29 Working Party,[29] or simply the Working Party, which is composed of representatives of the supervisory authorities on data protection of each of the Member States of the EU – including the UK's Information Commissioner, Christopher Graham. As shall be shown below, it is a strong, expert and authoritative

[29] Established by Articles 29 and 30 of the Directive

body unafraid of taking on the biggest of players in the market. It is the Working Party that is at the centre of the EU dispute with Google.

There are problems in the ways that the Data Protection Directive has been implemented across the member states – not just variations between states, but significant problems that appear to apply throughout the EU, most notably under-resourced enforcement, patchy compliance by data controllers and a low level of knowledge of their rights by data subjects.[30] Nonetheless, the first implementation report concluded that:

"Despite the delays and gaps in implementation, the Directive has fulfilled its principal objective of removing barriers to the free movement of personal data between the Member States. The Commission also believes that the objective of ensuring a high level of protection in the Community has been achieved since the Directive has set out some of the highest standards of data protection in the world."[31]

2.3 The Data Retention Directive

The Data Retention Directive (Directive 2006/24/EC)[32] arose from a very different source, one that has often been in tension with human rights – counter-terrorism[33]. The idea was essentially that governments and other agencies engaged in counter-terrorism might be able to find terrorists or potential terrorists through their actions (and data trails) on the internet. Data retention, therefore, requires those who provide 'publicly available electronic communications services' or 'public communications networks'[34] to keep records sufficient to identify individuals involved, for periods of up to

[30] See the first implementation report: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0265:FIN:EN:PDF
[31] Summary of first implementation report: http://europa.eu/scadplus/leg/en/lvb/l14012.htm
[32] Directive 2006/24/EC, available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006L0024:EN:HTML
[33] One trigger for the passing of the Directive was the terrorist attacks in London on 7th July 2005, referred to in paragraph 10 of the preamble.
[34] Data Retention Directive, initial description.

24 months, so that when necessary government agency can examine those records to try to catch the people concerned. Whilst data protection is essentially a European concept, data retention is something taken seriously all over the world, and in particular in the US.

The Data Retention Directive is an amendment to Directive 2002/58/EC,[35] the 'Directive on privacy and electronic communications', which is an adaptation of and replacement for Directive 97/66/EC,[36] 'concerning the processing of personal data and the protection of privacy in the telecommunications sector', which itself 'translated the principles' set out in the Data Protection Directive into 'specific rules for the telecommunications sector'.[37]

This somewhat complex evolution of directives demonstrates a number of things: firstly, that the EU takes the issue of data privacy seriously, and is constantly looking at what needs to be done to protect it; secondly, that its efforts have not been entirely successful, as directives have needed to be superseded and amended and have become increasingly specialised; thirdly, that this is probably in itself inevitable as the technology and its applications have been proceeding apace, and that regulation is almost always 'playing catch-up'. And fourthly, given the nature of the Data Retention Directive, this complex evolution demonstrates the sometimes-conflicting needs and motives at play within the EU – there are those at the EU who want to counter some of the most important principles of data protection in the names of either security or law enforcement.

2.3.1 The Scope and Terms of the Directive

Article 1 sets out the subject matter and scope and has a number of key elements:

[35] The 'ePrivacy Directive', downloadable from http://eur-lex.europa.eu/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf
[36] Downloadable from http://eur-lex.europa.eu/pri/en/oj/dat/1998/l_024/l_02419980130en00010008.pdf
[37] Directive 2002/58/EC, paragraph 4 of the preamble.

1) It applies to 'providers of publicly available electronic communications services' and 'public communications networks';
2) It requires the retention of 'certain data' 'in order to ensure that the data are available for the purpose of the investigation, detection and prosecution of serious crime, as defined by each member state in its national law'
3) It applies to 'traffic and location data on both legal entities and natural purposes and to the related data necessary to identify the subscriber or registered user'.
4) It does not apply to the 'content of electronic communications, including information consulted using an electronic communications network'.

The Directive is intended to ensure that email services, digital telephony services (including mobile telephony services) and their equivalents, and where appropriate internet access providers, keep sufficient information to enable law enforcement agencies to determine who has been communicating with whom,[38] and when and how – but not what they've actually been communicating. Precisely what that data should consist of is set out in the Directive for each of the main services covered by the Directive: fixed network telephony and mobile telephony, internet access, internet email and internet telephony.[39] On the face of it, it does not apply to search engines, but as shall be shown, that did not deter Google from using it for its own purposes.

The data must be made available to 'the competent national authorities in specific cases and in accordance with national law' – and it is left up to the national law to determine 'necessity and proportionality requirements', subject to the 'relevant provisions of European Union law or public

[38] Including records of unsuccessful call attempts – see Data Retention Directive, Article 3.2
[39] Data Retention Directive, Article 5(1).

international law, and in particular the ECHR as interpreted by the European Court of Human Rights.'[40]

It is in appearance a very simple directive – as befits one whose motivation for existence is to some extent political rather than technical or practical. One of the intentions of the Directive is to harmonise the data retention legislation that currently exists in a number of member states[41] – as the case study will show, a lack of harmonisation is something that international organisations (of which Google is a prime example) can use to attempt to effectively avoid legislation entirely. This laudable aim is unfortunately far from successful in that the terms of the Directive still allow member states significant leeway in key areas, not least in terms of the periods of obligatory data retention – the Directive allows retention periods from 6 months to 2 years.

Implementation of the Directive was neither straightforward nor consistent. As well as the various legal challenges noted below, 16 of the member states (including the UK) took advantage of provisions allowing for a postponement in implementation of the Directive[42] – an indication of the difficulties that implementation represented from a legal, technical and political perspective. The UK's implementation, the Data Retention (EC Directive) Regulations 2009,[43] was drafted to attempt to overcome the technical and financial objections of UK ISPs and phone operators who complained about being expected to provide (and pay for) technical solutions to store data, but in the process of drafting ran up against very significant concerns from both a legal and political perspective, drawing criticism from the Information Commissioner's Office[44] and other experts and political opponents.[45]

---

[40] Data Retention Directive, Article 4
[41] Data Retention Directive, Article 1, and preamble paragraphs (5) and (6)
[42] Declarations made by these 16 member states are included in the Directive
[43] Available online at http://www.legislation.gov.uk/uksi/2009/859/contents/made
[44] The ICO said that 'We have real doubts that such a measure can be justified, or is proportionate or desirable' – see http://www.ico.gov.uk/upload/documents/pressreleases/2008/proposed_government_database.pdf

2.3.2 Criticism and challenges

The Data Retention Directive has been subject to significant criticism, not least from the Working Party itself. Peter Hustinx, the European data protection supervisor, has said that 'The Data Retention Directive – turned the rules upside down'. In its official response, the Working Party said 'Traffic data retention interferes with the inviolable, fundamental right to confidential communications', and went on to question 'whether the justification for an obligatory and general data retention… …is grounded in crystal-clear evidence.'[46]

The Working Party's criticism was extensive.[47] In an opinion issued in 2005,[48] prior to the enactment of the Directive, it made observations and suggestions that it hoped would lead to changes in the final document. These included the observation that the aims of the Directive, although ostensibly to fight terrorism, in practice allowed data retention for 'the purpose of the investigation, detection and prosecution of serious crime, as defined by each Member State in its national law[49]' – function creep within a single directive, and precisely the kind of function creep that has characterised the worst kind of anti-terrorism legislation.[50]

The Working Party also noted that no evidence had been presented to clearly justify compulsory and general data retention. This is characteristic of anti-terrorism legislation, but is particularly relevant in this area, with the unprovable justification that to provide such evidence would require

---

[45] See for example http://news.bbc.co.uk/1/hi/technology/7410885.stm and http://news.bbc.co.uk/1/hi/uk/7409593.stm
[46] Opinion 113 of the Working Party, page 2, downloadable from http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/index_en.htm
[47] The Working Party had also issued strong opinions against previous, failed, attempts to institute data retention, in 2002 (Opinion 64) and 2004 (Opinion 99)
[48] Opinion 113, footnote 38
[49] Data Retention Directive, Article 1
[50] See for example GEARTY, C. A. 2006. *Can Human Rights Survive?,* Cambridge, Cambridge University Press., Chapter 4, for an exploration of the problems of anti-terrorism legislation.

revealing methods of detection and prevention that would compromise security. They also suggested that there are other methods of investigation or detection that infringe on privacy much less – notably the 'quick freeze-procedure' which involves surveillance and data retention only in well-founded cases, where law-enforcement agencies specifically request the storage of certain data.

The Working Party's suggestions for amendments to the Directive included the following:

1) That the Directive should suggest maximum rather than minimum periods of data retention, providing protection for citizens against over-zealous Member States.
2) A raft of specific safeguards intended to ensure proper scrutiny and control even if the major criticisms of the Directive outlined above were rejected. These included provisions against data-mining and further processing,[51] and against the use of retained data solely for public order purposes, and clear definitions of data categories to be held, and a 'no-contents' provision ensuring that only traffic data and not contents would be held. These last two (clear categorisation and 'no-contents') were included in the final Directive.

Other than a few of the specific safeguards in above, none of these criticisms resulted in changes to the draft Directive. All remain relevant and have been the basis for continuing criticism since the Directive came into force on 3 May 2006. The Working Party issued a further opinion, subsequent to the adoption of the Directive, again voicing its reservations, and urging 'uniform, European-wide implementation of the Directive'[52] and stressing the need for stringent and specific safeguards. The Working Party's concerns about consistency have proved well grounded at least in one way – Google used the inconsistencies and vagueness of the Data Retention Directive as one of its

[51] Both discussed in Chapter 5
[52] Opinion 119 of the Working Party downloadable from: http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/index_en.htm

tools in its attempt to rebuff the challenges of the Working Party, as will be shown below.

This criticism is further deepened when regard is had to the individual implementations of the Directive currently envisaged. In the Netherlands, the Dutch Data Protection Agency issued an opinion suggesting that the bill proposed to implement the Data Retention Directive breached Article 8 of the ECHR[53], while Digital Rights Ireland wrote to the Irish government threatening legal action to scupper the Irish Government's attempts to implement the Directive[54]. Further to this, the Irish Government has itself taken action to attempt to have the Directive repealed, effectively on the grounds that it was passed as though it was a trade directive, not a security directive[55] – the latter directives being much harder to pass, with more stringent requirements including a unanimous rather than simple majority vote.[56] In April 2008 an amicus brief was submitted to the European Court of Justice by 43 NGOs from 11 countries in support of this case, suggesting that data retention 'violates the right to respect for private life and correspondence, freedom of expression and the right of providers to the protection of their property'.[57]

In the light of the decision over Passenger Record Numbers[58], where a similar argument prevailed, this case appeared to have a reasonable chance of success – but in the event it failed on 10 February 2009. The Court suggested that the act of retention, as specified by the Directive, was simply an

[53] The Dutch Data Protection Authority's opinion is downloadable from http://www.dutchdpa.nl/downloads_adv/z2006-01542.pdf?refer=true&theme=purple
[54] See http://www.digitalrights.ie/2006/07/29/dri-challenge-to-data-retention/
[55] Under Article 95 of the Treaty Establishing the European Community, downloadable from http://eur-lex.europa.eu/en/treaties/dat/12002E/pdf/12002E_EN.pdf. Measures based upon Article 95 must have as their "centre of gravity" the approximation of national laws to benefit the functioning of the internal market.
[56] See http://curia.europa.eu/jurisp/cgi-bin/gettext.pl?lang=en&num=79939084C19060301&doc=T&ouvert=T&seance=REQ_COMM and for an analysis http://www.out-law.com/page-7310
[57] See http://www.vorratsdatenspeicherung.de/images/data_retention_brief_08-04-2008.pdf
[58] *European Parliament v. Council and Commission*, Joined cases C-317/04 and C-318/04. In May 2006, arrangements to transfer the "passenger name records" of air passengers from the EC to the US Bureau of Customs and Border Protection, which had been set up on the basis of Article 95, were held to be illegal and annulled.

obligation on service providers, and effectively that the Directive essentially harmonised that obligation, and hence could be viewed as a trade directive. The security aspects – the investigation, detection and prosecution of crime – were held to be a separate issue, independent of the Directive.[59]

In December 2007, more than 30,000 German citizens took up a case against the implementation of the Directive in Germany.[60] In March 2010, responding to the challenge, the German Federal Constitutional Court struck down their implementation of the Directive, describing the law as a "particularly serious infringement of privacy in telecommunications".[61] The Directive is not currently implemented in Germany as a result.

In Bulgaria certain terms of their government's implementation of the Directive were repealed in 2008, for their vagueness[62] whilst in Romania in 2009 the constitutional court declared data retention itself to be in breach of the fundamental right to secrecy of correspondence.[63] The exact impact of this declaration has yet to be determined – a new draft law was presented in June 2011, but reports suggest that it is substantially the same as the law that was previously rejected by the constitutional court, and that its fate may be the same.[64]

Criticism of the Data Retention Directive has continued. In December 2010, Peter Hustinx called for a European Commission review of the Data Retention Directive to prove that it had achieved results – by demonstrating

[59] The case is Case C-301/06, *Ireland v European Parliament and Council of the European Union*, and the judgment may be viewed at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62006J0301:EN:HTML
[60] See for example http://www.dw-world.de/dw/article/0,2144,3025009,00.html
[61] See for example http://news.bbc.co.uk/1/hi/world/europe/8545772.stm
[62] See for example http://www.edri.org/edri-gram/number6.24/bulgarian-administrative-case-data-retention
[63] See for example http://www.openrightsgroup.org/blog/2009/data-retention-rejected-by-romanias-courts
[64] See for example http://www.edri.org/edrigram/number9.13/new-draft-data-retention-romania

how many cases it had actually been needed on – or repeal it.[65] There is no sign yet, however, that there is any likelihood of repeal – and authorities show few signs of dampening their enthusiasm for the concept. Indeed, there are signs that the idea could be spreading: in the US, the Department of Justice has been renewing calls to introduce mandatory data retention for ISPs, something that has not appeared possible to date, and goes counter to the US principles of free trade.[66] Even to academics like Claire Walker, data retention seems to have become part of the landscape. As she put it:

"Communications data retention and interception have become a non-negotiable fact of modern life."[67]

3 Search engines and their role

Search data has become the first real battleground in the conflict between data protection and data retention. Two aspects of searching lie behind this: the nature of the data gathered, and the crucial role that search engines play in our navigation of the internet. This combination of factors not only gives search engines massive business opportunities but also raises significant concerns about potential risks to privacy and autonomy. Search engines to a great extent govern what people find on the internet – the choices that are made available to people about where and what to visit. The way in which they do this is crucial, particularly from the perspective of autonomy. The part they play in the development of the Symbiotic Web is very significant – as will be discussed below.

[65] Hustinx's speech can be found at http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/Publications/Speeches/2010/10-12-03_Data_retention_speech_PH_EN.pdf
[66] See for example http://news.cnet.com/8301-31921_3-20029423-281.html
[67] WALKER, C. 2009. Data retention in the UK: Pragmatic and proportionate, or a step too far? *Computer Law & Security Review*, 325-334. P333

3.1 How does Google make its billions?

The search market is competitive, and though Google is dominant (with close to 85% of the market according to some estimates,[68] though approximately 65% in the US[69]), there are other big players in the market, particularly in the US, where Yahoo (with approximately 16%) and Microsoft's Bing (with approximately 14%) provide strong competition[70] – and since Yahoo is now powered by Bing that means Microsoft have an effective 30% market share in the US. There are smaller players in the market too, such as Ask and AOL.

Google's business model is supremely effective. Its most recent financial results show revenue of $8.58 billion for the first quarter of 2011.[71] How is Google so successful? At first sight searching doesn't appear to be an obvious profit-maker. It's free to use, and looks as though it's just a service, provided out of some sense of public goodwill. Google's website, for example, is a paragon of clarity – a clear white page, devoid of advertising or distraction, with a blank box in which to enter your search data.[72]

The nature of that data needs to be carefully considered. 'Googling', a term now in many dictionaries,[73] is one of the first things that many people do if they are interested in something or in someone. As a consequence, search data are potentially some of the most revealing data of all, including the most intimate and private of information (looking for new jobs or romance, problems with health, sexuality etc) but also the most mundane of information – information that reveals a great deal about our ordinary lives, things like what movies we enjoy and film stars we admire, or what products we buy. Google and their competitors understand this, knowing first of all that if they know what we want to find they can do their best to provide it –

[68] Market shares for the search market are difficult to ascertain. The 85% figure comes from http://www.karmasnack.com/about/search-engine-market-share/ in May 2011
[69] See http://mashable.com/2011/04/11/bing-google-stats/ for example
[70] See http://mashable.com/2011/04/11/bing-google-stats/ for example.
[71] See http://investor.google.com/earnings/2011/Q1_google_earnings.html
[72] www.google.com
[73] For an online example, see http://www.merriam-webster.com/dictionary/google

and not just to help guide us to the places we would like to find but to the places that their advertisers would like us to find.

The intimate information can be important – and potentially highly damaging. From teenagers exploring homosexuality or researching abortion to almost anything connected with religion, there is great potential for harm. The mundane information, however, may have even more significant implications for autonomy, but to understand why, we need to look more closely at how a search engine functions.

There are many possible models for running a search engine, from the basic 'telephone directory' style search engine to the sophisticated data-gathering model used by Google and most of its competitors. The way that Google works begins to become clear when you complete your search – the advertising appears with the results, not with the initial search. The most significant part of Google's business model is its ability to provide *targeted* advertising, appropriate to the user concerned, and thus more likely to be successful. Because Google understands the nature of search data and knows how to use it, the targeted advertising works – and Google can prove it, because their records can show not only who looked at the advertising but also who clicked on it: who actually visited the sites of the advertisers. Moreover, when someone visits that advertiser's site, they can tell that they've come from Google. Advertisers have been convinced – Google's advertising revenue is the key to its success. As the case study will show, however, this ability to target advertising is not mentioned by Google's Privacy Counsel Peter Fleischer anywhere in his explanation to the Working Party as to why Google retains identifiable search logs.

Different degrees of targeting are possible, depending on the kind of information available and used for that targeting. The most direct uses only immediately available information; first and foremost the term searched for, but also things like the general location of the searcher, the time of day the search is made, the ISP that the searcher uses and so forth. The more

sophisticated models used by Google and others use much more information – profiles built up from the logs of all the terms that the user has ever searched for before, together with other data from Google services such as places looked for on Google Maps or Google Earth, that aspect of clickstream data which shows which links from our Google searches we actually follow, and for how long even the texts of emails sent through gmail. Other data purchased from third parties can be aggregated with this, building even more detailed profiles and allowing for even more accurately targeted advertising.[74] This advertising may appear not just on the search results pages, but on any pages provided by any of the many other Google services – including your gmail webmail pages.

It is crucial to remember that the first and most important reason that Google succeeds is because its search engine is very effective – when it first appeared it was streets ahead of the opposition, which is why it got so many users, and hence was so good for advertising. Google wishes to maintain that advantage in order to maintain their pre-eminent position in the market. Their collection and mastery of their database of personal information, personal searches and so forth is one of the keys to maintaining that advantage – and Google knows this, which is why it is not going to give it up without a fight.

3.2 The role of search engines in the Symbiotic Web

Some of the problems arising from the involuntary personalisation of the internet have been discussed in Chapter Two. They are relevant here for two reasons: firstly, because the nature and quantity of search data, as noted above, make it ideal for exactly the kind of profiling that forms the starting point for personalisation, and secondly because of the role that search engines play.

[74] An introduction to the effectiveness of profiling can be found in AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted*, London, John Murray.

The Symbiotic Web is to a great extent about the way that choices are offered to people – the way that people are offered things chosen for them, by others, on the basis of the information that others have about them. In particular, about the web pages people are guided to and that are suggested to them. Search engines play a crucial role in this – if the principal way people navigate the Web is by using a search engine, then the places that the search engines suggest to them are where they're likely to go. Already this has two direct implications. Firstly, when people 'google' something, they almost always only go to sites on the first page of search results,[75] and how a site gets to that first page is essentially up to the search engine. Secondly, as noted above, when search engine results are presented, the page is surrounded by advertisements and 'sponsored links'. Since 2010, Google has been blurring the boundaries between search results and such sponsored links, mixing the results where something might appear in both categories.[76] Moreover, as noted in Chapter Two, the issue of potential bias for commercial reasons in Google's search results, is currently under investigation by the EU in relation to the Foundem case.[77]

These two factors already make the role of search engines in 'shaping' the experience of web users – controlling them, reducing their autonomy – of considerable importance. That is just the first stage. The next is that not only advertisements but that actual content is chosen for people. Where Google is concerned, this could mean that the search results presented are not simply the results of searching for the terms that have been put in, but have been altered ('tailored') based on the individual's prior searching records – and not just the records of what they've previously searched for, but of which of the links that have been presented have actually been followed. As noted in

[75] As noted previously is estimated that 90% of users never look beyond the first page of search results on Google. See for example http://www.search-engine-marketing-australia.com.au/google-analytics.htm

[76] See Danny Sullivan's comments on the subject at http://searchengineland.com/google-blurs-the-line-between-paid-unpaid-results-again-36268

[77] See Chapter Two footnote 38

Chapter 2, Google now attempts to personalise search results unless the user specifically opts out.[78]

Most of this is not harmful in itself – indeed, much of it is helpful and makes a user's experience of the internet more rewarding. The problem is that it takes control out of the hands of the user, in a way that is often in effect dishonest, and puts it in the hands of the website's creators, in this case Google. For the most part, the motives of these creators are at worst pecuniary – but that may not always be the case, and even so it compromises the web-user's autonomy, controlling the options made available to them.

There is a potentially more sinister side to this kind of thing. Most people use Google as their prime method for navigating the internet, so the honesty and transparency of Google searching is critical to ensuring that the internet is something that can be used with the kind of openness, equality and neutrality that give it its strengths, and its possibilities as a tool to improve openness and freedom, rather than as one of limitation and control. If the links presented as the result of a Google search can change depending on the profile of the visitor, that gives scope for discrimination. This discrimination could arise through differential pricing and differentiated access to services based on anything from locale to ethnic origin, education or health. Websites might not even be visible to visitors who are not deemed 'suitable'.

As Ayres has suggested,[79] profilers believe (and they may often be right) that they know people's tastes better than they know them themselves, so the opportunity for people to have their online lives 'shaped' for them by others can be extreme. With the massive growth in Social Networking sites, they not only know someone, they know their friends, and can (and do, in the case of Facebook and Beacon, discussed in Chapter Four) use that information for their own purposes.

---

[78] see http://googleblog.blogspot.com/2009/12/personalized-search-for-everyone.html
[79] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray., Chapter 1

4 Google and the Article 29 Working Party

The first serious conflict between data protection and data retention has been over search data, and Google in particular.. The size of the search market and the importance of search engines in our everyday use of the internet, together with the nature of the search data as discussed above, is one of the reasons for this. Google's dominant position in that market – indeed, its dominant position in the internet world as a whole – made it an ideal target for the Article 29 Working Party. What is more, it exemplifies some of the most important problems arising from the tension between data protection and data retention – showing how in practice businesses can use these tensions to their own advantage, one of the reasons that more needs to be done to protect the individual.

There is a huge amount at stake, not just the tension between data protection and data retention. This includes the power-relationships between big business and the European regulators and the question of how an American corporation should react not just to a European regulator but to a European principle, for as noted above data protection is still essentially a European concept. European regulators have been locking horns with Microsoft for years and appear to have finally won, with Microsoft accepting their near half billion Euro fine[80]. Now, it seems, the regulators are going after the second American monolith of the internet industry, Google.

The Working Party did not choose Google as its target because Google is the 'worst' of the search engines providers – if anything, it is one of the more ethical providers. Certainly the records of Yahoo and Microsoft in China do not inspire much hope in terms of ethical practice[81] - while Google can justifiably claim that it was the only one of the major providers to resist the

[80] e.g. news.bbc.co.uk/1/hi/business/7056288.stm
[81] See the Amnesty International Report, 'Undermining Freedom of Expression in China: The Role of Yahoo!, Microsoft and Google', downloadable from http://www.amnesty.org/en/library/info/POL30/026/2006

US government's attempts to subpoena vast amounts of consumer data in the name of counter-terrorism[82]. Google's dominance of the market may, however, be of concern to the regulators. Searching is the primary method of navigating the internet and Google dominates the search market, so how Google works is of crucial importance in how the internet functions for the vast majority of users.

As we have seen, as services become more personalised and tailored, to a significant degree precisely because of the way Google uses the search data that is the subject of the dispute with the Article 29 Working Party, that 'neutrality' and independence is compromised. So too is Google's role as a disinterested navigator of the internet. Should there be any kind of obligation on Google to prevent this from happening? Or should market forces be relied upon so as to ensure that Google itself wants to avoid it? If the latter, is more effort needed to ensure that users of Google and other search engines understand the processes that Google is using, and how Google gathers and uses its data?

The EU Working Party may be also attempting to cement its place as the world's primary setter of data protection and privacy standards and policies. Standards around the world are far from consistent and harmonisation would be highly desirable – the EU Working Party would like to be at the heart of, and perhaps guiding, that harmonisation process.

It may also be true that this whole process may be seen as just a part of the internal conflict within the EU, between those who support the idea of data protection in the name of human rights and those who are striving for stronger data retention laws, in the name of the fight against terror. The vehemence of the opposition shown by the Working Party to the data

[82] See for example http://news.bbc.co.uk/1/hi/technology/4630694.stm

retention laws from their initial conception, as described in Section 2 above, lends some credence to this.[83]

It may be, however, that the most direct and obvious reason is the key – that the Working Party considers that Google's methods of gathering, holding and using data fail to comply with at least the spirit and intent of data protection legislation, and hence constitute a threat to the privacy and rights of members of the European Union. On the face of it this can certainly be argued. As the UK Government's *direct.gov* website says, personal information should be 'not kept longer than necessary',[84] and it cannot really be said to be 'necessary' that Google makes as much money as possible from us as individuals.

4.1 A dispute over search data

What the Working Party and Google were arguing about is the retention of search data in a form that can be linked – identified[85] – with the person who made the search. At the start of the dispute, as outlined above, Google kept a permanent record of everything that you searched for, building up a detailed profile of all your searches and thus of your interests, tastes and tactics in using the internet. The EU Working Party questioned whether such records should be maintained, and whether they were in compliance with EU data protection legislation.

The first salvoes were fired at the 28th International Data Protection and Privacy Commissioners Conference, held in London in November 2006,[86] where a strongly worded resolution on Privacy Protection and Search Engines was passed, making three key points: that search engines should

[83] The significance of the issue is reflected throughout the 'human rights world' – indeed, the title of Conor Gearty's 2006 book, 'Can Human Rights Survive?' shows how seriously the challenges to human rights from are taken – and the challenges from anti-Terror legislation form a key part of that book.
[84] www.direct.gov.uk/en/RightsAndResponsibilities/DG_10028205
[85] The issue of identity is a contentious one and is discussed further in Section 5
[86] See http://www.privacyconference2006.co.uk/

inform their customers up front about the data being gathered; that search engines should offer services in a 'privacy friendly manner'; and that data minimisation should be regarded as crucial.[87]

It was a general resolution, more a statement of principles than a specific call for action. It did signify intent – and was a sign of things to come in another way, in that the dispute took place in public, with both sides allowing their 'private' correspondence to enter the public domain as they appeared to try to win hearts and minds (of the other key regulators, the rest of the industry, the public and legal and other commentators) as well as making their legal points.

The next move came in the form of a letter from the Working Party to Peter Fleischer, Google's 'Privacy Counsel' for their European operations on 16th May 2007, and made public almost immediately online.[88] In it, the Working Party questions Google's need to keep identifiable search records, suggesting that Google's current practice 'does not seem to meet the requirements of the European legal data protection framework'.[89]

The areas of possible non-compliance arise essentially from principles rather than precise legal rules. The relevant principles, some of the key principles of data protection described in Section 2 above, are:

1. That the purposes for which data is kept must be specified
2. That data should only be held when it is 'strictly necessary' for the provision of the service
3. That services should be offered in a 'privacy-friendly manner'[90]

---

[87] The resolution is downloadable from http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_annex_16_05_07_en.pdf

[88] Letter from Peter Schaar to Peter Fleischer, 16 May 2007, made available online at http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_16_05_07_en.pdf (accessed 30 June 2010)

[89] ibid, page 1

[90] This principle was set out in the Resolution on Privacy Protection and Search Engines of the 28th International Data Protection and Privacy Commissioners Conference, November 2006

4. That data linked to an individual user should only be kept where that user has given his 'explicit, informed consent'
5. Data minimisation – that the minimum amount of data should be held and for the minimum amount of time

The Working Party asked Google to explain why the data was kept, and why for so long – at the time that the dispute began, it was being kept indefinitely, and by the time that the letter was sent their policy was to hold search log data for 18-24 months. The Working Party was asking Google to justify itself, and asking it publicly.

Peter Fleischer responded on 10th June 2007, with a six-page letter that was also made public immediately through Google's own blogs.[91] It was a long and detailed letter, but essentially boiled down to the following:

1. Google needs to keep its search data in order to keep improving the quality of its searches, and to fight fraud and abuse
2. Google believes it is in compliance with all legislation by keeping logs for 18-24 months
3. The principles of data retention mean that Google is obliged to keep its data for 24 months.

The first point appears to be a red herring – there is no need for data used to improve the quality of service to be kept in an identified form, and in general fighting fraud and abuse has to be faster than 18-24 months to be of significant use.[92]

The second and third points are more complex. One key part of Fleischer's argument is very hard to refute – that the laws and directives concerning data protection and data retention are vague and often contradictory. He

[91] Letter from Peter Fleischer to Peter Schlaar, 10 June 2007, made available at http://64.233.179.110/blog_resources/Google_response_Working_Party_06_2007.pdf (accessed 30 June 2010)

[92] The existence of the AskEraser (see Section 5 below) also shows that one of Google's competitors believes it can fight fraud and abuse without holding identitied data more than 'a few hours'

quoted back to the Working Party some of its own criticisms of the Data Retention Directive - in particular, that the EU Data Retention Directive lacks clarity in many ways, that its implementation across EU states is inconsistent, and hence that the rules that govern a provider like Google are far from clear. When you add to the equation regulations from other jurisdictions and Google's stated policy to provide 'one level of privacy protection for our users worldwide',[93] his argument does appear consistent and logical.

Fleischer claimed that there are few hard and fast rules, and that Google's 18-24 month policy is as close to compliance as they can get – and is proportionate to the risks involved. However, as a concession, Fleischer announced that Google would reduce their data retention from 18-24 months to a flat 18 months – a concession out of goodwill, for Fleischer reminded the Working Party that it might have to raise the limit again, as a result of 'future data retention laws'.[94] This small phrase may turn out to be the most important part of the whole letter – as will be explained below. Essentially, however, Fleischer's argument hinges on the supposition that data retention 'trumps' data protection.

4.2 The Working Party's Reponses

The Working Party responded almost immediately to Fleischer's letter, first of all by saying directly that the Data Retention Directive did not apply to search data – and hence that Fleischer's attempt to use data retention as a reason to hold the data did not apply.[95] It did not publish any further correspondence, but instead let it be known that the terms of its enquiry into search engines would be broadened to cover more than just Google, and that it would respond in detail in the early part of 2008.[96]

---

[93] ibid, page 1
[94] ibid. page 5
[95] Within one week, by 27th June 2006. See www.out-law.com/page-8179
[96] See e.g. www.news.com/EU-privacy-body-to-take-months-on-Google-probe/2100-1029_3-6212794.html?tag=html.alert.hed

This response appeared on 4th April 2008,[97] in the form of a comprehensive, published opinion covering searching and search engines generally rather than Google in particular. In it, the Working Party affirms its view that the Data Protection Directive does apply to search data while the Data Retention Directive does not.[98] The Working Party also conclude that "Search engine providers must delete or irreversibly anonymise personal data once they no longer serve the specified and legitimate purpose they were collected for…".

The opinion contradicts all the substantial arguments put forward by Google, going through each of the 'purposes and grounds' suggested by Google and others as reasons for their retention of data (improvement of services, securing the system, prevention of fraud, accounting requirements, personalising advertising, collection of statistics and law enforcement[99]) and rejects them as grounds for long term data retention. Most are 'too broadly defined to offer an appropriate framework to judge the legitimacy of the purpose', while the question of 'reprocessing' data collected for one purpose for another that might be incompatible with it is also considered a problem.[100]

That leaves two possible grounds for data retention – consent and contractual agreements, as discussed in section 2 above. On these, the opinion is unequivocal: ordinary, anonymous users cannot be considered to have given consent and the 'de facto contractual relationship' when using a search engine in its usual form 'does not meet the strict limitation of necessity as required in the Directive'.[101] The Working Party specifically reminded search engines of the Directive's requirements in this area, and noted that 'search engines should make clear to users what information is

[97] Working Party 'Opinion on data protection issues related to search engines', 00737/EN WP 148, Downloadable from http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf
[98] Working Party Opinion 148, Executive Summary, p3, and in detail on pp12-13
[99] Working Party Opinion 148, pp15-18
[100] Working Party Opinion 148, p16
[101] Working Party Opinion 148, p17

collected about them and what it is used for', something that search engines (and Google in particular) conspicuously fail to do.

In conclusion, the Working Party did 'not see a basis for a retention period beyond 6 months', and suggested that if search engine providers wished to retain data beyond that six-month period, they would have to 'demonstrate comprehensively that it is strictly necessary for the service'.[102] The opinion is a strong and comprehensive response, meeting Google's letter head-on and flatly contradicting it on almost every point.

4.3 Google's reaction – a regulatory result?

In the face of such an unequivocal response, Google eventually reacted – six months after the opinion was issued. In a blog post on 8th September 2008, Google said:

> "Today, we're announcing a new logs retention policy: we'll anonymize IP addresses on our server logs after 9 months. We're significantly shortening our previous 18-month retention policy to address regulatory concerns and to take another step to improve privacy for our users."[103]

This step could be seen as a 'victory' of a kind for the Article 29 Working Party. In effect, the pressure it exerted made Google change its policies concerning data, from holding it indefinitely first to holding it for 18 months and, then, here, to nine months. There are limitations to this change – the data is being anonymised, rather than deleted, and the extent to which that anonymisation is effective is somewhat debatable, as will be discussed in Chapter Five. Nonetheless, it does show that companies like Google do respond to pressure – when that pressure is appropriately applied, and is backed up by public opinion. That last point is, perhaps, the biggest key here.

---

[102] Working Party Opinion 148, p19
[103] http://googleblog.blogspot.com/2008/09/another-step-to-protect-user-privacy.html

Google understands the importance of the community, and of the community's views. If enough people care sufficiently about privacy, Google will do something about it. In the case of data retention periods, it did.
In its response, Google explicitly state the challenge:

> "…it's difficult to find the perfect equilibrium between privacy on the one hand, and other factors, such as innovation and security, on the other. Technology will certainly evolve, and we will always be working on ways to improve privacy for our users, seeking new innovations, and also finding the right balance between the benefits of data and advancement of privacy."

This is not just the challenge for Google – it is the challenge for everyone involved in the internet. Ultimately, it is the challenge of the Symbiotic Web – how to make sure that the symbiosis remains positive. What this case study suggests is that the process through which the balance is maintained is one that involves effort, thought and discussion between all the related parties. Google's reaction came through, as they put it:

> "…literally hundreds of discussions with data protection officials, government leaders and privacy advocates around the world…"[104]

This is symbiotic regulation in action. All the various relationships came into play – and ultimately, at least to an extent, the community produced a reaction. As shall be shown in the other case studies in Chapters Four and Five, this pattern has been repeated in other key areas.

5 Implications and ways forward

Should we, as private individuals, be concerned about these practices? First of all, it must be noted that whether they really threaten our privacy and

[104] http://googleblog.blogspot.com/2008/09/another-step-to-protect-user-privacy.html

hence our autonomy is much more significant than whether they comply with the precise terms of the Data Protection Directive.

Search engines play a vital role in the navigation and use of the internet for the vast majority of users. This means that there is a strong argument for people possessing a significant interest in changing search engines' current and possible future practices. Two questions arise – what is the outlook for the future? And what are the implications for attempts to find solutions? There are three connected issues to resolve – firstly the resolution of the specific dispute between the EU and Google, secondly the possibilities of finding a privacy-friendly future for internet searching in general, and thirdly resolving the overall tension between data protection and data retention. The three issues are intrinsically linked, and by finding solutions to one it may be possible to find ways forward in the others.

5.1 The resolution of the dispute between Google and the EU

The protracted negotiation that has taken place between Google and the Article 29 Working Party may have been part of Google's intention – to stretch out negotiations as long as possible while it investigated different options and business models. What appears to have been the final outcome has been a kind of workable compromise. Google may still be vaguely hoping that other parts of governments come to their 'rescue' – the mention of 'future data retention laws' in Fleischer's letter suggested this. In fact, he mentioned it twice – also suggesting that individual countries might implement stronger data retention laws.[105] The principles behind data retention – to be able to find and catch terrorists, organised criminals, paedophiles or pornographers through their internet use – apply as much to search data as they do to emails or telephone calls, no matter what the letter of the law currently says. Google may have been trying to remind Governments of that, to ensure that new, stronger data retention legislation is brought in, applying as much to search as to email, so that Google can use

[105] Fleischer letter p 3

that to trump privacy rights. To date, this has not happened, but there are certainly possibilities – though to an extent it appears that Google may have moved on, and realised that privacy is such an important issue that it needs to be embraced rather than resisted.

In the short term, though, not much could be expected from the legal process – and both sides appeared to know it, which is one of the reasons they did almost everything in public. Hearts and minds were as important as the legal process. Ultimately, much may depend on what the public want and politicians believe is possible. If the interests of individuals in privacy are to be protected, the arguments will need to be won at this level. As shall be discussed in depth in Chapter Seven, where the future of the internet is examined, there are both positive and negative signs about the potential outcome of this argument.

5.2 Privacy-friendly searching?

What options exist if there is to be a more privacy-friendly way to search the web? There are other search engines (see Section 3) but as noted above, if anything most of them have worse privacy policies and practices than Google. In addition, the whole industry knows that gathering private data is one of the keys to Google's success – a key that they're more likely to try to copy than to reject.

Nonetheless, one of Google's smaller competitors has shown one potential way forward. Ask.com offer a service called the 'AskEraser',[106] which effectively allows you to access Ask's search engine without having your data recorded – and also allows you to delete any search records that Ask already hold on you. It seems to satisfy most of the privacy requirements suggested in this thesis.[107]

[106] http://about.ask.com/en/docs/about/askeraser.shtml
[107] See the Electronic Frontier Foundation, analysis: http://www.eff.org/deeplinks/2008/03/search-privacy

There are caveats to the privacy provided by the service in that Ask do agree to cooperate with law enforcement agencies… "Ask.com must abide by federal, state, and local laws and regulations. Even when Ask Eraser is enabled, we may store your search activity data if duly required or requested to do so by law enforcement or other governmental authority. In such cases, we may retain your search data even if AskEraser appears to be turned on." This weakens the privacy – but if used in combination with the kind of 'quick freeze' approach to dealing with terrorism suggested by the Working Party in their Opinion 148 and discussed in section two above could offer a workable compromise.

Ask.com's market share is very small – less than 1%[108] – and this may lie behind their willingness to buck the industry trend and offer a real alternative to the mainstream. In addition, the AskEraser button is small and gray so unless a user knows what they want in terms of privacy they are very unlikely to find it or use it. What is more, the AskEraser is an opt-in service – you have to choose privacy, it isn't the default. It may even be that Ask.com offers it precisely for that reason – it satisfies the privacy lobby but makes little difference to the amount of data they can gather, the depth of their profiling and the effectiveness of their targeted advertising.

The existence of the AskEraser does show that if the search engines wanted to, they could offer the kind of services that might be approved of by privacy advocates. The question is whether the search engines want to provide these kinds of services – and the evidence so far seems to suggest that they do not.

5.3 A technological bypass?

There have been attempts to find ways to use Google's existing service in a manner that protects our privacy better. There are 'anonymisers' such as *proxify*[109] and TOR[110] which theoretically enable you to surf the internet and

[108] See e.g. http://www.karmasnack.com/about/search-engine-market-share/
[109] http://proxify.com/

to use Google without leaving your identity by routing the search through independent servers. These, though, need to be specially sought out and carefully checked before use – as these only provide limited protection.

Moreover, to help internet users in general, relatively obscure independent solutions are not enough – they've always existed, but only protect those people 'in the know', who are not likely to need protection in any event as they know good practice.[111] What is needed are more mainstream solutions – or the paradigm shift suggested in Chapter One so that privacy becomes the default option, and data gathering and data retention have to be 'opted into', or as the Working Party suggest, to operate only where users have given 'explicit, informed consent'.

Are mainstream technological solutions possible? One way could be if a browser provider built anonymous search into to its standard product. There are many technical problems to providing such a solution – not least the bandwidth and processing power needed to route the searches – as well as massive cost implications, and what might well prove insurmountable political and market pressures not to provide this kind of service, not least from Google whose business plan it would undermine. It is easy to imagine how governments would portray them – 'tools for terrorists' – and how strongly they would oppose them in the current political climate.[112]

5.4 Change from the community?

So, on the face of it, all that leaves is the possibility of change from community – norms, in Lessig's terms,[113] or more directly the kind of community action suggested by network communitarians such as Murray.

---

[110] See https://www.torproject.org/

[111] There are strategies to cut down the risks (e.g. http://www.eff.org/wp/six-tips-protect-your-search-privacy).

[112] This may have already happened at some US universities, where nodes for anonymising services were provided, but rumour suggests that there was pressure to remove them by law enforcement agencies. These rumours remain unsubstantiated – but the nodes have been removed.

[113] See LESSIG, L. 2006. *Code: Version 2.0*, New York, Basic Books.

This is where, as outlined in Chapter One, symbiotic regulation can come into play. Google has shown itself to be responsive to customers' needs and wishes, and if they believe that their users really desire privacy and anonymity they will find a way to provide it, building a business model that incorporates it. Google now earns more revenue outside the US than in the US[114], so pressure from the markets outside the US such as the EU will be likely to have more impact than in the past.

If sufficient community pressure is put on Google, the company might offer solutions itself – perhaps in the form of a 'premium' service for those willing to pay, offering varying degrees of anonymity, or by showing the options directly on their own website. This latter method could allow Google to make the case directly for non-anonymised search, by allowing them to explain to users (in a form of words agreed with the Working Party, for example) why Google would like to retain their search data, and how it can offer each user a better, more personalised search service as well as better, more appropriate advertising – something that people might appreciate. They might even find that by being more open, direct and honest about the way their search works that the public responds positively, and keeps using Google in its current form – or even in a more 'intrusive' form, giving away more personal details – but this time with full, explicit and informed consent.

Following this kind of logic, there may be other ways for the Working Party to proceed, even if it accepts that it may not be possible to win the argument directly or through the legal process. One way could be to say to Google 'you can have your data retention if and only if you provide good access – highlighted on your search pages themselves – to all the data you hold on the user, in a form they can read and understand, giving them a chance to correct or delete if appropriate'.

[114] Google's sales figures for the first quarter of 2008 showed 51% of sales outside the US, and this trend has continued. See for example http://news.bbc.co.uk/1/hi/business/7353677.stm

That could satisfy most of the Working Party's requirements, and if combined with up front information (perhaps as part of a short, blunt, log-in process) most of the rest of the problems could be resolved. The precise form of this information could be agreed between Google and the Working Party. This could benefit both:

- From the Working Party's perspective, the user's rights under data protection are satisfied much more – the right to know, see and understand what data are being stored is a fundamental data protection principle.
- From Google's perspective, it could actually allow their profiles to be improved, and thus allow advertising to be targeted even more accurately.

There are many objections to such a solution – and it could involve large costs to implement, though they would pale into insignificance compared to the near half-billion Euros the EU recently fined Microsoft. Nonetheless, it gives an idea of how some degree of horse-trading might be possible. Indeed, Google has gone some way toward implementing this through its 'Google dashboard',[115] providing access to some of the data that is held about individuals. This idea is looked at in more detail in Chapter Four – where a broader concept, 'Collaborative Consent', is introduced to suggest a way forward.

5.5 The future of Data Retention

The struggle over data retention is far from over – indeed, there are still strong forces working in both directions, as the emphasis on counter-terrorism shows few signs of dissipating and the security challenges to privacy remains both present and powerful, while the privacy lobby seems to have some momentum of its own, as the strength of the Working Party's recent opinion has shown. Public interest and concern over privacy appears

[115] See http://www.google.com/dashboard

to be growing all the time, and if as interest grows so does understanding of the issues, it is likely that there will be a growing level of opposition at least to mandatory, generalised data retention.

Some aspects of the problem have not even been addressed. One of these is the issue of profiles themselves – even when search data etc is deleted, it is possible that profiles derived from that data could be retained and considered 'current' rather than past data, and thus not covered by any requirement to delete data older than a certain age. These could include categorisations or more sophisticated descriptions of individuals, and other kinds of profiling data. How does data protection legislation cover this? At the moment this seems far from clear – but it should, and pressure should be put on regulators to ensure that it does.

At the moment the fight is over data protection and data retention – but perhaps it should go further, and move on to the right to have data about you deleted, and not just if the data are not true. As noted above, there are many reasons to care about the amount of data being gathered and processed. The combination of these and the risks of data vulnerability that will be discussed in Chapter Five make a strong case for a right to delete personal data.

6 Conclusions and rights-based solutions

The first and clearest conclusion from the case study is that the tension between data protection and data retention exists, and that it is an issue that has not yet been resolved. What is more, it is an issue that is of serious concern not only to the Working Party but also potentially to everyone. Currently, data retention applies only to traffic data arising from telephony, email and their equivalents – which in itself is a significant concern, as the Working Party has signalled – but if it were to be extended to cover search data, as Google has suggested that it might, that concern would be much magnified. Further extensions to cover other forms of data could make this

even worse. Such extensions would offer threats to our privacy and our autonomy, and should be opposed.

These threats to autonomy fit directly within the terms set out in Chapter One – they are essentially covert, have a real and increasing effect on both real life and online activities, can be discriminatory and reinforce imbalances of power and are open ended in scope.

The question then is whether the compensating 'goods' that arise from data retention are sufficient to override these concerns over autonomy. The 'goods' arise in three areas; economic benefits to business, security benefits to governments, and 'service' benefits to consumers – the positive aspects of better targeted advertising and providing more appropriate links. The first of these is difficult to argue – there are effective business models that don't rely on this kind of data retention, even if they might not have the potential for quite such astronomical profits. The second, the security question, is hard to answer fully– but the Working Party questioned 'whether the justification for an obligatory and general data retention… …is grounded in crystal-clear evidence,'[116] and offered suggestions of alternative ways to solve these security problems. The third of these goods, 'service' benefits to consumers, is undoubtedly beneficial – but would be just as beneficial if it was done with consultation and consent. On the surface, therefore it seems that the 'goods' are not sufficient to justify this kind of constriction on our autonomy but it is acknowledged that further investigation is needed.

6.1 A Right to Roam with Privacy

The way that search engine providers currently gather, hold and use search data represents a risk to both privacy and autonomy. If a balance between the interests of businesses, governments and individuals is to be found this needs to change. A balanced, rights-based approach is one way that this

[116] Opinion 113 of the Working Party, page 2, downloadable from http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2005/wp113_en.pdf

change could be brought about. Here, that could mean the establishment of a right to roam the internet with privacy.

Many of the problems surrounding search engines could be solved by such a right, a right that could be viewed as springing directly from Article 8 of the ECHR. In essence, it would be a right to roam the internet without data being gathered about you.

This 'right to roam with privacy' would cover all the principle methods of navigating the web – it would, for example, not only cover search data but also clickstream data, whether it is gathered by ISPs or directly from websites monitoring where their visitors have come from and are going to. The right would not be an absolute right, but would be the default position. If a website, ISP or search engine wishes to gather data, they would need to get genuine, informed consent.

This could effectively require the search-engine providers to offer two levels of service – a basic, effectively anonymised one as a default, and a personalised, data-gathering one as an option. The basic one could operate similarly to the AskEraser – with data kept only for hours unless there is a specific reason to hold it for longer, allowing targeted rather than general data retention. The premium service could work as Google does now, but require a registration and log-in process that would be set up clearly enough to satisfy informed consent requirements – this could be designed to allow the search engine to actually gather more information than before, and potentially a business benefit from it, but without compromising privacy and our autonomy.

When applied to ISPs, there could be a similar two level approach – a basic service without data gathering, and a 'super' service with data gathering, but one that has to be opted into. That 'super' service could be a cheaper – but it should be made clear to subscribers why it is cheaper, and what the subscribers are giving up in order to get their discount. It may well be that

the vast majority of subscribers would choose cheapness over privacy – but they should be given that choice.

The right to roam with privacy would work in balance with the interests of security and of economics. An apposite analogy is that of CCTV, something that should not be used universally but can be useful where it is really needed. For security purposes, where information leads the authorities to be concerned about a particular individual, that individual's activities (including search and clickstream data) can be followed. Similarly, if certain websites are of interest, those sites can be monitored and visitors to them to be 'tagged' and tracked. For economic purposes, commercial sites can track people – so long as appropriate, standardised warnings are placed on those sites to alert visitors, the equivalents to the signs saying 'CCTV is used in this area'. Where sites want to gather more than the basic data, they would require registration and login procedures, again with appropriate warnings beforehand.

6.2 The benefits of a rights-based approach

Having established and coherent rights of this kind has a number of benefits. First of all, it could help in the process of harmonising legislation. As the case study showed, where there are conflicts or inconsistencies, businesses will take full advantage. Fleischer managed to bring into his argument not only the conflict between EU data protection and data retention regulations but conflict between European and US regulation, and even with specific rules in Germany.[117] Clear and consistent rights have a better chance of producing clear and harmonised legislation.

Secondly, it could help to ensure that legislation is clearer, unambiguous and more precise. Particularly in such a technical area, legislation needs to be carefully and clearly drafted. The Data Retention Directive was not – it has

[117] See Fleischer letter page 4

holes big enough to drive a juggernaut through, and Google did just that.[118]. Such lack of clarity again allows businesses to take advantage.

Thirdly, it could help regulators to work better together. The fact that Google was able to use the EU's own data retention regulations as what was, in effect, an excuse not to comply with data protection regulations highlights the problem here. Those drafting the data retention regulations should have responded better to the Working Party's opinion on the matter. Again, clear and strong rights can reduce the chances of internal disputes between regulators.

Fourthly, it could help provide a better approach to the crucial and contentious issue of consent – clarified rights to informed consent, with both principles and examples, could provide better protection for individuals and better guidance to the courts. Consent is examined in further depth in Chapter Four.

6.3 Solutions in the real world

How all of this might work in practice will be set out in Chapter Six. The kinds of rights being considered here would operate in balance with the interests of businesses and government – and recognising that businesses and governments can play a positive role. Privacy advocates have a tendency not to recognise this – there seems to be a movement to demonise Google,[119] to present them as the 'bad guys' in every way. Not only is that missing the point (for Google has made using the internet far easier and more productive) but it appears highly unlikely to produce results. Google has created an image of itself as publicly responsible and responsive to consumer wishes – the most likely way to get Google to change is to work with it, not fight against it.

[118] Ibid page 3 for a comprehensive attack on the clarity of the Data Retention Directive
[119] Websites such as http://www.googlizationofeverything.com/ paint Google almost entirely in black, and as the enemy of the consumer.

Similarly, a key lesson to learn from the case study is that regulators do sometimes work for the benefit of their constituents – the Working Party have been bold enough to take on not only one of the biggest players in the internet world, but to stand up to another part of their own system, the data retention lobby, who are backed by much stronger, more powerful and better organised supporters. Privacy advocates should recognise this – and applaud the Working Party for their efforts so far and encourage them to continue to fight for their corner.

The dispute between Google and the Working Party has been a crucial test, but it is one that has still not finished. Will Google react positively and imaginatively and come up with some kind of a radical solution that both protects privacy and enables them to build their business? Will they take what might be called 'the Microsoft approach', and essentially ignore almost everything that the regulator says, and accept the fines when they finally come along? Will those who wish to extend data retention to cover more and more areas rescue them from their dilemma?

The final and most important question for this chapter is whether the solutions that are being suggested have any chance of real success. In particular, in this case, is a right to roam with privacy conceivable from a business perspective? That issue will be examined in depth in Chapters Six and Seven, but as discussed in Section Three of this chapter there are other business models available for search engines, and opportunities to target advertising without retaining identified data. Another appropriate analogy could be that of the banning of smoking in pubs. A few years ago such a ban would have been almost inconceivable, and people running pubs said that their businesses would be unsustainable as a result – but the ban is now in place, and there is little evidence of pubs going out of business. Asking providers for basic, privacy-friendly services alongside their premium, personalised ones would not be as much of an imposition – it would be more like requiring pubs to provide 'no-smoking areas'. It would require adjustments to their business models, but they would be far from

insurmountable. If there is sufficient public interest and political support, it could be possible.

Chapter 4 – Behavioural Targeting and Consent

1 Introduction

Behavioural targeting epitomises some of the key aspects of the Symbiotic Web. In simple terms, behavioural targeters gather data from people browsing the web and use that data to tailor the information that those people receive, usually in order to make a profit. That, together with the fact that behavioural targeting is one of the fastest growing features of the internet,[1] makes their examination particularly useful in the study of the implications and problems associated with the symbiosis. The story of the most controversial of the behavioural targeting systems, Phorm's 'Webwise' is one that illustrates many of these problems dramatically – an examination of that story is central to this chapter.

As shall be shown, Phorm's Webwise has to most intents and purposes failed. The failure and the reasons behind it are salutary lessons in the need for businesses, governments and others to understand symbiotic regulation. Seen through the lens of symbiotic regulation, it was a failure to understand the complexity and nature of the regulatory matrix in which Phorm operated that lay behind many of their mistakes, and in the end led to the failure of their business idea. Ultimately, Phorm failed because it did not understand the nature of the Symbiotic Web. Symbiosis succeeds when both sides of the symbiosis benefits from the relationship – as happens with models like those of Google, Facebook and others. With Webwise, only Phorm stood to benefit – in effect, they were offering a parasitic rather than a symbiotic model. Painfully, almost tortuously, that parasitic model failed – the parasite was rejected and effectively purged from the system. That purging was a painful process, and has had some serious implications, not all of which are positive. Indeed, the whole of behavioural targeting is under threat as a consequence

[1] According to eMarketer.com, spending on behavioural marketing in the US in 2009 was around US$1.1 billion, and predicted to rise to US$4.4 billion by 2012. See http://business-newsarticles.com/business_articles/2011/05/behavioural-targeting-takes-off-for-retailers-215794.htm

of the saga – and a technologically innovative and potentially beneficial stream of business ideas may be lost, delayed or hamstrung as a result. This might have been avoidable if the situation had been better understood.

As will be shown, part of the existing misunderstanding relates to the kind of rights that people think they should have – and the rights that behavioural targeters and others believe that they have. People appeared to believe that they had the kind of privacy rights that would prevent monitoring and tracking of the sort performed by Phorm's Webwise, while Phorm stuck by a legal interpretation of rights, as set out in the Data Protection Act, which they interpreted in their own particular way. This highlights one of the principle problems in the field – and one of the most important reasons to establish clearly understood rights: the gap between detailed law and principled rights is currently a significant one. Accordingly, a new right is set out here: a right to monitor the monitors.

1.1 The crucial role of consent

One of the most important things to come out of this case study is the need for the symbiosis to be a consensual one – not only must both sides of the relationship benefit, but they must understand that benefit, and consent to the process. Phorm and other behavioural targeters, have so far ridden somewhat roughshod over the idea of consent, dealing with it on at best a superficial level, and often effectively avoiding it entirely. If a way can be found for this to be changed, it could not only protect the rights of the individuals and in particular their autonomy, but also provide an environment in which business ideas have a better chance of success. The last part of this chapter suggests a way that this might be achieved. The internet provides hitherto unparalleled opportunities, which, if grasped, could mean a whole new level of consent becomes possible.

A new concept will be introduced in the last section of this chapter, that of 'Collaborative Consent'. Collaborative Consent has two key aspects. Firstly, it

treats consent not as a discrete, one-off decision but as a process, and secondly it looks at consent as a two-way agreement – so that the consenter is allowed and enabled to see what they have consented to, to monitor, modify or withdraw that consent in real time, and where the enterprise seeking the consent must communicate properly with the consenter not just at the start of the process but throughout. Collaborative Consent requires a dialogue between the enterprise and the individual. The internet provides a medium for immediate and interactive communication that allows such a process to be possible – and for the symbiosis of the Symbiotic Web to remain both benign and consensual, this kind of consent is not only apt but the only kind of consent that is appropriate. Moreover, it is a form of consent that would help enable the right to monitor the monitors set out in this chapter.

As this chapter will show, the story of Phorm was fraught with problems, from technical and legal challenges to political, economic and public relations complications, and the ultimate failure of its business should not be viewed even by privacy advocates as a triumph, though in many ways those advocates might see the final result as a victory. The technology that lies behind Webwise is interesting and innovative – and could, if implemented in a positive, consensual way, bring significant benefits to both users and businesses. This chapter will conclude with a look at how this could happen, and, indeed, how this kind of a system could play a crucial role in bringing about a future for the internet a lot closer to the positive models suggested by Berners-Lee and others than might otherwise be the case.

2 Phorm

The story of Phorm demonstrates many of the issues that are of concern in this thesis. In essence, Phorm attempted to implement a system that would monitor and analyse people's entire web-browsing behaviour, using that analysis to profile those people and target them for advertising. The way that it intended to do that monitoring, the extent to which it was done

surreptitiously, and without the meaningful consent of those being monitored, made it highly controversial. As shall be shown by the case study, that controversy and the way that it was managed, eventually brought about the failure of Phorm's business plan in the UK. The manner of that failure, and the roles played by the online community, by other businesses, by privacy advocate groups, by politicians and lawmakers both in the UK and Europe, needs to be considered carefully. The role of the online community and privacy advocates in the process was particularly crucial: they highlighted what Phorm was doing and why it was of concern, they lobbied and brought other groups onto their side, alerted politicians and lawmakers – including those in Europe, eventually leading to the abandonment of Phorm by its erstwhile business allies, and the collapse of Phorm's business model. The story provides a graphic demonstration of how regulation on the internet can function in practice – and how and why Murray's theory of symbiotic regulation can be applied, as suggested throughout this thesis. It also helps to outline the role that rights can play in the process.

2.1 The origins of Phorm

Phorm was controversial from the beginning. The company that became Phorm began as '121Media', in what the BBC describes as "the murky world of adware and spyware".[2] Phorm are at pains to point out that their origins lay in adware rather than spyware[3] though their principal product, PeopleOnPage, was classified as spyware by F-Secure, one of the principle anti-spyware software providers.[4] As Kent Ertugrul, Phorm's founder, put it:

[2] See the BBC technology pages "Phorm: your questions answered", on http://news.bbc.co.uk/1/hi/technology/7283333.stm

[3] Adware is effectively a way to acquire software – a user can download a product 'for free', but must accept advertising as a consequence. Spyware is a more intrusive version, which, generally unbeknownst to the user, installs software onto their system to monitor their activities and serve 'appropriate' advertisements to them, while at the same time often slowing or corrupting their computer systems. For an analysis of the differences between them, see http://www.webopedia.com/DidYouKnow/Internet/2004/spyware.asp

[4] See http://www.f-secure.com/sw-desc/peopleonpage.shtml

> "…what happened was it became very clear to us that there was no distinction in people's minds between adware - which is legitimate - and spyware. So we did something unprecedented which was we turned around to our shareholders and we shut down all our revenues. We weren't sued, we weren't pressed by anyone, we just said "this is not consistent with the company's core objectives."[5]

121Media withdrew all its products and reformed as 'Phorm'[6] and began to develop its 'Webwise' system, in the emerging field of behavioural targeting. Behavioural targeting refers to systems that collect data on web-browsing behaviour – data which might be searches made, sites visited, or more detailed clickstream data such as the time of browsing and so forth – in order to select which advertisements to display. Behavioural targeting in one form or other is already common on the net – amongst others it is already used by Google,[7] Yahoo,[8] AOL[9] and Microsoft[10] as well as by a number of specialist advertising and marketing companies such as AudienceScience,[11] Omniture[12] and Netmining.[13]

The marketing industry are highly enthusiastic about behavioural targeting, suggesting that not only does it work well for advertisers[14] but that it gives customers what they want, 'improving user experiences' and suggesting that audiences actually welcome this kind of activity. Mark Wilmot, writing in *Marketing Daily* in 2009, said "Something amazing happens when marketing

---

[5] From an interview with Kent Ertugrul in *TheRegister*, on http://www.theregister.co.uk/2008/03/07/phorm_interview_burgess_Ertugrul/
[6] http://www.phorm.com/
[7] Most directly through Google AdSense, see https://www.google.com/adsense/static/en_US/AfcOverview.html?sourceid=aso&subid=ww-ww-et-pubsol&medium=link
[8] See http://advertising.yahoo.com/adsolution#product=Behavioral
[9] See http://advertising.aol.com/advertiser-solutions/targeting/behavioral-targeting
[10] See http://advertising.microsoft.com/ad-programs/microsoft-targeting
[11] See http://www.audiencescience.com/
[12] See http://www.omniture.com
[13] See http://www.netmining.com/
[14] A US study commissioned by the 'Network Advertising Initiative (NAI), a coalition of online marketing companies, suggested that in 2009 behaviourally targeted advertising was more than twice as effective than ordinary 'Run of Network' advertising (measured by average conversion rate (6.8% vs 2.8%). See BEALES, H. 2009. The Value of Behavioral Targeting. Network Advertising Initiative.

efforts are actually relevant to people. We see this step as initiating that crucial dialogue. And shoppers, for their part, are replying; essentially giving permission to marketers to learn their habits and respond accordingly".[15] What he means by 'essentially giving permission' reveals a great deal about the way that the advertising industry views the issue of consent – as something that customers do automatically, implicitly, just by participating in their programmes or accepting their services, without any kind of real debate or discussion.

Privacy advocates, unsurprisingly, take a diametrically opposite view, seeing behavioural targeting as a pernicious and potentially dangerous practice.[16] A 2009 study by a group from the University of Pennsylvania and the Berkeley Center for Law & Technology suggested that the American public is closer to the views of the privacy advocates than those of the marketing industry. This research suggests that most are against behavioural targeting – and even more against it when they know the details about how the data concerned is gathered and used. This study suggested that 66% of adults did not want marketers to tailor advertisements to their interests, rising to between 73% and 86% when they were informed about the methods involved.[17] If even the simpler forms of behavioural advertising are contentious, the more sophisticated and more pervasive forms are even more so.

[15] See for example Mark Wilmot, "Put out the welcome mat", in *Marketing Daily*, July 28, 2009, http://www.mediapost.com/publications/?fa=Articles.showArticle&art_aid=110489. As he puts it "Something amazing happens when marketing efforts are actually relevant to people. We see this step as initiating that crucial dialogue. And shoppers, for their part, are replying; essentially giving permission to marketers to learn their habits and respond accordingly".

[16] See for example Privacy International's view on http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-559082 or the Center for Democracy & Technology's guidelines on http://www.cdt.org/privacy/pet/Privacy_Controls_IPWG.pdf

[17] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania., particularly p3

2.2 Webwise

With Webwise, Phorm took behavioural tracking to a new level, not just tracking particular aspects of surfers' web activities, or activities on particular websites or uses of particular web services, but attempting to track their entire web activity - every website visited, every click made, every service used. Moreover, though their initial service (and presumably their business model) was based on targeted advertising, from the start their ambition was greater: as their homepage suggested, they wanted to provide 'a personalised internet', including both advertising and content.

> "…Webwise will automatically start working for you by understanding your interests from the pages you visit, matching them to the content of millions of websites, and providing you with personalised content and relevant advertising."[18]

Webwise can be seen, therefore, as epitomising the Symbiotic Web – gathering personal data, monitoring individuals' behaviour, and attempting to tailor their whole web experience on the basis of that behaviour. As we shall see, the 'content' part of the service was introduced much later, when Phorm already appeared to be in decline, which indicates in part at least that the advertising aspect of the business was the driving force behind it. Nonetheless, Phorm both knew and hoped to benefit from the potential for tailoring content and indeed the whole internet experience.

Achieving this depth of monitoring involved two key things: some very inventive technology and an alliance with cooperative ISPs. A full technical analysis of the technology is beyond the scope of this thesis – though detailed work has been done on the subject, particularly by Richard Clayton of the Computer Laboratory at the University of Cambridge.[19] As Clayton puts it:

---

[18] From http://www.phorm.com/consumers/index.html

[19] First of all in CLAYTON, R. 2008. The Phorm "Webwise" System. http://www.cl.cam.ac.uk/~rnc1/080518-phorm.pdf: Online., and then continuing analyses in Clayton's blog on http://www.lightbluetouchpaper.org/

> "The basic concept behind the Phorm architecture is that they wish to take a copy of the traffic that passes between an end-user and a website. This enables their systems to inspect what requests were made to the website and to determine what content came back from that website. An understanding of the types of websites visited is used to target adverts at particular users."[20]

The way Webwise does this can be broken down into three key phases: monitoring, analysing, and targeting. Issues arise through each of these stages, so it is important to understand them and the distinctions between them.

Phase 1: Monitoring

In the monitoring phase, a user's web activity is monitored, by intercepting the instructions given by users at the ISP level and attempting to filter out what it believes is to be useful from that activity, building up a profile of the individual user. The mechanism through which it performs this monitoring is technically complex, and as shall be shown its legality has been challenged in a number of ways. It should be noted that Webwise does not monitor all internet traffic produced by a user – it very specifically monitors only web traffic, rather than email, file transfer or other non-web traffic. Moreover, it doesn't monitor all web traffic – it makes an attempt to ensure some degree of privacy by ignoring 'known' webmail sites (based on a list maintained by Phorm) and sites that are private enough not to permit search engines to examine them.

This monitoring mechanism involves putting a 'false' cookie onto the user's computer by masquerading as the website which the user wishes to visit – for example, if the user sends out a command to visit www.bbc.co.uk, a computer run by the ISP will pretend to be www.bbc.co.uk, and send back a

[20] Ibid. p2

cookie to the user's PC as though it was www.bbc.co.uk. That cookie will contain an individual identifier – specifying the individual user – that is then used by Phorm to monitor the activities of that user on the www.bbc.co.uk domain. That identifier (known as a UID) is used as the principle way of identifying a user throughout the Webwise process, and is one of the ways in which Phorm intends to maintain privacy – it is a randomly generated number, with no connection to the individual, and as such maintains (in Phorm's opinion) anonymity and privacy. All the information within the Phorm system is linked to this UID and not to any other identifiable key such as an IP address. The efficacy of this anonymity is a complex issue – and a crucial one, Phorm uses it not only as what they believe is a way to ensure that they are not covered by data protection law (since the data they hold is not linked to an individual, just to a UID) but as a way of portraying themselves not just as a 'privacy friendly' company but as a company in the vanguard of the fight *in favour* of privacy rather than as an enemy of it.

At first glance this UID system would appear to offer the kind of anonymity that Phorm suggests, at least from a technical perspective – but this is hotly disputed. Nicholas Bohm, in a legal analysis of the Phorm system that followed Clayton's technical analysis, said that:

> "If parts of the visited site use the HTTPS protocol for secure browsing, the cookie containing the Phorm UID will be sent to the site, where the UID can be read; and if a webmaster wishes to do so, he can read the UID in any case using Javascript. The result is that any site which holds any personally identifying information about a user, and many do, can associate that information with the Phorm UID and indeed also with the user's IP address visible to the site. In view of this, Phorm's claims for the anonymity of its processes are, to put it no higher, a considerable exaggeration."[21]

[21] BOHM, N. 2008. The Phorm "Webwise" System - a Legal Analysis. Foundation for Information Policy Reseach., p2

Moreover, what may be technically possible now is not necessarily an indicator of what may be possible in the future – and the ingenuity and technical skills of hackers and other computer experts should not be underestimated. As Ross Anderson, Professor of Security Engineering at the Cambridge University Computer Lab has said; historically anonymising technology has never worked.[22]

There is another side to the question of anonymity for Phorm – and for all behavioural trackers. That is the interaction with the real world, which can effectively bypass all these attempts at anonymity. Suppose, for example, a web-surfer has browsed for something they would wish to keep private – a relatively harmless example could be one partner secretly researching wedding venues, not wishing their partner to know about it. The Phorm system may not know who that person is, or be able to link their UID to anything identifiable like an IP address, but anyone who *in the real world* sees that person served with a personalised advert while they browse will be able to identify them. The link can be made in the world of atoms, even if it isn't made in the world of bits. What is perhaps even more direct, several people might share the same computer, and not always log in using different accounts – a system like Webwise would mean that they might be served by advertisements targeted at each other, and hence learn things about each other that might compromise their privacy. Phorm may provide technical and legal anonymity – but that anonymity could be fatally flawed in the real world.

Phase 2: Analysis

The analysis phase works by examining the individual web pages visited by a particular user. It operates in a way that is similar to the way that search engines inspect web pages, extracting and examining the words on the page, comparing them with other pages, and attempting through this kind of

[22] Quoted, for example, in the Evening Standard, in March 2008 http://www.thisislondon.co.uk/standard-home/article-23449601-web-users-angry-at-isps-spyware-tie-up.do;jsessionid=D5AA1541C91446314EAD7013363AB159

analysis to determine what kind of a page it is – for search engines, this is then used to determine which search terms should lead to the listing of a page, while for Webwise it is used effectively to determine the profile of a person. The processes are similar, the results converse. There is also one very big difference – when a page is analysed by a search engine, the owner of that page should at least in some senses benefit from that analysis, as their pages will be more easily found by the right kind of people. When a page is analysed by Phorm, the owners of websites whose pages are scanned do not benefit – the prime beneficiaries are Phorm and their clients, who will be able to find potential targets better.[23]

Once the analysis of a page visited has been performed, the Webwise system distils it into a short record: the URL of the page, the top ten words with which it can characterise the page, the search terms the user might have used to find the page at first, and the UID of the user.

Phase 3: Targeting

The targeting phase works on the basis of these short records. They are matched against 'channels' that have been set up by Phorm on instructions from Phorm's advertisers. If this profile matches with the specification of the channel, then that UID becomes eligible for an advertisement from that channel – and a further record is made, simply recording the name of the channel and the UID, and a 'date-stamp' of when the match was made. The earlier record, containing the visited URL, the search terms and the frequent words, is immediately discarded. Then, if the user visits a participating website (one that takes advertisements from Phorm), the system will look up

[23] This raises issues, most directly the recurrent issue of the extent to which the internet can be considered a public or private space. Phorm assumed that if a website permitted search engines to examine it, then it would also consent to Phorm's analysis. As Phorm's representatives told Richard Clayton "we work on the basis that if a site allows spidering of its contents by search engines, then its material is being openly published. Conversely, if the site has disallowed spidering and indexing by search engines, we respect those restrictions…" See Richard Clayton's analysis, in CLAYTON, R. 2008. The Phorm "Webwise" System. http://www.cl.cam.ac.uk/~rnc1/080518-phorm.pdf: Online. There are also parallels with the arguments made in the much discussed case of *eBay v. Bidder's Edge, 100 F.Supp.2d 1058 (N.D. Cal. 2000)*

which channels a UID is matched to, and an 'appropriate' advertisement from one of those channels will appear on that website. Precisely which advertisement appears depends on the rules set up by the website and by the relevant channels, and if more than one advertisement is eligible, by a form of auction, whereby the advertisement that makes Phorm the most money is placed.

Phorm's channels are in effect profiles: categories into which individuals are being put, which are used to determine what kind of content will be delivered to them. These categories are defined and set by Phorm and their clients, not by the individual being categorised. And, more importantly, the classification takes place for the benefit of Phorm and their clients, not for the individual. What makes the nature of this kind of profiling even clearer is that, explicitly, where there is an option, the option chosen is the one that financially benefits Phorm the most. That is the essence of this kind of profiling - and precisely the kind of risk outlined in Chapter Two's discussion of the Symbiotic Web.[24]

2.3 Phorm's Webwise in practice

In many ways Webwise is very similar to more conventional behavioural targeting systems – Phorm's system of channels, for example, is similar to the profiles or 'personas' (sic) used by other systems. The way that advertisements are served to participating websites is effectively identical.

There is, however, one significant difference. As noted above, Webwise monitors and analyses all an individual's activities, not just those on a particular site or system. Google's behavioural targeting, by comparison, works only on data gathered through searches made using Google's search engine and other Google services – so if a user searches with Yahoo,

[24] This method of analysis and targeting is very simple - boiling down a web page into a ten word record, then matching it against an existing set of pre-created channels, rather than designing and creating channels based on an analysis of users' behavioural patterns. It would be false to assume, however, that the simplicity of the analysis is something that would be true either for Phorm as it develops or for future behavioural targeting systems.

Microsoft or ASK instead, that data will neither be available nor be used by Google for their analysis. Similarly, Amazon analyses a customer's activity on the Amazon site, and Facebook the activities on the Facebook site – though Facebook's recent 'Beacon' service extended this through the formation of an alliance of online retailers, allowing each of them to share the data gathered during visits to the others' sites.[25] Webwise takes it to a new level, gathering data not just from a small selection of sites and services, but from all sites and services except those that specifically and actively opt out of the system.

The way that Webwise does this is by working at the ISP level. From a practical perspective it is a system that needs to be deployed by an ISP, so that it can be in a position to intercept all the web-surfer's activities – and from a business perspective, this means that the ISP must cooperate very closely with Phorm. Indeed, the key to the Phorm business model, as it first became apparent, was that Phorm was aiming to work with three of the UK's largest ISPs: BT, TalkTalk and Virgin Media.

From a legal perspective, what does this mean? Nicholas Bohm, of the Foundation for Information Policy Research, in his aforementioned legal analysis of Phorm,[26] suggested that the deployment by an ISP of the Phorm architecture would involve four different forms of illegality, for which the ISP would be primarily liable, and for which Phorm would be liable as an inciter:

1 Interception of communications, an offence contrary to Section 1 of the Regulation of Investigatory Powers Act 2000 (RIPA). This relates to the monitoring phase, where the Phorm architecture, as managed and operated by the ISP, intercepts the instructions sent by the surfer to a website (in the example above, to www.bbc.co.uk) in order to copy them.[27]

[25] Beacon is discussed further below.
[26] BOHM, N. 2008. The Phorm "Webwise" System - a Legal Analysis. Foundation for Information Policy Reseach.
[27] Ibid. pp3-11

2 Fraud, an offence contrary to Section 1 of the Fraud Act 2006. This relates to the way in which the Phorm server masquerades as the target server, in order to make the surfer's web browser accept the Phorm cookie (in the example above, that the Phorm server would falsely represent itself to be the BBC server).[28]

3 The risk of committing civil wrongs actionable at the suit of website owners – Bohm in his analysis gives a number of examples how this might work. He suggests the example of the Bank of England, which like many other websites states categorically in its published privacy policy that it does not "use cookies to collect information about you." When Phorm is in action, it would look to most users that the Bank of England is doing precisely that – though the monitoring cookie would actually have been placed by Phorm, it would look to a user as though it were a Bank of England cookie. The owner of the site might therefore have civil remedies for false implication – or even defamation, or potentially passing off or trademark infringement, where the name in the cookie includes a trademark.[29]

4 Unlawful processing of sensitive personal data, contrary to the Data Protection Act 1998.

These four issues are not just legal issues, but hint at the deeper issues that lie behind not just Phorm but other behavioural targeting systems – and indeed other forms of data gathering on the internet. The first issue, the RIPA issue, concerns the privacy of an individual's actions on the net to start with – whether people want or expect their web browsing, the instructions they put into their browsing software, to be private or not. The second and third

[28] Ibid. pp11-12
[29] Ibid. p16

issues are issues of good faith – when a surfer visits a website, can they expect that their interactions with that website just to be with that website, and not with another, unconnected third party? All three of these issues relate to another, even bigger matter, one that underlies the whole of this thesis, and will be dealt with in more detail in Chapter Seven: the extent to which the internet can be considered a private or a public space.

The last of these legal issues, the data protection issue, is the most complex and perhaps the most important. Browsing activities can be some of the most personal, most intimate, and most sensitive of activities – concerning everything from personal tastes to relationships, jobs and finance, plans for the future, even personal peccadilloes. Is there an expectation that this kind of thing is considered private? From a legal perspective, as outlined by Bohm, there are a number of different ways in which the processing of data by Phorm might be considered illegal.

2.3.1 Phorm and RIPA

Firstly, if the issues relating to RIPA and the Fraud Act above are accepted, then the purpose for which the data is being gathered and processed cannot be legal, and hence the processing itself cannot be legal. Whilst Bohm's arguments appear strong, however, they have not been tested in court and despite lobbying the authorities have not taken action, and the nature of the technology is sufficiently complex for their to be significant doubt as to whether they would succeed.

Phorm themselves sought advice from the Home Office on the RIPA issue. They asked two questions: firstly, do Phorm's actions constitute 'interception of communications' or not, and secondly if they do, is it lawful interception. There are delicate legal issues such as whether a 'person' has intercepted the communications, as required for RIPA to apply according to Section 2(2) of RIPA, or whether the contents of the communications have been 'made available' to that person, as required by Section 2(8) of RIPA. On the second

question, if interception is deemed to have taken place, then it can only be lawful if both the sender and the intended recipient of the communication have consented to that interception. Phorm relied on the idea that surfers have consented to their service in some form (either through the terms and conditions of their ISP, or through some kind of direct consent whose precise form would be determined when their service comes into action) and on the assumption that if a website consents to be spidered for search engine purposes, then they have consented to have communications to them intercepted for Phorm's purposes. As noted above this latter assumption is highly debatable.

The memo that they received in response has become available on the internet, and is far from conclusive on either question, but ultimately suggested that Webwise would be legal if the users gave explicit consent.[30] As shall be shown in Section Two below, however, the nature of the correspondence between Phorm and the Home Office, as well as the advice that was eventually issued, gave rise to significant controversy and played a part in the eventual fall of Phorm.

2.3.2 Phorm and Sensitive Personal Data

More importantly, particularly in the context of autonomy, if all web browsing is intercepted, that web browsing will be likely to include information about the browser that would be classified as 'sensitive personal data' according to Section 2 of the Data Protection Act 1998. This is defined as:

> "…personal data consisting of information as to —
>
> (a) the racial or ethnic origin of the data subject,
>
> (b) his political opinions,
>
> (c) his religious beliefs or other beliefs of a similar nature,

[30] See for example http://cryptome.org/ho-phorm.htm. The memo itself was 'leaked' to the web. See for example http://www.guardian.co.uk/technology/blog/2008/mar/12/homeofficeonphormitslegal

(d) whether he is a member of a trade union (within the meaning of the [1992 c. 52.] Trade Union and Labour Relations (Consolidation) Act 1992),

(e) his physical or mental health or condition,

(f) his sexual life,

(g) the commission or alleged commission by him of any offence, or

(h) any proceedings for any offence committed or alleged to have been committed by him, the disposal of such proceedings or the sentence of any court in such proceedings."[31]

Given that the data may include this 'sensitive personal data', one of the conditions set out in Schedule 3 of the Data Protection Act must be met. In particular, Schedule 3 demands specifically *explicit* consent to the processing, and the processing may only be lawful if it is 'necessary' in the sense that it is needed to protect the 'the vital interests of the data subject or another person',[32] or is necessary to exercise a right or obligation conferred or opposed by law[33] or a specific legal process or in the administration of justice,[34] or (for necessary processing of data for medical purposes) for use health professionals or equivalents,[35] for monitoring the existence or absence of equality of opportunity or treatment between persons of different racial or ethnic origins.[36] There are a few exceptions – where the information has already become public,[37] where the personal data are processed in circumstances specified in an order made by the Secretary of State,[38] to allow not-for-profit groups established specifically for political, philosophical, religious or trade-union purposes to process data relating to its own members, without any disclosure of such data without consent of the data

[31] Data Protection Act 1998, S2
[32] Data Protection Act 1998, Schedule 3, paragraph 3(a)
[33] Data Protection Act 1998, Schedule 3, paragraph 2
[34] Data Protection Act 1998, Schedule 3, paragraphs 6 and 7
[35] Data Protection Act 1998, Schedule 3, paragraph 8
[36] Data Protection Act 1998, Schedule 3, paragraph 9
[37] Data Protection Act 1998, Schedule 3, paragraph 5
[38] Data Protection Act 1998, Schedule 3, paragraph 10

subject.[39] Overall, these conditions are strong – essentially requiring that 'sensitive personal data' should only be processed where necessary, rather than when a processor can benefit for business or financial purposes.

The idea that sensitive personal data should require more stringent conditions – and indeed a great many restrictions – is one that makes a good deal of sense, particularly from the perspective of advocates of autonomy and of human rights. The developing techniques of data aggregation and profiling mean that it needs to be considered much more carefully. According to the rules set out in the DPA, as outlined above, data concerning whether a person suffers from diabetes would be classified as 'sensitive personal data'. Data about whether the subject is a regular purchaser of sugar-free chocolate, or has ordered books about treatment for diabetes would not. Similarly, data about whether a man was a homosexual would be considered to be 'sensitive personal data', but data suggesting that their preferred holiday destinations were San Francisco or Sydney, that they were members of the Barbra Streisand fan club, and that they spent large sums of money on hairdressing would not. None of these facts specifically indicate that the individuals are diabetics or homosexuals respectively – but if profiling is applied, even automatically, the chances of the individuals being classified within categories that consist almost entirely of diabetics or homosexuals respectively would be high. What is more, these examples show only the more obvious and intuitive kinds of connections that could be made, and any kind of 'sensitive' data can be inferred from what appears to be non-sensitive data. With detailed processing and large scale data aggregation, even the most seemingly innocuous data, from sports followed or the kinds of news items read to choice of snacks or time of surfing on the internet can become highly significant. The data themselves are not sensitive personal data but are capable of revealing sensitive personal data.

In the case of Phorm and other behavioural targeters – and for many other data gatherers operating on the internet – though the data they gather may

[39] Data Protection Act 1998, Schedule 3, paragraph 4

not necessarily fall within the precise definitions set out in the DPA, it is very likely that sensitive personal data could be derived from their data. The chances that web-browsing behavioural data would neither contain sensitive data nor sufficient data from which sensitive data could be derived would be very small indeed. That, therefore, could be argued to imply that the conditions set out in Schedule 3 to the DPA should apply – ethically, if not according to the letter of the law. More importantly, what this shows is one of the weaknesses of the category system set out in data protection legislation. The categories of personal and sensitive personal data are becoming blurred and confused – and on top of this the separation between personal and non-personal data is being further confused, as the idea of anonymisation,[40] with Phorm's UID system a prime example, comes under contestation. The data protection categories were developed when data was considered in a much more static form – while data in the current era has a much more dynamic nature, in terms of how it is gathered, processed and held. Data can shift between the three categories depending on time, processing and context. This is an issue that will be discussed in depth in Chapter Six – the implications are significant, particularly in terms of how future data protection or equivalent legislation should be framed.

If the considerations for sensitive personal data are considered to apply to Phorm, the only term that could apply is if specific, explicit consent had been gained. As shall be discussed in detail in the second half of this chapter, this issue of consent is the crux of the whole story. There are many aspects to it directly here – issues of identity, for example, as discussed above. The person who is actually browsing the internet and having their details processed may well not be the person who has entered into a contract with their ISP or the person who has clicked 'OK' at any point where asked by Phorm, if indeed that has happened at all.

That raises another key issue - the question of whether to make a system 'opt-in' or 'opt-out'. How consent for the system would be gained was initially

[40] And de-anonymisation, something that will be looked at in Chapter Five

left to the ISP to decide as and when they implemented it, though the initial impression gained from Phorm was that opt-out might be the norm.[41] After the privacy furore that arose when Phorm's nature began to go public, as detailed below, and after an opinion by the ICO,[42] Phorm eventually decided to insist on an opt-in system.

3 Does any of this matter? Isn't it just about advertising?

Phorm, and other forms of behavioural targeting, are primarily used for advertising, and as a consequence it is possible to consider this issue not to be one of great significance. The regulatory suggestions to date have largely only concerned the advertising industry – either through self-regulation, or through industry specific legislation. That, however, does not take into account the full nature of current uses or the future potential of the techniques and systems that have been developed.

The first point to make is that while targeting is currently about advertising, the techniques developed and systems used will be equally useable for other kinds of targeting. Precisely the same systems used to target people to sell to can be used to target people for scams or other identity related crimes – the more accurately scammers can target their scams, the more likely those scams are to be successful. Similarly, governments and others can use targeting techniques for their own purposes, both appropriate and contentious. Here, as in other areas, both criminals and governments are likely to 'piggyback' on commercial efforts, both copying techniques and actually accessing and using commercial data and systems for their own purposes – something that will be looked at in more depth in Chapter Five when we consider data vulnerability. So though targeting is currently largely concerned with advertising that is not necessarily going to remain the case. Indeed, as Phorm's 'Webwise discover' service indicates, the advertisers

[41] See for example an interview on the BBC at http://news.bbc.co.uk/1/hi/technology/7283333.stm
[42] See http://www.theregister.co.uk/2008/04/09/ico_phorm_tougher/

themselves see advertising only as a starting point, and intend to use their targeting and tailoring techniques much more broadly.

Secondly, it should be noted that advertising on the internet has a fundamental difference from most forms of advertising – it is advertising that can be proved to be successful. Advertisers can and do know whether people have followed their advertisements, and can then follow through to see whether they have made a sale. Moreover, much of the advertising on the net is on a 'pay per click' basis, so that advertisers only pay when potential customers actually click on their advertisements and visit their sites. This in turn means that advertising providers like Google will only place advertisements that people will follow – and because of the enormous scale of Google and others, it means that they have accurate and analysable records of which kind of advertisements work, why, when and on whom. Ultimately, that is likely to mean that advertising not only can but will be more successful and more persuasive, something that in itself needs to be understood.

3.1 Profiling

The profiling used by targeters has very little to do with the kind of psychological or even sociological profiling that people might immediately assume, but instead is a mathematical, analytical system, and as a consequence significantly more accurate. It starts from the kind of thing that an Apple iTunes user might see when they connect their iPod, recommending something by Crosby, Stills and Nash because there's a song by Joni Mitchell in their music library. iTunes doesn't recommend Crosby, Stills and Nash because Apple employs specialists in folk rock to tell them what people who listen to Joni Mitchell might like, but because they've built up a database of millions of users' musical tastes, and it shows that many people who have Joni Mitchell on their system also have Crosby, Stills and Nash on their system. It is a mathematical exercise, not an aesthetic one.

This has a number of key implications. Firstly, it means that for accurate profiling of this kind you need extensive databases – as noted in Chapter Two, this is one of the reasons for the gathering of such vast quantities of data. The data thus gathered are used not only for information about the individual concerned, but to help target others more accurately too. Secondly, it means that profilers need powerful computers, computers capable of working with the immense databases thus compiled. These first two factors together mean that the best and most successful profilers (and targeters) tend to be the largest companies, with the largest number of users and the deepest pockets – companies like Google, Apple, Amazon and so forth. This in turn has regulatory implications – firstly it puts particular emphasis on the need and ability to regulate these biggest of companies, and secondly it means that if ways can be found to ensure that the biggest of companies behave ethically, appropriately and in ways that are positive for the individuals, then significant strides will have been made to solve the overall problems.

The next key implication of this method of profiling and targeting is that it has the potential to be much more accurate than 'traditional' psychological, sociological or 'human expert' profiling. Professor Ian Ayres, in *Supercrunchers: How Anything Can Be Predicted*, gives a series of examples of how this kind of statistical analysis outperforms human experts in even the most surprising and often counter-intuitive fields, from success in sports to assessing the potential quality of a particular year of Bordeaux wine.[43] As the amounts of data and the power of the computers have increased dramatically in recent years, and as the analytical techniques have been developed and refined, these profiling and predictive systems have become more and more accurate. This trend seems certain to continue, and though Phorm's current methods of analysis and targeting are simple, even crude, future systems are likely to be more refined, and potentially far more accurate.

[43] AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted*, London, John Murray., particularly pp1-10

Further to all of this, another level of analysis becomes possible when data are aggregated – and not just similar and apparently closely related data, but seemingly unconnected and dissimilar data – to create even greater data sets and allow much more complex and detailed profiling. For example, if you start with data about musical taste, then add age and location data, and then add to that data about income, data about occupation, education and marital status, the number of further possible inferences that can be made increase exponentially. The sample sizes are sufficiently large that statistical analysis is very powerful, and applying the kind of predictive techniques discussed by Ayres make targeting even more effective.[44]

3.2 Imperfect profiling

This kind of profiling and targeting is by its nature imperfect – particularly for simple analysis systems like that currently used by Phorm – but that from a commercial perspective this does not really matter. An advertiser does not need all its targets to follow its advertisements, or to make a sale to all those who follow its advertisements – it just needs enough of both to find sufficient real customers. For other purposes, however, imperfect targeting can have significant implications – those engaged in counter-criminal activities, for example, need to ensure that they do not round up the innocent with the guilty, while serving an inappropriate advertisement may have implications wider than just failing to make a sale, particularly where products might have sensitivities, such as books about religion or sexuality.

One specific aspect of this kind of imperfection needs particular attention – the problem of imperfect identification. The issue of identity has many implications and complications, but it is particularly important when considering how targeting and profiling are performed through the internet. When data are gathered from one particular computer, how can it be possible to tell the identity of the person using that computer? Is it the same person who generally uses that computer? Where computers are shared, for example

[44] Ibid. pp135-138

in a family household, how can the website visited tell which member of the family is using the computer at a particular time? If there are casual or visiting users, will the data gathered from their use of the computer be merged in with that by the regular user? There are systems on most computers to allow individuals to set up secure, personalised profiles – but are they always used? Can it be appropriate for a website visited or other profiling system to assume that they are being used?

That does seem to be what Phorm were trying to do. In an online interview with the BBC, Phorm's spokesperson replied to a question about 'surprise' advertising appearing – e.g. the aforementioned person researching wedding venues whose partner is then bombarded by advertisements for dresses and rings – that "most people have a separate login if they are sharing a computer…"[45] This is a bold assumption to make – and the consequences of it being wrong could be significant.

3.3 Predictive profiling

Profiling can have further uses – and new uses are being explored all the time. One particular area of concern to those interested in autonomy is the use of profiles to predict people's behaviour, and the use of those predictions to control that behaviour. A prime example of this is also detailed in Ayres: the use by Harrah's casinos of profiling information to effectively manipulate its customers into spending more money.[46] What Harrah's does is try to assess an individual's 'pain point': the amount of money a particular gambler can lose in one session and still enjoy the experience. This pain point is calculated based upon a combination of personal and demographic details and Harrah's own data gathered about all their customers' gambling habits. Harrah's regular customers use a swipe card system to gamble, so Harrah's knows exactly how much money a customer has lost at any time. When that customer is nearing their pain point, Harrah's sends a representative called a

[45] In http://news.bbc.co.uk/1/hi/technology/7283333.stm
[46] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray. pp30-31

'luck ambassador' to interrupt them, offer them a free drink or a meal – and Harrah's also knows what kind of food or drink that customer usually buys when they visit Harrah's casinos. This break cuts off the pain, and leaves the customer refreshed and ready to spend more money in Harrah's after they've had their drink or food.

The Harrah's example is an 'offline' example – but the techniques used give some idea both as to what can become possible with this kind of system and the dangers of businesses having too much information about their customers. Do Harrah's customers know what is happening to them? Did they consent to be monitored in this way, and manipulated in this way, when they signed up for the swipe card system that has gathered all their data? From a legal perspective, almost certainly – the small print on the agreement that they signed will have covered all such eventualities. Whether they really understood that agreement, let alone how Harrah's would be using the data that they had agreed that Harrah's would be able to gather, is another question.

For all these reasons, what has happened with Phorm is much more significant than as though it was 'just' about advertising. The advertising industry has barely scratched the surface of the possibilities that could be developed from behavioural monitoring, targeting and profiling – though even those scratches are thought provoking. It is crucial that lessons are learned.

4 The Rise and Fall of Phorm

Phorm raises a wide range of issues, from the technological nature of its interception and inspection systems and the various technical legal issues highlighted by Bohm's legal analysis to the deeper and less concrete concerns over people's every activity being monitored and exploited for financial gain. What is more interesting is how these various issues have played out as Phorm has attempted to bring its Webwise service into action. Indeed,

whether the legal issues put forward by Bohm and others have technical merit (or would actually succeed in court) does not appear, in practice, to have been as important as the part that their existence *as challenges* has played in what appears to be the ultimate demise of Phorm.

The story reveals a great deal about the reality of regulation in this kind of an area – and the relative strengths and importance of the various interest groups involved, from Phorm and its business allies to the privacy advocacy groups, the hackers and computer-users' community, the various governmental groups at different levels and from different locations – and, just as importantly, the many other businesses who have their own interests in the field. All these different interest groups have played key parts in the process, and if we are to understand how and why Phorm has, so far, failed to deliver its product, we need to understand these parts. Still more importantly, we need to understand it better if we are to find better ways to deal with this kind of an issue in the future.

4.1 A public dispute

The controversy over Phorm has been played out in public, to a great extent, and more particularly and appropriately over the internet itself. Hackers, digital rights and privacy groups reacted strongly from the moment the proposed service became known. The Open Rights Group, perhaps the most respected of these groups, started a 'Stop Phorm' campaign,[47] while Professor Ross Anderson, referring to the three ISPs who were at that stage proposing to implement Phorm's Webwise, said "The message has to be this: if you care about your privacy, do not use BT, Virgin or Talk-Talk as your internet provider"[48]. Tim Berners-Lee told the BBC that he would change his ISP if it introduced a system like Webwise.[49] The technical and legal analyses

[47] http://www.openrightsgroup.org/campaigns/stop-phorm
[48] Quoted in the Evening Standard, 6th March 2008, at http://www.thisislondon.co.uk/standard-home/article-23449601-web-users-angry-at-isps-spyware-tie-up.do;jsessionid=D5AA1541C91446314EAD7013363AB159
[49] See "Web creator rejects net tracking" at http://news.bbc.co.uk/1/hi/technology/7299875.stm

provided by Richard Clayton and Nicholas Bohm respectively, as discussed above, underpinned this campaigning, which was followed in detail by online technical news providers like *The Register*[50] and *Wired*.[51]

One of the most contentious issues, through which much of the initial drama was played out, was the discovery that in 2006 and 2007, prior to the existence of Phorm's Webwise becoming public, BT had carried out 'secret' trials of the system, involving tens of thousands of end-users. These trials were carried out without the consent of the end users, and when their existence became public, through a report leaked onto the internet, there was not just an outcry from privacy groups but the City of London Police met with BT representatives to informally question them about the trials. The City of London Police decided not to pursue a formal investigation, suggesting that there was no criminal intent on behalf of BT, and, crucially, that there was 'implied consent' by the end-users[52]. As noted above, this latter claim is a highly contentious one, and the whole issue of consent is key to the challenge of Phorm, while Bohm suggested that the police claim that there was no criminal intent was simply a misunderstanding of the legal requirements for criminal consent[53]. Nonetheless, no police action followed. Moreover, though nothing specific has materialised from the controversy of the secret trials, the outcry about them caused BT significant embarrassment, provided a powerful weapon for anti-Phorm campaigners, and added to the impression that Phorm itself was somehow 'underhand', secretive and potentially illegal.

4.2 Phorm's defence and government involvement

Phorm's defence to these attacks included a PR campaign that included founder Kent Ertugrul talking directly to the media, including being

[50] See http://www.theregister.co.uk/2008/02/29/phorm_roundup/ for a summary of all the various Phorm stories covered by The Register
[51] See for example http://www.wired.com/epicenter/2009/04/uk-web-spying-f/ for one of the more recent stories.
[52] See for example http://www.theregister.co.uk/2008/09/22/bt_phorm_police_drop/
[53] Also see http://www.theregister.co.uk/2008/09/22/bt_phorm_police_drop/

interviewed by the BBC,[54] *The Guardian,*[55] and *The Register,*[56] as well as attempting to engage directly with the UK Government, specifically to ask the Information Commissioner's Office to confirm that Phorm's UID anonymity system meant that it was compliant with the Data Protection Act. Effectively, Phorm believe that data protection does not apply to their system, as the data they gather, process and use do not constitute 'personal data'. The ICO did, effectively, confirm that this was the case, though it also expressed the view that 'opt-in' consent would be required for any trials and for any eventual rollout of the service, and suggested that they would be continuing to monitor the situation very closely.[57]

This was just one part of the involvement of the UK government in the Phorm controversy. As noted above, when faced by the legal analysis produced and distributed by Bohm, Phorm sought advice from the Home Office concerning RIPA. The issuing of this advice became the centre of another controversy, as emails between the Home Office and Phorm were released that appeared to show that the company had helped to edit this draft legal interpretation of Phorm by the Home Office, in an attempt to ensure that the service would be seen as appropriately 'legal'. Baroness Sue Miller, the Liberal Democrat spokeswoman on Home Affairs, accused the Home Office of 'collusion', calling the exchange of emails 'jaw-dropping', and said that "The fact the Home Office asks the very company they are worried is actually falling outside the laws whether the draft interpretation of the law is correct is completely bizarre."[58] Both the Home Office and Kent Ertugrul vigorously denied this interpretation of the exchange of emails.

Further complications followed. In response to 'several questions from UK citizens and UK Members of the European Parliament', the European

[54] http://www.bbc.co.uk/blogs/ipm/2008/03/phorm_an_interview_with_kent_e.shtml
[55] http://www.guardian.co.uk/technology/blog/2008/mar/06/yourquestionspleaseforkent
[56] http://www.theregister.co.uk/2008/03/07/phorm_interview_burgess_Ertugrul/
[57] See ICO press release, at http://www.ico.gov.uk/upload/documents/pressreleases/2008/new_phorm_statement_040408.pdf and reporting at http://www.theregister.co.uk/2008/04/09/ico_phorm_tougher/
[58] See http://news.bbc.co.uk/1/hi/technology/8021661.stm

Commission inquired into how the UK government had responded to the complaints about Phorm by users.[59] EU Telecoms Commissioner Viviane Reding sent a letter to the UK Government – this time to the Department for Business, Enterprise and Regulatory Reform ('BERR') – asking for an explanation as to how Phorm's technology conformed with EU data protection and privacy laws. BERR replied, after a delay, providing an explanation whose key points depended on Phorm's UID-based anonymity, together with a confirmation of the requirement in the Home Office memo that explicit consent would be required.[60] Phorm, therefore, in BERR's opinion, complied with EU privacy law.

This view was immediately challenged by the Open Rights Group and others, in part based upon Bohm's legal analysis. They noted specifically the requirement for both sides of a communication to need to consent to an interception of a communication – so not only did web surfers need to consent, but website owners, and stressed once more the inadequacy of Phorm's UID-based anonymity. As a result of this and other responses, the EC inquiry concluded that if the UK believed that Phorm complied with UK privacy law, then that law must not be a correct implementation of the relevant EU directives. After much communication with the UK Government, in April 2009 the Commission launched an action against the UK government, calling for changes in UK law. In the words of Viviane Reding.

"We have been following the Phorm case for some time and have concluded that there are problems in the way the UK has implemented parts of EU rules on the confidentiality of communications." [61]

---

[59] http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/570&format=HTML&aged=0&language=EN&guiLanguage=en

[60] BERR's reply to Commissioner Reding was not made public, but BERR did disclose to The Register the key points, which were then published on the internet at http://www.theregister.co.uk/2008/09/16/phorm_eu_berr/

[61] See again http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/570&format=HTML&aged=0&language=EN&guiLanguage=en

This action has yet to be concluded. Both Phorm and the ISBA ('the voice of British advertising') tried to dissuade the EC from continuing their action,[62] while the Open Rights Group and others actively supported it, and publicised the existence of the action through the media. In October 2010 the Commission confirmed that as the UK Government has not changed the law as the Commission suggested, they would be taking the UK to the European Court of Justice to force it to do so.[63] That action is still ongoing.

4.3 A rancorous dispute

During all these disputes, Phorm has been portraying itself as a 'privacy friendly' company, suggesting that rather than being a threat to privacy, Phorm would be providing something that was positive for privacy. Webwise, according to Phorm, meant that you could have the targeted advertising that people wanted without the need for gathering or holding personal data. Phorm engaged a specialist consultancy service, 80/20 Thinking, to perform a 'Privacy Impact Assessment' on the service. That assessment appeared largely positive. As 80/20's Simon Davies put it to the BBC: "We were impressed with the effort that had been put into minimising the collection of personal information." The Privacy Impact Assessment was subsequently used by Phorm to demonstrate their 'privacy-friendly' credentials. This was not without issues, however. Simon Davies, as well as being CEO of 80/20 Thinking, is a noted privacy advocate and one of the founding members of Privacy International – and Kent Ertugul tried to suggest that 80/20's positive assessment of the Phorm system meant that Privacy International had endorsed Phorm, something that he later had to retract.[64]

[62] See for example http://www.isba.org.uk/isba/news/657
[63] See http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1215&format=HTML&aged=0&language=EN&guiLanguage=en
[64] In a live webchat on Phorm's own blog. See http://www.webwise.com/how-it-works/transcript_080311.html

The disputes between privacy advocates and Phorm became increasingly rancorous as the affair wore on. A number of 'anti-Phorm' websites appeared such as Badphorm,[65] Dephormation[66] and the Anti-Phorm League.[67] In response to some of the more vociferous of anti-Phorm campaigners Phorm set up their own campaigning site, Stopphoulplay.com. Phorm were forced to admit to 'overzealous' editing of their Wikipedia entry, after having deleted sections critical of Phorm and links to some further stories.[68] In the words of BBC correspondent Darren Waters, "This is a battle with no sign of a ceasefire, with both sides settling down to a war of attrition, and with governments, both in the UK and the EU, drawn into the crossfire."[69]

This, then, was the far from simple background. Legal challenges, technical disputes, serial campaigning, possible police action, EU action against the UK government, smear campaigns and propaganda, and all the while Phorm attempting to get its business into action. The result of it all began to become clear in 2009. Though before that stage it had looked as though Phorm was likely to succeed, with the UK government apparently firmly behind it, three of the biggest ISP's planning to use its service, an endorsement of sorts from noted privacy advocates and a guarded approval from the Information Commissioner's Office. Then business reality began to kick in, as other businesses and other government departments began to respond seriously to the furore generated by the whole affair.

In April 2009, Amazon.com announced that it would not allow Phorm to scan any of its domains.[70] Others followed, including the Nationwide Building Society.[71] Then, the hammer blow fell when BT announced that it would not be implementing Phorm – followed immediately by Talk-Talk, and then Virgin Media. Phorm's shares fell 40% on the announcement, and it looked as

[65] http://www.badphorm.co.uk
[66] https://www.dephormation.org.uk/index.php
[67] http://www.antiphormleague.com/index.php
[68] See http://www.theregister.co.uk/2008/04/08/phorm_censors_wikipedia/
[69] http://www.bbc.co.uk/blogs/technology/2009/04/phorm_hoping_to_stop_phoul_pla.html
[70] See http://news.bbc.co.uk/1/hi/technology/7999635.stm
[71] See http://www.guardian.co.uk/business/marketforceslive/2009/jul/21/phorm

though Phorm's business model was in danger of total collapse. Phorm responded by bringing out new products, notably its content-tailoring service, 'Webwise discover', which "will allow visitors to any website to automatically find content within that site based on their interests from across the web", but as this will also depend on an ISP choosing to implement Webwise's system, and none are currently planning to, it is hard to see what future there is in it at this stage. Then, in August 2009, the Office of Fair Trading announced that it was investigating the use of personal information in internet advertising, questioning the use of tailored advertising and the possibility of tailored prices based on personal information.[72] Phorm's share price fell once more, this time more than 20%, as a result of the announcement of that investigation.[73] The All Party Parliamentary Communications Group ("apComms") has also undertaken its own inquiry into internet traffic, covering amongst other things, behavioural advertising – the report was issued in October 2009.[74] This report came out with strong conclusions, including the recommendation that:

> "…the Government review the existing legislation applying to behavioural advertising, and bring forward new rules as needed, to ensure that these systems are only operated on an explicit, informed, opt-in basis."[75]

In September 2009 potentially the final blow fell with the resignation of Phorm's Chief Technology Officer, Stratis Scelparis.[76]

---

[72] See http://www.guardian.co.uk/media/2009/aug/20/internet-targeted-advertising-oft-investigation
[73] See http://www.guardian.co.uk/media/2009/aug/20/advertising-digital-media
[74] See http://www.apcomms.org.uk/uploads/apComms_Final_Report.pdf
[75] Ibid. p21
[76] See http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/media/6209787/Phorm-loses-technology-chief.html

5 Phorm and Regulation

The Phorm saga has had an impact on more than just the company itself. The aforementioned investigations by the OFT and apComms are just part of the fall out. Phorm was the focus, but it drew attention to the whole of behavioural advertising, adding controversy to a developing practice, leading to serious contemplation of regulation on both sides of the Atlantic. The European regulatory fall-out detailed below arose directly from Phorm, as outlined in 4.2 above.

The route to the contemplation of regulation in the US was a little less direct – it was a reaction to the whole practice of behavioural targeting rather than just to Phorm. Nonetheless, Phorm played an important part, at least from the perspective of the lobbyists who have supported the idea. As Jeff Chester, of the US-based Center for Democratic Technology, put it in May 2008: "This is such an important story… In the UK, there's been a huge firestorm over Phorm. But there's been close to nothing here."[77] That 'firestorm' catalysed the pressure groups into action, and eventually the US authorities began to consider action.

5.1 Regulation in the US: Do Not Track

In September 2009 a coalition of privacy and consumer rights groups (including the aforementioned Center for Democratic Technology) wrote an open letter to the House Committee on Energy and Commerce calling for the regulation of behavioural advertising.[78] A 'comprehensive privacy bill' was drafted by Congressman Rick Boucher, who headed the House Energy and Commerce Subcommittee on Communications Technology and the Internet,

[77] Quoted in The Register at http://www.theregister.co.uk/2008/05/16/congress_questions_nebuad/

[78] See http://arstechnica.com/tech-policy/news/2009/09/privacy-advocates-want-regulation-of-behavioral-advertising.ars for a discussion of the issues, and http://www.uspirg.org/uploads/Lh/2Y/Lh2Y_vpDJ2A5maDU214SFw/WaxmanBartonLetterSEPT091.pdf for the letter itself.

and was put before Congress in 2010.[79] Another bill, the 'Best Practices Act' was introduced by Congressman Bobby Rush in August 2010.[80] Both attempt to legislate to control behavioural advertising – and both have been strenuously opposed by the advertising industry.[81] Neither bill has yet passed through either House, nor shows any sign of doing so.

A further bill, the 'Do Not Track Online' bill was introduced into the Senate in May 2011.[82] This Bill takes a slightly different approach – one that might mean that it has more of a chance of success. It can be seen as part of a bigger initiative, the 'do not track' initiative, which involves not just law-makers but the internet industry – and most directly, the makers of browser software.[83] The idea, in essence, is to allow users to control whether they allow behavioural targeting through their browser settings. Behavioural advertisers would be expected to design their advertising systems to look for these settings and then follow them.

The makers of some of the most used browser software are working in collaboration with this plan. In particular, Microsoft and Mozilla have introduced versions of Internet Explorer[84] and Firefox[85] respectively which implement this, and Apple have announced that the next version of Safari will

[79] See for example http://thehill.com/blogs/hillicon-valley/technology/97685-boucher-privacy-bill-will-not-inhibit-advertising-new-draft-due-in-june
[80] The 'Building Effective Strategies To Promote Responsibility Accountability Choice Transparency Innovation Consumer Expectations and Safeguards Act' or 'BEST PRACTICES Act', available online at http://www.house.gov/apps/list/press/il01_rush/h_r_5777_the_best_practices_act_2010.pdf For analysis, see for example http://www.loeb.com/congressionalcommitteeholdshearingonboucherprivacyproposalandnewhr5777/
[81] See for example 'Tech firms warn privacy bill will harm economy' on CNET, at http://news.cnet.com/8301-31921_3-20011435-281.html
[82] The short title of the Bill is the 'Do-Not-Track Online Act of 2011', and it is available online at http://commerce.senate.gov/public/?a=Files.Serve&File_id=85b45cce-63b3-4241-99f1-0bc57c5c1cff
[83] See http://donottrack.us/
[84] See for example http://online.wsj.com/article/SB10001424052748703363904576200981919667762.html
[85] See http://blog.mozilla.com/blog/2011/02/08/mozilla-firefox-4-beta-now-including-do-not-track-capabilities/

do likewise.[86] Google is going a slightly different route for its Chrome browser – and perhaps more importantly, is part of the opposition to California's version of the Do Not Track law, which is further down the line than the national law.[87] Together with Facebook and a number of other key industry representatives, they have suggested that the Bill would 'create an unnecessary, unenforceable and unconstitutional regulatory burden on internet commerce.'[88]

From the perspective of privacy and autonomy, however, there are flaws in 'do not track' that work in the opposite direction. The first is that they are 'opt-out' rather than 'opt-in' – an issue that will be discussed in depth in section 6 below. The second, and perhaps the most fundamental flaw of all, is that they rely on the goodwill of the trackers in order to function.[89] These limitations are fairly fundamental – but even so, the strength of feeling behind the initiatives and the willingness of both lawmakers and browser-providers to engage with the issue is something that suggests hope for the future.

5.2 Regulation in Europe: the 'Cookies Directive'

In Europe in September 2009 Meglena Kuneva, the consumer affairs Commissioner, told a gathering of ISPs, major websites and advertising firms that they were violating "basic consumer rights in terms of transparency, control and risk", through data collection and behavioural targeting.[90] Following this, in October 2009, came the adoption of an EU directive that has become known as the 'Cookies Directive'.

---

[86] See for example http://online.wsj.com/article/SB10001424052748703551304576261272308358858.html
[87] See for example http://www.theregister.co.uk/2011/05/05/google_backs_do_not_track_opposition/
[88] The letter in opposition to the bill can be found at http://static.arstechnica.com/oppositionletter.pdf
[89] See for example http://www.pcworld.com/article/217556/donottrack_in_chrome_and_firefox_different_approaches_same_fatal_flaw.html#tk.mod_rel
[90] Reported in http://www.theregister.co.uk/2009/03/31/kuneva_behavioural/

This piece of legislation has such far-reaching implications that Struan Robertson, the editor of OUT-LAW.COM and a respected blogger in the field, has called it 'breathtakingly stupid'.[91] It modifies existing European legislation[92] to effectively require that any cookie can only be stored on a user's computer, or accessed from that computer, with that user's explicit, informed consent. This would cover not just such things as advertising but any kind of web analytics – indeed the functioning of most modern websites, as cookies are used by most websites in one form or another for example to store user preferences. The Directive was due to come into force in all 27 member-states of the EU by 26th May 2011. In the UK, it has been enacted as the 'Privacy and Electronic Communications (EC Directive) (Amendment) Regulations 2011'.[93]

On the face of it, the Directive appears to be a massive sledgehammer to crack a not particularly large nut – and also a hammer blow to the beneficial aspects of the Symbiotic Web. It would indeed offer some protection to individuals, but at a potentially huge price. As soon as the Directive was agreed, bloggers like those on OUTLAW.COM started looking for ways to mitigate that price – hoping for incomplete implementation of the directive into national law, or for 'business-friendly' interpretations of its terms by local regulatory bodies – or even suggesting that businesses might be better off simply ignoring the law and accepting any penalties that arise as a result.[94] When respected lawyers like those of Pinsent Masons who provide OUTLAW.COM to suggest such an approach it gives an indication of quite how serious the implications of this Directive initially appeared be.

[91] http://www.out-law.com/page-10510
[92] The Directive, labelled PE-CONS 3674/09, modifies Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (the ePrivacy Directive) and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws. It is available at http://register.consilium.europa.eu/pdf/en/09/st03/st03674.en09.pdf
[93] Available online at http://www.legislation.gov.uk/uksi/2011/1208/contents/made
[94] http://www.out-law.com/page-10510

Advertising bodies, initially alarmed by the Directive, soon suggested that 'there was 'nothing to fear' from the new laws, as the preamble suggested that compliance could be through browser settings[95] - an opinion that the ICO seemed to agree with in July 2010.[96] By March 2011, however, after some prompting from the Article 29 Working Party, who unequivocally stated that the Directive requires an opt-in system and cannot be complied with as the advertisers suggest, through simple modification of browser settings,[97] the ICO's tone was changing somewhat. The same month, in a speech at the ICO's annual Data Protection Officer conference, the Information Commissioner suggested that 'UK businesses must 'wake up' to new EU law on cookies'.[98] The ICO suggested that businesses were not doing enough to prepare for the new regulations – something that caused some distress in the industry, as many had assumed from the ICO's initially softer line that the Directive wouldn't really make much difference. At a session organised by the Society for Computers and Law on 'Privacy by Design', the ICO's representative came under sustained attack for the inconsistency of the advice, by lawyers and industry representatives alarmed by the new, harsher interpretation of the law.[99]

In May 2011, the ICO issued guidance as to how businesses can comply with the new law.[100] This guidance appears on the face of it to tread a bit of a middle path – it follows the Article 29 Working Party's stance on browser settings, but in a slightly milder form, and tacitly admits that the ICO is not clear how this will really work in practice, saying that 'for now' they are

[95] See for example http://www.research-live.com/news/government/iab-europe-says-nothing-to-fear-from-eu-cookie-rules/4001534.article
[96] See for example http://www.research-live.com/news/government/ico-urges-easy-opt-out-of-online-tracking-but-eu-demands-opt-in/4003097.article
[97] Based on their Opinion 2/2010 from June 2010, WP171, and in particular section 4.1.1. The opinion is available online at http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf
[98] The speech was summarised in a news release at http://www.ico.gov.uk/~/media/documents/pressreleases/2011/data_protection_officer_conference_news_release_08032011.ashx
[99] The meeting was on 16th March 2011, attended by the author. See http://www.scl.org/site.aspx?i=ne19845
[100] Available online at http://www.ico.gov.uk/for_organisations/privacy_and_electronic_communications/~/media/documents/library/Privacy_and_electronic/Practical_application/advice_on_the_new_cookies_regulations.ashx

'advising organisations which use cookies or other means of storing information on a user's equipment that they have to gain consent some other way.' It then lists the potential ways that consent could be gained, from pop-ups and terms and conditions to what they describe as settings- and features-led consent, which build the consent system into the operations of the website itself, so that a user 'consents' to the cookie when they change a particular setting or choose a particular feature.

Many of these ideas are interesting suggestions. Indeed, they could be seen as making small steps towards the concept of 'Collaborative Consent' that is introduced in Section 6.6 of this chapter, showing an imaginative approach to the ways in which 'real' consent might be achieved, though without the key elements of continuity and collaboration. However, the overriding result of the guidance appears to be a lack of clarity. The ICO gives people a great many options – and the tone of the guidance shows how the ICO themselves do not know where this is going. They state that they 'will be keeping the situation under review', and that there will be a 'phased approach to the implementation of these changes'. The ICO has said that they will be issuing separate guidance on how they intend to enforce the regulations – at the time of writing, this guidance has yet to appear, but the relatively conciliatory tone of the guidance does not suggest that enforcement will be strict or harsh.

In the short term, however, the ICO's guidance did little to dampen industry fears – and by the end of May the government, in the shape of the Department for Culture, Media and Sport, took the unusual step of issuing an 'open letter' to try to reassure the industry. The letter, from Minister Ed Vaizey, tries to clarify the situation and emphasise that the UK's implementation will be 'light touch' and 'business friendly'. Vaizey say that the letter 'makes clear beyond any reasonable doubt how in the view of the

UK Government the regulation should be treated'.[101] Whether the industry views it in the same terms has yet to be seen.

5.3 Uncertainty and the future

The ICO's lack of clarity over the issue reflects the confusion that the Directive has produced – most of the industry appears still to be far from clear what needs to be done, and indeed whether it really can be done. That confusion appears to be reflected throughout the European Union. By the required date of May 26th 2011 only three members of the EU had implemented it – the UK, Estonia and Denmark.[102] The law looks difficult to comply with, and there appear to be very few in the industry that support it, or have yet made serious attempts to comply with it – and yet it addresses a real concern about privacy, which is why it got through the European Parliament. The industry is unhappy about what is being proposed – but the businesses themselves had done nothing to address this real concern. If they had – and if they had understood or taken sufficient account of people's concerns in the area – then perhaps this kind of law, a law they dislike, might not have come into existence.

The story is similar to that seen in the US, where there is a real gap between what the industry wants and what the lawmakers seem to be proposing. There are two key differences between the approaches. Firstly, in the US, the law has not yet been enacted, and may well not come into action for a long time, if at all. Secondly, in the US there has at least been an attempt (and a partially successful one) to get the industry on board. In both cases the principal result is uncertainty – and uncertainty is a real problem for any business, particularly in times of financial difficulty.

---

[101] 'Open letter on the UK implementation of Article 5(3) of the e-Privacy Directive on cookies', p5. Downloadable from http://www.dcms.gov.uk/images/publications/cookies_open_letter.pdf

[102] See for example http://www.cio.com/article/682917/EU_Countries_Ignore_New_Law_on_Internet_Privacy?taxonomyId=3089

This uncertainty arises in part at least from a problem in trust: as a result of the practices of Phorm and others, many people have lost trust in the advertising industry over such things as behavioural advertising. The advertisers, too, have lost trust in the ICO and the European regulators. That gap in trust, and the uncertainty that follows it, cannot be good for the beneficial aspects of the web symbiosis.

5.4 Symbiotic regulation as best practice?

Whether the effective failure of Phorm is an individual incident or representative of an overall movement has yet to be seen – but the regulatory crackdown does suggest the latter. Why is this happening? And what, if anything should be done about it? Murray's theory of symbiotic regulation can help to provide some of the answers to these questions.

Looked at through the lens of symbiotic regulation, it can be argued that what is happening to Phorm is happening to a great extent because Phorm has failed to fully understand the complexity of the regulatory matrix. From this perspective, Phorm appears to an excellent example of how symbiotic regulation really works, and why, if a good regulatory result is to be achieved, it needs to be harnessed. The kinds of rights that are being put forward in this thesis – in particular, the right to roam the internet with privacy, as introduced in Chapter Three, and expanded and extended below to include a 'right not to be monitored', together with better developed and implemented consent rights which will be set out in this chapter – are intended to strengthen the ability of individuals to engage in this symbiotic regulation. As will be set out below, well-constructed and clearly expressed rights would not simply assist the individuals – they could also help both businesses and governments to avoid getting into the kinds of messes that have enveloped both Phorm and the UK government as a result of this affair.

The regulatory matrix in which Phorm operates is complex. As the story related above has shown, many of the different relationships within it have

had their impact: Phorm's relationship to their customers, Phorm's relationships with their business allies – and with their competitors, all the various different parts of the UK Government's relationships both with Phorm, and with the people, the hackers and the advocacy groups' relationships with people, with other businesses, with the UK government - and with the EU and the EU's relationship with the UK. Finally, as the culmination of all these things, other businesses' relationships with their customers – for that, in the end, must have been what caused BT and the other ISPs to withdraw, and hence for Phorm to fall.

It appears that Phorm took too simplistic a view of the regulatory environment, relying on its ability to lobby and negotiate with government, to form alliances with businesses, and to provide an inventive technological solution. They looked to find solutions that met the letter of the law – or at least could be argued to meet with the letter of the law, for the arguments that Phorm put forward about compliance with data protection law have some substance to them, enough to convince the ICO, the Home Office and BERR to give Phorm their support and approval. It is a similar misunderstanding to that made by Kent Ertugrul when he tried to make the distinction between legal adware and illegal spyware – then, as for Phorm, he did not understand sufficiently that what people understood and felt was more important than the letter of the law. The public does not like adware, even if it's legal, and sees very little difference between it and spyware. Effectively, Kent Ertugrul was saying that the public was 'wrong' not to distinguish between the two, but that he would have to bow to their 'wrong view' in abandoning his system. He did not appear to accept that the public might actually be right – in the sense that they didn't like adware because it interfered with what they considered to be their rights. The letter of the law was not what the public cared about, rather what they thought to be right.

5.4.1 The role of the community

Compliance with the letter of the law, moreover, is not enough when norms and markets come into play. Phorm drastically underestimated the feelings of the community with regard to privacy – as the 2009 study of American attitudes to behavioural advertising has shown dramatically[103] – and the power of the community to influence other parts of the regulatory matrix. Ultimately, through the various advocacy groups, through public campaigning, and through the EU, the community managed to get its view across. Phorm became perceived as 'anti-privacy' and this perception gathered momentum, regardless of Phorm's efforts to portray itself as a privacy-friendly company.

Whether these perceptions actually lay behind the key events in Phorm's ultimate downfall – BT's withdrawal; the refusal of websites like Amazon.com to be scanned – is questionable, for in both cases there were other factors involved. In Amazon's case, letting Phorm scan their website could potentially have robbed them of some of their crucial competitive advantages, as their ability to monitor their own enormous customer base is a key part of their business model, while for BT, it might simply have been a matter of not wanting to throw good money after bad. In Amazon's case, however, the fact that they chose to talk about the privacy issues as a part of their reasoning was very revealing. BT did not mention privacy – in fact, they said very little except that they were no longer pursuing Phorm as an option. For BT admitting that privacy issues played a part in their decision would have been very difficult - suggesting that they had at best initially misjudged the situation, and at worst broken the law in their earlier trials of the service, but the adverse publicity and overall image of Phorm cannot have helped the cause of BT's continued participation.

[103] See TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania.

5.4.2 Facebook's Beacon and other services

This analysis of Phorm's failure is made with the benefit of hindsight, but the story of Facebook's Beacon, which has a number of similarities to the saga of Phorm, adds weight to the suggestion. Through Beacon, Facebook shared data with an alliance of online retailers, allowing each to use the other's information about individuals in order to better target advertising and services.[104] Beacon was contentious to say the least, and just as for Phorm the public outcry was vociferous. Facebook's initial response was to change the way Beacon operated – primarily to change it from 'opt out' to 'opt in' – but ultimately Facebook was forced to abandon the system completely, after settling a class-action law suit that had been brought in California accusing not only Facebook but a number of its allied retailers of breaching various US wiretapping and privacy laws.[105] Just as in the case of Phorm, community reaction was strong enough to bring about the end of a service that went beyond what people thought was right. Furthermore, just as in Phorm's case, it was through the manipulation of all the various relationships in the regulatory matrix – relationships between individuals and Facebook, between individuals and governments, through the use of the law, through working with businesses – that this result was brought about. Just as in the case of Phorm, a lot of the trouble could have been avoided if Facebook had been more aware of both public opinion and of the ability of the public to bring that opinion to bear.

Some other services – Google StreetView being perhaps the best example – have produced somewhat similar reactions from privacy advocates, and in some ways appear even more intrusive and yet have not suffered the same

[104] For an analysis of how Beacon worked, see http://www.facebook.com/beacon/faq.php,
[105] Facebook eventually abandoned Beacon on 21st September 2009. For an examination of the class action suit, see http://www.wired.com/threatlevel/2008/08/facebook-beacon/ , and for the suit itself see http://www.wired.com/images_blogs/threatlevel/files/facebook_beacon_complaint0812081.pdf . It is worthy of note that the settlement, currently pending approval in the U.S. District Court of the Northern District of California, includes Facebook's putting US$9.5 million "to create a foundation to fund products that promote online privacy, safety and security". See http://www.concurringopinions.com/archives/2009/09/facebook-settles-beacon-lawsuit.html

fate as Phorm and Beacon, at least within the UK.[106] The reasons for this are not simple – but in symbiotic regulation terms, the regulatory matrices in which they operate at different. From the start, Google has both a stronger base position and a better reputation with the public, and, it appears, a better grasp of how to get the community on its side. Moreover, StreetView offers a service that is both useful and attractive to users – a benefit, in exchange for the intrusion, and an example of how the symbiosis of the Symbiotic Web can function. Even though StreetView has had a lot of very adverse publicity, particular in relation to the revelation that cars taking photos for use by StreetView had been 'scraping' data from private Wi-Fi networks, something that will be discussed in Chapter Five, that adverse publicity has not dampened public enthusiasm for the use of the service.

6 Ways forward and rights-based Solutions

The Phorm affair has caused the UK government considerable problems. It faces a lawsuit from the EU and accusations of collusion with what is perceived to be a 'dodgy' business, and being portrayed itself as riding roughshod over people's privacy and rights – and all of this to back a business idea that has ultimately ended in failure. If it had had a better idea of the likely outcome, and a better understanding of what it was that mattered to people – why, in the end, people were sufficiently distressed by Phorm to bring about its downfall – then the government could have avoided the whole farrago. That, again, would be an advantage of clearly expressed and understood rights, rights that are more closely aligned with what people really want and need than with the current strict legal form.

[106] Action has been taken against Google StreetView in other countries – Switzerland is one example, see http://www.techradar.com/news/internet/swiss-take-legal-action-over-google-street-view-650241, while in Japan there are other concerns about the misuse of images – see http://www.searchenginejournal.com/google-street-view-in-japan-faces-various-complaints/13048/ This, however, emphasises the point that the regulatory matrix is different in different regimes – and Google StreetView appears, in general, to have been accepted in most countries.

Further, it is not just Phorm and the UK government who have found themselves in difficulties, but the whole of the online advertising industry. They're facing a regulatory crackdown not only in Europe but potentially in the US as well – a crackdown that could potentially damage their entire business models. That crackdown has yet to fully materialise, but at the very least they are faced with the need for some serious lobbying – and at a time when finances are being stretched to breaking point for many, that is a distraction and a drain on resources that they can little afford. There are many who have most of their eggs in the behavioural targeting basket, and if the eventual result of the Phorm farrago is that this basket is broken, their businesses could break with it. That, too, could potentially have been avoided if the situation had been better understood – and if the rights that people believed themselves to have had been properly appreciated.

6.1 The right to monitor the monitors

Behavioural advertising – and Phorm's Webwise in particular – highlights the need for rights governing surveillance, monitoring and tracking. The reaction to Phorm, and the successful resistance to its implementation of Webwise, suggests that people feel very strongly about this – and indeed that they have a right not to be monitored.

Phorm made much of its system of anonymisation, and of the fact that it did not hold data, that of the vast bulk of the data gathered were immediately deleted, and that its system was therefore compliant with data protection legislation. One key question is whether even if their system of anonymisation had worked perfectly, even if the data had been immediately deleted, would monitoring of an individual's internet activities then have been acceptable? Data protection legislation covers gathering of data – should similar protection be provided against monitoring, even if data isn't subsequently stored?

This crystallises one of the problems with data protection legislation – though it purports to be privacy legislation, it focuses on data rather than on privacy. Privacy should be concerned with whether we have the right not to be watched, not just with whether those who watch us should be able to record and store the images of what they see. Having our internet activities monitored amounts to having our private actions watched. As Berners-Lee put it:

> "To allow someone to snoop on your internet traffic is to allow them to put a television camera in your room, except it will tell them a whole lot more about you than the television camera".[107]

Moreover, it is not just privacy advocates and internet experts like Berners-Lee who are concerned about this kind of monitoring. In the 2009 survey of American attitudes to behavioural advertising, 68% of adults stated that they would definitely not allow tailored advertising that results from 'following the websites you visit and the content you look at' even in a manner that keeps them anonymous, with a further 19% saying that they probably would not allow it.[108]

Taking this into account, along with the strength of the reaction to Phorm, there is a strong case for a right not to be monitored. This would be an extension of the 'right to roam the internet with privacy' set out in Chapter Three. It would have similar limitations to those discussed in Chapter Three – allowing for example security services to monitor if they have demonstrated that they have sufficient reason to monitor particular people or particular websites, and have obtained appropriate authorisation.

In this thesis, however, it is suggested that this kind of a right should be taken a step further. Not only do people have a right not to be monitored, but when

[107] To a meeting at Parliament organised by Baroness Miller, as reported in The Register at http://www.theregister.co.uk/2009/03/11/phorm_berners_lee_westminster/
[108] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania. P16

they are monitored and accept that monitoring, they have a right to monitor those who are doing the monitoring, and control how and when that monitoring takes place. In this context, 'monitoring' has to be carefully defined. It would need to include such things as intercepting, inspecting, viewing, analysing, comparing, categorizing or otherwise examining or processing information that may be linked to an individual or their activities, even where records of or data arising from that monitoring are not gathered or held, or where the data are immediately anonymised. Definitions would need to be sufficiently technologically neutral to ensure that the rights would still apply as technology develops – so as to avoid the kind of 'side-stepping' of legislation attempted by Phorm's system claimed in relation to data protection legislation.

6.2 Consent rights

If people have the right not to be monitored, when, then, can such monitoring occur? And when can systems like behavioural targeting come into action? The answer is simple: when the people being monitored and targeted consent to that monitoring and targeting. The rights connected with consent are crucial, not only in the cases examined in this chapter, but in the examples developed throughout this thesis – from the search engines discussed in Chapter Three to many of the case studies in data vulnerability that will be looked at in Chapter Five.

A solution to the problem of how to achieve a form of consent that both allows innovative online businesses to be developed and satisfies people's essential understanding of 'true' consent is suggested here: *Collaborative Consent*. This is intended to provide better-expressed, better-understood and better-implemented system of consent, more appropriate for this kind of situation, and taking better account of both the capabilities and the real uses of the internet.

What Phorm's Webwise (and to a lesser extent Facebook's Beacon and other equivalent systems) do are technologically innovative and potentially useful both for the company involved and the individuals who use it – but there are clear problems concerning the impact it has on the individuals' privacy and autonomy. In the same way that the overall challenge for regulators and technologists is to find a way to keep the beneficial symbiosis of the Symbiotic Web in balance, the most important question arising from the Phorm saga is whether a way can be found to harness and support these kinds of technological innovation and these kinds of potential benefit in a way that does not impact upon privacy and autonomy in too detrimental a fashion. The key to the solution is consent: if a way can be found for the individuals concerned to have a real, autonomous and consensual choice as to whether and how to participate in programmes like these, all sides could benefit.

Consent is a more complex issue than it seems – it is not simply a matter of getting a user's consent before doing something. The Data Protection Directive talks about 'express, informed consent' – and all three of those terms need to be considered carefully. Moreover, there are some things that it should simply not be possible to consent to – where it may be deemed that if someone appears to consent to it, then either they cannot have been properly informed, or that something must have prevented them from making an appropriate decision, whether it is duress, lack of mental capacity or something similarly significant.

It is crucial that the issue of consent is understood better and engaged with directly. At present, it is an issue that is often sidestepped or treated in such way as to make is mere legal form rather than having any real connection with what would be understood in any 'real world' sense as 'consent'. On the internet, and indeed when dealing with computer software in general, the kind of consent generally gained is by a user scrolling down a long page of writing that they do not read and then clicking 'OK' at the end to confirm that they have 'read and understood' the terms and conditions. The information

thus presented (but rarely read[109]) is deemed to make the consent 'informed', while the clicking of OK is deemed to make it 'express'. This 'click-wrap consent' has been generally found to be legally acceptable[110] – but if autonomy is taken at all seriously it is close to meaningless. The kind of 'browse-wrap' consent used by Google and others is less legally compelling than click-wrap consent, as the Article 29 Working Party has suggested. As discussed in Chapter Three where search engines are concerned, ordinary, anonymous users cannot be considered to have given consent and the 'de facto contractual relationship' when using a search engine in its usual form 'does not meet the strict limitation of necessity as required in the Directive'.[111] Moreover, it is even less understood by users, and hence of more concern in respect of autonomy.

Many internet providers stretch the consent issue even further, setting their terms and conditions so that by 'signing in' to one service, a user consents to having his or her data gathered and aggregated for other services provided by the same company. If a user signs in to Gmail, for example, and then subsequently uses any of the other Google services (from Google Search to Google Maps etc), then data are gathered about what is searched for or any places examined in Google Maps, and aggregated with the data record of the individual signed in to Gmail. As Google puts it in its privacy policy:

> "We may combine the information that you submit under your account with information from other Google services or third parties, in order to provide you with a better experience and to improve the

[109] A dramatic example of this was demonstrated by the April Fool's joke played by the games company Gamestation who changed their terms and conditions to include a term through which customers gave the company their immortal souls. Nearly 90% of them did so, even though they were given a gift of a £5 voucher if they chose to opt out. See for example http://blogs.telegraph.co.uk/technology/shanerichmond/100004946/gamestation-collects-customers-souls-in-april-fools-gag/

[110] For example in *ProCD v. Zeidenberg*, 86 F.3d 1447 (7th Cir. 1996), *Specht v. Netscape Communications Corp.*, 150 F.Supp.2d 585 (S.D.N.Y. 2001) and *Hotmail Corp. v. Van$ Money Pie*, No. 98-20064, 1998 WL 388389 (N.D. Cal. Apr. 16, 1998)

[111] See Working Party Opinion 148, p17

> quality of our services. For certain services, we may give you the opportunity to opt out of combining such information."[112]

Under current law, this appears to be legal. Google, however, do not define precisely which services it is talking about – and not all 'Google services' are labelled with the Google name, YouTube.com being the most obvious example. If you are signed in to Gmail, does that mean that you have consented to having your YouTube.com activities monitored and aggregated with your Gmail data? YouTube.com has separate accounts that can be individually signed in, but will data still be aggregated? A further question is whether people understand the linkages between the different services, even if they all bear the same labels – are they aware that by signing in to Gmail they are giving Google the legal green light to gather data from all of their services? It does not seem very likely, except for the most 'savvy' of surfers. What is true of Google is equally true of the other big internet companies such as Yahoo and Microsoft, all of whom have a raft of different services with the capability to monitor and gather different kinds of information – and, under current law, seemingly entirely legally if consent has been given through sign in to just one of their services.

The main issue, therefore, is not what kind of consent is currently legal, but the more fundamental issues and how we should consider using law to make sure that legal consent more closely resembles 'real' consent, in the sense that it relates to having made an informed, autonomous decision.

6.3 When is consent required, and when can consent be assumed?

The question of when consent is required is, again, both a legal question and an ethical one. From a legal perspective, the determining factor is generally whether assuming consent is 'reasonable'. In RIPA interception of communications can be lawful if there are 'reasonable grounds' to assume

[112] http://www.google.co.uk/privacypolicy.html ,

the communicators' consent.[113] The question, therefore, becomes what is 'reasonable'? At the very least, it is clearly unreasonable to assume that someone will consent to things that most members of a society reject. Where there is doubt, further questions must be asked and further information sought before consent may be assumed – and while doubt exists, consent cannot be assumed, particularly in a business context.

There is one kind of exception to this, but it is an argument that is both highly contentious and could only apply in very unusual circumstances. It can be argued that there are situations where the needs of society as a whole outweigh the rights of individuals to choose. One specific examples is the suggestion that organ donation should be opt-out rather than opt in – could society's need for organs for medical purposes outweigh the need for each individual to make an active choice, when that choice would only come into play once the individual is dead, and when the most common reason to refuse donating one's organs is a strong, positive and often religious objection? Another is the idea that vaccination against certain diseases should be compulsory rather than optional, where that kind of vaccination only functions properly if the vast majority of people are vaccinated. For something like this to apply, the benefit to society must be overwhelmingly clear, and also something that society in some way consents to as a whole. In a democratic society that might be through the decisions of government, or through obligations under international treaties – there are things that even if the majority of individual members of a society would like them, international obligations might prevent. One of the clearest examples in the UK would be the use of the death penalty.

Is the benefit to society of the existence of a slickly functional and profitable internet business sector so clear and overwhelming that it should override the individual's right to choose? There may be some who would argue this, but it is an argument that is very hard to sustain if privacy is taken at all

[113] RIPA Section 3(1)

seriously, and moreover, a decision like this would need to be made at the highest political level.

6.4 Opt-in or Opt-out?

As well as determining whether consent can be assumed or needs to be individually sought, 'Societal consent' should determine whether a system or service needs 'opt-in' or 'opt-out' consent – if it is 'normal' in a society for something to be acceptable, then using an opt-out consent system might be acceptable, but if it is normal in society for something to be unacceptable, then an 'opt-in' system is crucial. As before, where there is any doubt about whether something is acceptable or unacceptable, the rights of the individual should get the benefit of the doubt, and only an 'opt-in' system should be possible.

This kind of an approach attempts to tread a middle path between the needs of business and the demands of privacy advocates. Some privacy advocates suggest that only opt-in systems could ever be acceptable – Privacy International, for example, in response to Phorm's suggestion that they had endorsed Webwise, stated that they wouldn't support any opt-out system.[114] That, however, runs the risk of going down the path of the kind of legislation recently passed by the EU, requiring all cookies to be specifically consented to – legislation which, as noted, legal experts have called 'breathtakingly stupid'.[115] In terms of the Symbiotic Web, it would mean breaking the symbiosis rather than trying to maintain the positive balance. It is a delicate balance to keep – and as noted, the benefit of any doubt should always go to privacy, and hence to 'opt-in'.

Behavioural advertisers appear to have a very different understanding of the underlying level of consent to their systems – what level of societal consent exists in their context. The behavioural targeting systems of the main players

[114] As reported in the interview with Kent Ertugrul in the Guardian, available at: http://www.guardian.co.uk/technology/blog/2008/mar/06/yourquestionspleaseforkent
[115] http://www.out-law.com/page-10510

in the internet world, Google,[116] Yahoo[117] and Microsoft,[118] all work on an opt-out basis. When Phorm first mooted their system, they left the question of opt-in/opt-out up to the ISPs, seemingly not considering it that important a question and leaving the impression to many that they thought opt-out was probably the most likely solution. The actions of all of them suggest that they believe that, in general, behavioural tracking is not just acceptable, but in fact would be supported by users – while the views of privacy advocates, backed up by the 2009 report on behavioural advertising, suggest the opposite very strongly.[119] Indeed, that survey makes it clear that society as a whole, at least in America, does not assume that behavioural tracking, and the advertising associated with it, would be, is general acceptable. The survey covers attitudes in America, but in the absence of similarly convincing studies of attitudes in Europe or the UK, at the very least the opposite – that society accepts and supports behavioural tracking – cannot be assumed. The logical consequence, therefore, is that opt-in rather than opt-out systems are at least currently a necessity.

Google's Global Privacy Counsel, Peter Fleischer, speaking at the Computers, Privacy and Data Protection Conference in Brussels in January 2010, suggested that the question of 'opt-out, opt-in' is a bit of a red herring, for two reasons. Firstly, because even opting in is often not very meaningful, as people just scroll and click, without understanding, and secondly, because you cannot expect one company (in his case Google) to take the opt-in route unilaterally, as it would be shooting itself in the foot. The second objection would be easily by-passed by a legal requirement for opt-in rather than opt-out. The first objection is a much more important one – but the consequence of it is not that the idea of opt-in should be abandoned, but that a way needs

[116] Google AdSense, https://www.google.com/adsense/static/en_US/AfcOverview.html?sourceid=aso&subid=ww-ww-et-pubsol&medium=link
[117] http://advertising.yahoo.com/adsolution#product=Behavioral
[118] http://advertising.microsoft.com/ad-programs/microsoft-targeting
[119] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania.

to be found for opting-in (and indeed all forms of consent) to become more meaningful. If that way can be found, then the objection just melts away.

In its 'analysis' phase, Phorm made strong assumptions about the public nature of the internet – essentially that by allowing their websites to be searched, website owners are giving anyone freedom to examine, analyse and potentially make profits from those websites. By assuming that no specific consent would be required for this, Phorm were assuming that society – in this case 'web society', including both individual web-surfers and those who provide websites – have consented to this on an overall level, and hence do not need to 'opt in' to their system. Has web society done this? It is a big assumption to make, and one not currently supported by convincing evidence. If behavioural trackers – whether for advertising systems or for other systems – wish to make their services opt-out rather than opt-in, then they should need to provide convincing evidence that this is what society in general supports.

The first step towards getting this kind of acceptance of behavioural tracking could be to promote a better understanding of the positive aspects of the symbiotic relationship of the Symbiotic Web. That, however, would place a duty on the commercial enterprises to be honest about how and why they gather data. At present they appear to wish to short-cut the process, to assume consent before those they are asking to consent have even begun to understand what they are consenting to – perhaps for fear that if the 'consenters' do understand what is going on, they will withdraw their consent, as evidence from cases like Phorm and Beacon, and surveys like the 2009 University of Pennsylvania study[120] suggest they might.

6.5 Informed Consent

One of the strongest principles of the Data Protection Directive is the requirement for informed consent. That raises an immediate question – what

[120] Ibid.

does it mean for consent to be 'informed'? There are two very different ways to look at it – does 'informed' just mean that information has to be given, or does it mean that an 'informed decision' needs to be enabled, a decision where the information has not only been given but has been understood, and that understanding has been confirmed. The former, where information is given, is what generally happens on the internet – the information that users scroll down without reading before clicking 'OK' can be said to have been given, but it is rarely read, let alone understood. The latter, where information is not only given but understood, and a genuinely informed decision is enabled, is what anyone interested in autonomy would demand.

How can this kind of an 'informed' decision be enabled? In the field of medical law the concept of informed consent has been investigated and discussed in depth. Harvey Teff introduced a concept he called 'collaborative autonomy' to find a way through the maze of ethical and medical problems surrounding the need for and meaning of 'informed consent'. In it, Teff suggests a process of communication, a dialogue, through which more complex issues are discussed until they are understood, and as the situation develops and the patient's understanding and views develop, the decision as to whether to continue with treatment or change direction can be made in a manner that is both better informed and more flexible. As he puts it:

> "What many patients seek is sufficient understanding to reach an 'informed' decision in the fuller sense of the term; this can seldom be achieved without the kind of dialogue, and the kind of relationship, to which the collaborative model of medical practice alone aspires."[121]

Though the issues involved in medical consent are somewhat different to those on the internet, there are many similarities, and a similarly collaborative model is possible. The idea that consent should be a dialogue, a process rather than a one-off decision based on fixed, provided information,

[121]TEFF, H. 1994. *Reasonable care : legal perspectives on the doctor-patient relationship,* Oxford, New York, Clarendon Press ; Oxford University Press., p198

is something of particular relevance in cases like Phorm, and can be taken a step further – for what is being consented to is a continuing process rather than a single discrete event, and that places particular demands on consent. Moreover, the internet itself is a communications medium, and one that lends itself ideally to communicative processes. When the model of the Symbiotic Web is considered, that suggestion becomes even more emphatic. For a beneficial symbiosis, both sides must benefit – and that means that what is required is active collaboration between the users and the enterprises gathering the data.

6.6 Collaborative Consent

In the internet context, therefore, consent should be looked at in a collaborative way. The unparalleled communications opportunities presented by the internet can be harnessed to produce a different kind of consent – a form of consent that allows informed decisions, and a real opportunity for those decisions to be expressed. Using the communicative potential of the internet can allow consent to become much more of a collaboration between those gathering data or monitoring users and the users themselves – and one of the cornerstones of a collaborative form of the Symbiotic Web.

The starting point is to ensure that contracts, Terms and Conditions, End User Licence Agreements and so forth are written in plain, understandable language, using agreed and standardised terms for certain forms of activity, explaining technical language like Deep Packet Inspection in terms that explain the impact in a way that ordinary users might be able to understand. These contracts would be designed as much to inform the user – to communicate – as they would be to satisfy legal obligations. This subject will be dealt with in detail in Chapter Six. The work of the creators of Copyleft,[122] and more directly and recently the code of practice for Privacy Notices issued

[122] See http://www.gnu.org/copyleft/ - copyleft licenses are designed with standardised terms, and are intended to be readable and useable by non-lawyers.

by the Information Commissioner's Office[123] give some clues as to how this kind of thing might work. Both the privacy notices themselves and the code of practice concerning them are intended to be communicative, and to begin the process of using the communications opportunities of the internet in a positive way. As the ICO puts it:

> "It's a lot easier to actively communicate a privacy notice in an online context than in a 'bricks and mortar one. You should make full use of the technology available to you to promote transparency and fairness."[124]

Collaborative Consent would take this a step further, not just using the online context to communicate such things as privacy notices, but to include the whole consent process. The provider of a service would engage in a direct dialogue with the user, telling that user all the relevant information as it happens, alerting the user to important changes as they happen, and needing to get direct responses before taking any action. This dialogue would be supported by further, back-up information – in particular, information about what data is being gathered (and has been gathered) and how it is being used – taking the Data Protection 'access' principle into the interactive age.

Google has already begun providing some of this kind of information to those users perspicacious enough to search for it; their *dashboard* system allows Google account holders to see what data have been gathered about them from which Google services,[125] while their *Google Ads Preferences* allow users to see some of how Google has used this data in terms of what 'interest categories' Google has placed them in for advertising purposes.[126] Collaborative Consent would also allow for interactive discussions where

[123] Downloadable from http://www.ico.gov.uk/for_organisations/topic_specific_guides/privacy_notices.aspx
[124] ICO Privacy notices code of practice p9
[125] The Google dashboard is accessed through the Google Privacy Center (http://www.google.com/privacy.html). For a brief discussion of how it works, see http://googlesystem.blogspot.com/2009/11/google-dashboard.html
[126] See http://www.google.com/ads/preferences

possible, and would give the user opportunities to withdraw and modify their consent at any opportunity – allowing, as Google Ads Preferences do, users to modify their profiles, enable or disable the receipt of targeted advertising, decide whether or how their data might be shared and so forth. If Google can do this, why not other data gatherers? And if Google can make the information available indirectly, through a set of links via their privacy centre or their advertising system, why should they not make it available immediately and directly?

Collaborative Consent could be a key enabler for the 'right to monitor the monitors' set out earlier in this chapter. When looked at in the context of the Symbiotic Web, the principal features of this kind of an approach – most directly that it should provide a regular reminder that monitoring is taking place, and give the user the option to withdraw consent – provide a strong step towards supporting the continuation and development of the positive aspects of the symbiosis. If those who are monitoring and targeting people require continued consent from those being monitored and targeted, then they will need to communicate the benefits that those being monitored and targeted are getting. In order to communicate that benefit, they first need to ensure that a benefit really exists – and hence that the symbiosis is a beneficial rather than parasitical one. That, viewed from this perspective, was the problem with Phorm's Webwise.

6.7 A future for behavioural tracking?

There are lessons to learn from this analysis for all forms of consent on the internet – from the unsuitability of browse-wrap or click-wrap consent in many cases where it is currently considered both legal and appropriate to the idea of plain-English contract with standardised terms. This is something that will be considered in Chapter Six, where the full implications of the kinds of rights being considered here will be explored in depth, and in Chapter Seven, which will look at the future of the internet as a whole.

Phorm in its current form appears to have failed – but there are potential advantages to the kind of technology that it has developed, advantages that should not (and almost certainly will not) be lost. There are ways that they can be used without sacrificing autonomy, ways that can put their benefits into the hands of the users rather than just into the hands of Phorm itself and its advertisers.

One way it could work would be to recast Phorm as a profiling service - you would volunteer to be profiled, to allow Phorm to monitor your web browsing and use its extensive databases to provide a content tailoring service. That profile could be put under your own control - used for tailoring content not just to Phorm's benefit, but for your own. This might be what Phorm was suggesting (albeit unwillingly) with their subsequent 'Discover' service, which was intended to allow users to decide when to use the service to scour websites to find relevant content. If Phorm believed their own publicity, turning a tailoring service this way around could be very popular.

What is more, it could be part of a reshaping of the internet, helping to bring about the kind of positive model of Web 3.0 suggested by Berners-Lee. Berners-Lee's 'intelligent agents', the key to the semantic web, could be programmed using behavioural tracking. A surfer could switch on their behavioural tracking system, surf the net, then turn the tracking off and examine the profile that the tracking system has produced, before handing that profile to the intelligent agent and sending it on its way. While the profiling methods currently employed by Phorm, as described above, are relatively crude, the potential is there, and as the algorithms are improved and data are aggregated they are likely to become more accurate. Ayres suggests that computerised, automatic profiling, based on mass statistics and well-developed algorithms tends to be far more accurate than those produced by individuals themselves – as Ayres has shown, people are often remarkably bad at making estimates and at knowing their own

preferences.[127] This way, the advantages of the technology could be brought to bear without threatening privacy or autonomy – and the more positive future that the beneficial aspects of the Symbiotic Web could bring might become possible.

[127] See AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray., Chapter 5

Chapter 5 – Data Vulnerability and the Right to Delete

1 Introduction

On 20 November 2007, the Chancellor of the Exchequer stood up in the House of Commons to make an announcement: Her Majesty's Revenue and Customs had lost two computer discs. These disks were being sent to the National Audit Office as a part of that Office's normal compliance responsibilities, but they had never arrived. What made this event of such significance was what those discs contained. As the Chancellor put it:

> "The missing information contains details of all child benefit recipients: records of 25 million individuals and 7.25 million families. Those records include the recipient and their children's names, addresses and dates of birth, child benefit numbers, national insurance numbers and, where relevant, bank or building society account details."[1]

The Shadow Chancellor put the whole thing into context:

> "Let us be clear about the scale of this catastrophic mistake: the names, addresses and the dates of birth of every child in the country are sitting on two computer discs that are apparently lost in the post; and the bank account details and national insurance numbers of 10 million parents, guardians and carers have gone missing."[2]

The fall out from the loss of these two discs was significant – amongst other things, the Chairman of HMRC, Paul Gray, resigned even before the Chancellor made his official announcement. Data loss, something that before this announcement would have been considered obscure and unimportant,

[1] See Hansard, 20 Nov 2007, online at http://www.publications.parliament.uk/pa/cm200708/cmhansrd/cm071120/debtext/71120-0004.htm#07112058000527
[2] Ibid.

worthy of a brief mention on the financial or business pages of a newspaper, became front page news, and not just for a day or two. Even more than two years later, as smaller scale data breaches come to the public's attention they continue to hit the headlines.

The question of why this particular event had such an impact upon the public has no simple answer. It may have been a matter of scale – the numbers of people involved were huge, and most people must have either been directly included in the lost data or knew someone who was. It may have been a matter of timing – this data breach emerged at just the time that people had reached a critical level of understanding and concern about the dangers surrounding data loss. It may have been related to the nature of the data – people are acutely concerned about financial data and about risks to their bank accounts. It may have been related to the fact that the data was lost by the government – and in particular the HMRC – and it should have been the kind of data that was kept the most carefully, the most securely. If the HMRC can't be trusted to keep data secure, who can? It may even have been connected to the fact that the data dealt with children, and that people are particularly sensitive about protecting their children. To date, the discs have not been recovered – and it is not known whether any damage was actually done as a result of their loss.

In practice it was probably a combination of all these factors and more. The revelation of the HMRC disc loss brought many other data losses to light, revealed partly by the more concerted efforts of civil servants and the Information Commissioner's Office and partly by the focus brought to bear by the media and politicians of all colours. In the year that followed these revelations about the HMRC disc loss the number of data breaches reported to the ICO 'soared' (the word used by the ICO) to 277, including '80 breaches by the private sector, 75 within the NHS and other health bodies, 28 by

central government, 26 by local authorities, and 47 by the rest of the public sector'.[3]

Why did it seem to matter so much? Ultimately, it is a question of autonomy. If people's most personal information can be so easily lost, and potentially put into the hands of criminals or others who could or would wish to use it against them, people feel in danger. If their data are vulnerable, the people themselves are vulnerable. If their data are threatened, people themselves feel threatened. As is being shown throughout this thesis, the use (and misuse) of data can result in direct threats to autonomy – but it is perhaps equally important to understand that there is a *feeling* of a threat to autonomy that is of great importance too. If the problems are to be addressed, they must be addressed at both levels – people must both *have* more control over their data and they must *feel* that they have this control.

The nature of the threat to autonomy presented through the vulnerability of data is qualitatively different from the majority of the threats discussed in this thesis so far. In most of those cases, there was at least some degree of intentionality behind the gathering and use (or misuse) of the data, certainly in terms of the data gatherer. Search engines intend to gather data from those who search – and those who search put their search terms into the relevant box intentionally, and intend the search engine to use that data to present them with appropriate possibilities. When someone browses a website, they do in general intend to look at the site – and could be expected to understand that the owners of that site may know that someone has visited, even if they don't know what else might be told and what might be done with that information. When data vulnerability is examined, that intent is missing. What happens to the data was not, at least in general, envisaged either by the person who initially gathered the information or by those who are the subjects of the data. This represents a new level of lack of control – and a new set of potential risks, including risks to autonomy.

[3] ICO press release dated 29 October 2008, downloadable from http://www.ico.gov.uk/upload/documents/pressreleases/2008/data_breaches_29_october_2008.pdf

1.1 Data minimisation and the right to delete

One of the most important things to understand – as will be demonstrated in this chapter – is that ultimately, where data exist, they are vulnerable, in a wide variety of different ways. This, in turn, can provide one of the keys to regaining control and autonomy – for if data do not exist, they cannot be vulnerable. If a way can be found to take more control over what data actually exists – to make the data protection principle of 'data minimisation' become something that has a real impact – that could make a significant difference to the vulnerability of data.

This chapter sets out some of the ways in which this may begin to happen. The first step is a change in the paradigm that governs the retention of data: those who hold data should have to justify their holding rather than wait to be challenged by those about whom the data is held. The corollary to this would be the establishment and realisation of a 'right to delete' personal data. Through this change in paradigm and the establishment of this right, businesses could be encouraged to develop new business models, models that do not depend so much on the holding of vast quantities of personal data for such long periods.

This idea of a right to delete is subtly but importantly different from the idea of a right to be forgotten, a right currently under discussion by European Regulators for inclusion in the forthcoming revision of the Data Protection Directive.[4] Quite how such a right might work out in practice was not entirely clear when the idea was first mooted – but as shall be discussed in Section Four, the connotations of the name of the right are of concern, and have been subject to criticism. A right to be forgotten looks like the rewriting or erasing of history, or a kind of censorship. The right to delete is about the control of data, not about censorship – and is not in conflict with freedom of expression

[4] Viviane Reding, Vice-President of the European Commission, responsible for Justice, Fundamental Rights and Citizenship Privacy matters, has been promoting the idea – for example in a speech in November 2010, available at http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/10/700&format=HTML&aged=0&language=EN&guiLanguage=en

or the freedom of the press. It should not be seen as a way to rewrite or conceal history or as a tool for celebrities or politicians; it is rather a basic and pragmatic right that should be available to all.

1.2 A theoretical and pragmatic right

At a broad-brush level, the interest theory of rights that underlies this thesis as set out in Chapter One supports the assertion of such a right: the issue of data vulnerability indicate that we have a strong interest in keeping our data secure, and hence a strong interest in deleting data that concerns us. That interest implies the existence of a right of some sort – and this chapter attempts to set out how that right might be.

The second reason is a more basic one – that we have a claim to such a right because we *believe* that we *should* have it. That is a bold assertion, and more work would need to be done to fully endorse it, but it does have support from one of the very few empirical studies in the field to date: the 2009 University of Pennsylvania's study of attitudes to behavioural advertising referred to in Chapter Four. In that study, 92% of those surveyed believed that there should be a law that requires "websites and advertising companies to delete all stored information about an individual, if requested to do so" [5] - the single strongest finding of the entire study. Drawing firm conclusions from surveys such as this is a perilous business, but a finding like this does at least indicate that it is something that needs to be considered seriously.

There are good reasons to believe that it is an appropriate right to assert. One of the most important of these comes down to the deeper question of to whom the data really belongs. If it is in any real sense 'ours', then when we allow others to gather and hold that data, we trust the people to whom we hand it over to deal with it responsibly – but we still retain some rights over it. One approach that has been suggested, most notably in the US, is to

[5] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania. P20

consider a 'property rights' approach to personal data: to grant individuals some kind of property right in their personal data that would 'enable individuals to bargain over which personal data to reveal to which firms for what purposes'.[6] That approach, offered in part in an attempt to use a market-based solution rather than the strict legal control approach of European data protection, has been analysed in some depth in the academic literature. A particularly pertinent analysis comes in Samuelson's 1999 article 'Privacy as Intellectual Property?'[7] which suggests a form of 'licensing approach' rather than a direct form of property right.

There is something to be said for these arguments – certainly the idea that individuals should be able to bargain with those who gather and use their data is attractive. However, making this workable appears very difficult, as it could very quickly become highly complex, and as for consent, as discussed in Chapter Four, all that is likely to happen in practice is that people will scroll through whatever rights they are being offered and just click 'OK' to whatever is being suggested. One of the most important aspects of any system to give individuals control or more autonomy is that it should be simple to understand and simple to use. Complex legal licensing systems or property rights, however legally precise or superficially attractive they might appear, are highly unlikely to be simple in practice.

That is one of the strengths of a 'right to delete' approach. It is simple and direct, and makes it clear what is important – and it shifts the balance of power in any subsequent 'bargaining' process over rights over data and its use. Property rights fail to capture the distinctive nature of the relationship between a person and their personal data. Personal data are not like other property – something owned, something separate from the individual,

[6] Discussed in SAMUELSON, P. 2000. Privacy as Intellectual Property *Stanford Law Review,* 52**,** 1125-1175. Samuelson goes through the arguments made elsewhere by people such as Kenneth C Laudon, Patricia Mell and Richard S Murphy and others, advocating a property right in personal data, before outlining her own 'licensing' approach.
[7] Ibid.

something that can be bought and sold on the marketplace.[8] Instead, personal data can be looked at as part of what might be described as an 'extended person'.

1.3 Personal data as part of an extended self

Personal data is different from property both in qualitative terms and in terms of its impact and potential use. First of all, data has little or no intrinsic value without its connection to its owner – indeed, much of its value is in its connection to its owner. A car, for example, is of value in itself. The information that one particular person owns one particular kind of car has only minimal value unless something is known about the person who owns it – and is of most value of all if exactly who that person is becomes known, for then it can reveal further information used in selling that kind of car, or in selling further things to the person who owns the car. If, for example, it is known that the author owns a Peugeot 308, then the more that is known about the author (his age, his family situation, where he lives, what other cars he has owned in the past) then the more that initial piece of information gains value. Peugeot can more accurately focus their marketing – and other companies can attempt to advertise or sell more accurately to the author. The value of personal data is as much in the 'personal' as in the 'data'.

Secondly, the loss of control over it can have much more impact than the value of the property itself. Indeed, the value of the data to the person might often only exist in any real terms in connection with loss of control over it. Continuing the example above, the information that the author owns a Peugeot 308 is all-but valueless – but if its leaking could subject the author to an avalanche of unwanted junk mail, cold calls or spam email, then the value of keeping control of it and preventing that avalanche might be significant.

[8] For an examination of the way that attitudes to personal data has been moving towards an overly 'material' form, see Simon G Davies *'Re-engineering the Right to Privacy: How Privacy Has Been Transformed from a Right to a Commodity'* in AGRE, P. & ROTENBERG, M. 1997. *Technology and privacy : the new landscape,* Cambridge, Mass., MIT Press. The changes that Davies highlighted in 1997 have become even more pronounced in the years since then.

For other, more sensitive kinds of data, and indeed for any kind of data when combined or aggregated with other data, the value of keeping control could be much more important.

This combination of factors makes the connection between the person and the personal data much stronger and more 'entwined' than is the case with a conventional property relationship. Psychologists like Clark and Chalmers have investigated the concept of an 'Extended Mind',[9] attempting to answer the question 'Where does the mind stop and the rest of the world begin?'. They have looked at the interaction between the mind and the perceived environment, particularly in the context of computer (and other technology) use, and suggested that the barrier is not as clear-cut as it might at first seem.

A full analysis of this argument is beyond the scope of this thesis – but what can be suggested is that the connection between a person and their personal data, particularly insofar as it is used in the online environment, is a stronger one than that of conventional property, or indeed of that that has existed in the past between a person and the information held about them. That connection is close and complex – and important enough to demand protection. Further, it is through that data that the person interacts with and impacts upon the online environment – and the accumulation of that data makes up how the person is perceived and experienced by the online environment and by other people operating in that online environment. It is in that sense, too, that a person's personal data can be seen as part of what might be considered their 'extended person' – and if autonomy is to be taken seriously, then autonomy needs to be exercisable over this extended person.

2 The reality of data vulnerability

Before setting out possible solutions, the various ways in which data is vulnerable must be considered, by examining a number of the recent examples of data loss and data vulnerability.

[9] CLARK, A. & CHALMERS, D. J. 1998. The Extended Mind. *Analysis,* 58**,** 10-23.

2.1 The HMRC data loss

In the aftermath of the HMRC loss, the government commissioned a full investigation into the affair, performed by Keiran Poynter, Chairman of PricewaterhouseCoopers LLP. The report ('the Poynter Report') revealed a complex combination of individual errors, communications failures and institutional deficiencies at HMRC. The National Audit Office (NAO) undertakes audits of the operations of HMRC, and in the course of those audits had requested information from HMRC.

In March the NAO requested details of 'all new and terminated cases' for the financial year 2006/07, but what was really wanted was a sample extract of data, and to discuss the format and nature of that sample.[10] HMRC sent a sample of 12 full records extracted from their most recent '100% Scan' of the Child Benefit Computer System (a scan which they undertake regularly as a part of their normal business). The NAO looked for ways to reduce the data being sent, but HMRC, under pressure to reduce costs, did not want to go to any extra effort, so simply sent the whole of the 100% scan, on two CDs, which were then sent on to the NAO, all without effective scrutiny by the senior management of the HMRC.[11]

This March 2007 data transfer went without a hitch – though there were a large number of weaknesses, none of them resulted in any problems. That then led directly to the October 2007 data loss. The NAO employee again asked for data, and this time was challenged as to the need for a full copy of the data. This challenge was met with a citation of the need for continuity with the March 2007 data transfer, which was accepted.

This time, however, the NAO employee asked that the data scans be sent to the NAO office in London. After a good deal of confusion, partly as a result of

[10] Full details of the correspondence between the various employees of the NAO and HMRC can be found in POYNTER, K. 2008. Review of information security at HM Revenue and Customs. *In:* TREASURY, H. (ed.). London: HMSO., Part I Section IV.
[11] There were many other weaknesses in the system highlighted by Poynter, including the discs spending a weekend in the house of one employee.

an apparent misunderstanding as to which tray was used for normal post and which for secure, traceable post, two CDs containing the whole of the database were sent by 'untraceable internal mail' from HMRC in Washington, Tyne and Wear to the NAO offices in London. That internal mail system had been contracted out to the courier company TNT – thus giving another party access to the data. Precisely what happened to the discs is still unclear – what is clear is that they did not arrive at their destination. Further copies of the discs were subsequently made – adding another level of vulnerability – and these eventually did arrive at the NAO offices.

Poynter summarises the events leading to the October 2007 data loss as follows:

> "…the events giving rise to the October 2007 loss were predicated by the custom and practice from March, though a lack of clarity in communications and failure to involve sufficiently senior HMRC staff were contributory factors in both cases."

In all, according to Poynter "more than thirty HMRC staff from four different departments, and a number of NAO staff, played some part in the story." Poynter points to such factors as placing operational concerns above security risks, failures to keep to official procedures, failures to either seek or consider appropriate authorisation for removal of data off-site, the routine use of insecure methods of data storage and transfer, and also institutional HMRC factors such as an insufficiently strong and poorly communicated information security policy and a lack of awareness and training on information security among HMRC staff.[12]

[12] POYNTER, K. 2008. Review of information security at HM Revenue and Customs. *In:* TREASURY, H. (ed.). London: HMSO. pp8-11

2.2 MOD Personal Data Loss

Before drawing conclusions from both the events surrounding the HMRC data loss and the Poynter review, it is worth examining another highly public loss of data: the theft, in January 2008, of a Royal Navy recruiter's laptop, which contained the unencrypted personal records for more than 600,000 recruits and potential recruits. This too was the subject of a detailed government review, this time undertaken by Sir Edmund Burton for the Permanent Under Secretary, Ministry of Defence ('The Burton Review').[13]

The events were much simpler than those of the HMRC data loss – the Royal Navy recruiter left his laptop overnight locked in the boot of his car, from which it was stolen. The main issues that arise from that event are firstly why a laptop would be left in a car, and secondly (and far more importantly) why such a laptop should include such potentially highly sensitive personal details, and in an unencrypted form.

The first issue is quite easily understood. The recruiter concerned was 'in clear breach of physical security rules'.[14] Further investigation revealed that this was not a unique occurrence – since 2003 nine other recruiters' laptops had been stolen and in at least three of those cases the thefts had been under similar circumstances, taken from parked cars. Security policies were simply not being applied.

The second issue is more complex. Looking first at the issue of encryption, the Burton Review revealed there had been an intention that all laptops should be properly encrypted, but through a mixture of administrative changes and technical difficulties that intention had been delayed a number of times and various proposals watered down. The original intended laptop encryption compliance date was April 2006 – but by the time of the laptop theft, nearly two years later, compliance was not expected and had not been

[13] BURTON, E. 2008. Report into the Loss of MOD Personal Data. MOD.
[14] Ibid. Part 1 p8

achieved. Even those laptops that had the encryption packages on them did not have them fully installed or operational.[15]

Looking next at the issue of which data should be held, there is an uncomfortable parallel with the events of the HMRC data loss. In essence, it appears that the laptops all contained all the data, largely because that was the easiest way to do it – which meant that the details of about a million people (recruits, potential recruits, 'next-of-kin' and so forth) were being held on all the MOD recruitment laptops. As a matter of design, the laptops synchronised their databases with the full database on the MOD main server.[16] This is an almost exact reversal of the crucial data protection principle of data minimisation.

The Burton Review's main conclusions in terms of data were strong: in overall terms, that "[t]he Department is not treating information, knowledge and data as key operational and business assets", that "[i]nformation risk is not being formally managed at executive boards across the Department, with a small number of exceptions", and that "…there can be little assurance that information is being effectively protected." Ultimately, as Burton puts it, "[a] serious security event of this nature was inevitable."

The parallels with the Poynter Report into the HMRC data loss are very strong. The problems start with management and management systems, and work all the way down. Most importantly of all, in both cases, the data minimisation issue is fundamental to the severity of the problems. The HMRC discs should have held only a small sample of the child benefit data, but in fact held information for all 25 million people, while the MOD laptops could only have held relevant data but actually held the entire database.

[15] Ibid. Part 1 p8
[16] Ibid. Part 1 p9

2.3 Other government data losses – common themes and conclusions

Many other examples of government data loss have come to light since the HMRC data loss, revealing the same kinds of failings as those identified by both the Poynter Report and the Burton Review – failures of policy and of implementation of policy, individual errors, failures of sub-contractors to fulfil their duties and so forth. Examples include: a box of data and files from the Department of Work and Pensions found by a motorist near a roundabout in Devon, which had apparently been lost by courier company TNT,[17] up to 6,000 NHS 'smart cards' used to give access to confidential patient records going missing,[18] payroll documents for 182 staff members in the NHS found dumped in a street in Stevenage.[19] These latter documents had been in the custody of another third party subcontractor, Capita.

There have also been significant losses in the UK by local governments: in 2008, a BBC survey revealed that personal data about members of the public has been lost or wrongly revealed by 13 London councils in the last year, including sensitive information about children in care, stolen when a youth worker took files into a bar, and files containing court reports and a review of a statement of special educational needs stolen from a bag while a worker was in a pub.

The cumulative impact of these various data losses and related problems has been recognised as significant by the authorities. As well as the Poynter Report and Burton Review into the HMRC and MOD data losses respectively, in June and July 2008 there were a number of other official investigations and reports. The Independent Police Complaints Commission investigated the HMRC data loss, and reported on that investigation.[20] At the direct request of the Prime Minister, Sir Gus O'Donnell produced a report into 'Data Handling

[17] See http://news.bbc.co.uk/1/hi/england/devon/7198043.stm
[18] See http://news.bbc.co.uk/1/hi/health/7230512.stm
[19] See http://news.bbc.co.uk/1/hi/7319293.stm
[20] Downloadable from http://www.ipcc.gov.uk/final_hmrc_report_25062008.pdf

Procedures in Government'.[21] July 2008 saw the production of the Data Sharing Review, by Richard Thomas and Mark Walport.[22]

In November 2008, the Information Commissioner's Office produced its own report 'Taking stock, taking action', which examined all of the above investigations and reports and attempted to draw appropriate conclusions from them all.[23] There are common themes from all the reports – the need for better management systems, more responsibility and at more senior levels, better information risk management, better training for staff and so forth. The ICO report's conclusions are clear:

> "While we recognise that much work has been done in improving information governance over the last year and are realistic in recognising that one can never completely eliminate the risk of data loss, the fact that the ICO continues to receive significant numbers of notifications of information losses indicates that both the public and private sector have to continue to improve in this area."[24]

Perhaps the most important point arising from the ICO report is the recognition that 'one can never completely eliminate the risk of data loss'. Even when procedures are in place they are not always followed, and communication between individuals and departments are rarely completely clear. Another common theme is the role that third party contractors have played. The ICO report suggests that "[w]here a contractor provides information management services, there should be an agreement between the relevant parties which details responsibilities and reflects the policies of the contracting organisation", but in reality the use of third parties can add

[21] Downloadable from http://www.cesg.gov.uk/products_services/iatp/documents/data_handling_review.pdf
[22] THOMAS, R. & WALPORT, M. 2008. Data Sharing Review Report. London: Ministry of Justice.
[23] ICO 2008b. 'Taking stock, taking action'. London: Information Commissioner's Office. Downloadable from http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/ico%20position%20paper%20on%20data%20loss%20reports.pdf
[24] Ibid. p7

another possibility of weakness or risk. Realistically, however, the use of contractors is not only regular current practice but almost certain to remain so in the foreseeable future.

2.4 Commercial Data Risks

A key argument made in many of the reports is that public bodies need to take a more business-like approach to data. That isn't really surprising – business very often drives practice elsewhere, and businesses have been in the forefront of the development of the understanding of the value of data and information. There are, however, as discussed in Chapter Two, significant additional risks associated with the way that businesses deal with data.

2.4.1 The T-Mobile data-selling scandal

On 17 November 2009, the Information Commissioner's Office issued a press release[25] to highlight a particularly pernicious problem that had arisen. Employees of a mobile phone company appeared to have sold details relating to customers' mobile phone contracts, including their contract expiry dates. These had been sold on to competitors who had then used the material to cold call those customers in an attempt to induce them to transfer to those competitors. The ICO investigation revealed that 'substantial amounts of money' had changed hands. The identity of the company involved was revealed very swiftly in the press as T-Mobile, and that thousands of their customers were involved.[26]

Information Commissioner Christopher Graham used the events to highlight the bigger issues:

[25] ICO press release 17 November 2009, downloadable from http://www.ico.gov.uk/upload/documents/pressreleases/2009/mobile_phone_records_s55_171109.pdf

[26] See http://www.telegraph.co.uk/technology/mobile-phones/6591726/T-Mobile-customers-hit-by-data-sale-scandal.html

> "More and more personal information is being collected and held by government, public authorities and businesses. In the future, as new systems are developed and there is more and more interconnection of these systems, the risks of unlawful obtaining and disclosure become even greater."[27]

He went on to suggest that greater punishments were needed for this kind of offence – in particular, custodial sentences. That call got support from customer rights' groups and opposition politicians both from the Conservative and Liberal Democrats. It is hard not to agree that the penalties in force at the time, fines of a few thousand pounds, were inadequate punishments to act as any kind of a deterrent.

The key point is that it was not the ICO who uncovered the issue, but T-Mobile, after an internal investigation following customer complaints. [28] The immediate question that arises is how many other similar breaches have gone undiscovered or unreported? For T-Mobile the whole story was an embarrassment – to have to admit to having had such a lapse in security is something which could not do anything but damage to the company: would other companies in similar circumstances be so quick to come forward?

2.4.2 Vulnerability through bankruptcy – XY Magazine

The US owner of XY Magazine and its associated website, whose target readership and subscriber database consisted of young homosexual boys, filed for bankruptcy in 2010. XY's creditors have applied for possession of the company's user database, quite logically from their perspective, since it is probably the company's most valuable asset. That database includes around a million users and by its very nature reveals some of the most personal and sensitive of information of people in an enormously sensitive and vulnerable

[27] ICO press release 17 November 2009, downloadable from http://www.ico.gov.uk/upload/documents/pressreleases/2009/mobile_phone_records_s55_171109.pdf

[28] As reported in The Register at http://www.theregister.co.uk/2009/12/09/tmobile_ico/

situation. The potential for harm is huge – and the legal situation far from clear. Under European Law, data protection would apply, but in the US there is no such protection. The Federal Trade Commission in the US has expressed concerns, and suggested that the sale of the database could be in violation of the legal prohibition of "unfair or deceptive acts or practices", but at the very least this is something that could be subject to significant legal argument.[29]

XY Magazine is a graphic example, and one that has had a great deal of media attention, as well as drawing in privacy advocates such as Simon Davies of Privacy International.[30] It is almost certain, however, that it is just one of many examples of data being sold or transferred as a result of bankruptcy or similar procedures. Where data is so obviously 'dangerous' the kind of attention shown to XY Magazine might potentially prevent the transfer of data, but less clearly sensitive data could be far more likely just to be sold as part of the normal sale of the assets of a company – particularly in a legal environment like that in the US, without data protection laws.

Many of the kinds of weaknesses highlighted in the various government reviews noted above are likely to exist in commercial organisations as well. Organisational issues, senior management responsibility issues, staff competence and staff training issues, pressures on time and on costs, the use of third-party contractors and so forth exist in commerce as much as in government, and indeed often without the kinds of safeguards and detection processes existing in government organisations. The converse is also true: that the weaknesses and risks more commonly thought of in connection with commercial organisations can also exist to an extent in government departments and connected bodies. As the value of data becomes better understood and appreciated pressure will grow upon government bodies to take advantage of that value.

[29] See for example the BBC report on http://www.bbc.co.uk/news/10612800 and the New Statesman at http://www.newstatesman.com/magazines/2010/07/gay-magazine-creditors-legal

[30] Quoted in both the media reports noted above

2.5 Vulnerability to Governments

Vulnerability of data held by governments or government agencies is one thing, the vulnerability to government action of data held by others, whether they are commercial, academic or other organisations, is another. Governments often wish to use data gathered and held by others. What is more, they have used a wide variety of means to acquire this data.

2.5.1 Direct legal action – Subpoenas

Firstly, and most directly, governments can attempt to acquire data using legal action. In August 2005 for example, the US Department of Justice (DoJ) filed a request for data to Google, requesting not only search data but related web addresses. The Department of Justice wanted the data as a part of its attempts to combat child pornography[31]. Initially, the DoJ requested a substantial amount of information – both all the websites that could be located using particular search terms and details of all the specific searches that were made on those terms over the two month period covering June and July 2005 – but this was eventually narrowed down only 5,000 sample searches from Google's search log. Google resisted the initial government requests, and actively defended the subpoena, resulting in a decision at a California District Court[32] that was somewhat inconclusive – Google was ordered to supply some of the URL data, but not the search data. Expert opinion as to who exactly 'won' the case is divided, but Daniel Solove concluded that 'Overall, I view this opinion as a victory for information privacy'.[33]

The case is interesting from many different perspectives. Firstly, it demonstrated that the DoJ saw the potential for the use of search data – both

[31] Specifically for its defence in the case ACLU v. Gonzales, No. 98-CV-5591, pending in the Eastern District of Pennsylvania. The case involved a challenge by the ACLU to the Child Online Protection Act (COPA), 47 U.S.C. § 231
[32] Gonzales v. Google, Inc., No. CV 06-8006MISC JW (Mar. 17, 2006)
[33] As a part of his analysis of the case at http://www.concurringopinions.com/archives/2006/03/the_google_subp.html

Google's database of URLs and its record of searches made. Secondly, its initial approach to Google was simply to 'ask' for the data rather than go through a legal procedure. It emerged during the case that other search engine providers, notably Yahoo, Microsoft and AOL, were not subject to similar subpoenas because they probably did provide the information asked of them or at least came to some sort of accommodation with the DoJ behind closed doors.[34]

What remains unknown is how many such requests have occurred in similar circumstances and how often such requests are just quietly complied with, and how often governments get access to data without the outside world even being aware that they have asked for it. If Google had not resisted the initial government request, none of this information might ever have become public.

2.5.2 Direct government action – the use of legislation

Governments all over the world use the law to provide themselves with access to information. In the UK, there are two particularly relevant pieces of legislation: the Data Retention (EC Directive) Regulations 2009[35] and the Regulation of Investigatory Powers Act 2000[36] ('RIPA'). Data Retention has been discussed in depth in Chapter Three, while RIPA has been touched upon in Chapter Four insofar as it relates to the interception of communications. RIPA also contains extensive powers relating to the use of communications data – data which data retention law has required communications providers to retain.

A full discussion of RIPA is beyond the scope of this thesis, but there are two aspects that are particularly of note here. The first is that RIPA grants the power to use communications data to a very wide variety of public bodies,

[34] See for example Phillip Lenssen's blog at http://blogoscoped.com/archive/2006-03-18-n45.html
[35] Available online at http://www.legislation.gov.uk/uksi/2009/859/contents/made
[36] Available online at http://www.legislation.gov.uk/ukpga/2000/23/contents

from the various arms of law enforcement to local government bodies, the Charities Commission, various bodies within the NHS and so forth. It also grants powers to use direct surveillance to a similarly wide set of organisations. The extent to which such powers are authorised depends on the relevant body – but there have been controversies about their misuse, from local councils using surveillance to determine whether parents were lying about where they lived to get their children into particular schools[37] or to catch those who let their dogs foul the grass.[38] The scope for misuse of these powers is extensive.

The second aspect of importance is that RIPA contains a provision requiring the disclosure of encryption keys to allow authorities to access encrypted data. These powers have been brought into action – and convictions have taken place of people who refused to supply the keys.[39] The combination of these powers – data retention, access to communications data for a wide variety of public bodies, and the power to force people to hand over encryption keys provides the authorities with a powerful set of tools.

Another important vulnerability relates to the way that data held in one country can be vulnerable to laws passed in other countries. Perhaps the most direct of these relevant in the UK relates to the powers set out in the United States through the USA PATRIOT Act.[40] Section 215 of the Act, which revises the Foreign Intelligence Surveillance Act of 1978 (FISA), provides that designated FBI personnel may apply to the FISA court for an order requiring the production of business records relevant to an investigation concerning international terrorism or clandestine intelligence activities. Applications to the FISA court do not require 'probable cause', but simply a claim by the FBI that the records are needed for an ongoing investigation into something that can loosely be classified as related to international terrorism or intelligence

[37] See for example http://news.bbc.co.uk/1/hi/england/dorset/7343445.stm
[38] See for example http://news.bbc.co.uk/1/hi/england/northamptonshire/7414382.stm
[39] See for example http://www.theregister.co.uk/2009/08/11/ripa_iii_figures/
[40] The Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001, downloadable from http://thomas.loc.gov/cgi-bin/bdquery/z?d107:hr03162:%5D

activities. What is more, the terms of the law demand that individuals served with a warrant under FISA rules may not disclose, under penalty of law, either the existence of the warrant or the fact that records were provided to the government.[41]

This makes data gathered and held in the US particularly vulnerable – but there is another possibility that is relevant here. The USA PATRIOT Act's requirements cover US companies, and allow the US authorities to access data held by those US companies – and this may be so even when that data relates to and was gathered from a non-US company. When in 2008 the US defence company Lockheed Martin bid for the contract to run the 2011 UK census it was speculated that the USA PATRIOT Act might have allowed US authorities access to all the data gathered by Lockheed Martin as a part of that UK census. Questions on the subject were raised in the Commons Treasury sub-committee, and whilst Angela Eagle, the then Treasury Minister, told the committee that she had received legal assurances that this would not happen, the doubt and the question remain. Angela Eagle's exact words were that she was "pretty confident" that there would be robust safeguards on the security of data.[42] Lockheed Martin did eventually win the contract, after strict assurances were made that the data would remain in the UK and under the ownership of the Office for National Statistics but the larger issue remains – if data are gathered by a company owned under one jurisdiction and operating under another, the laws of both jurisdictions might apply.

There is another scenario to consider too – that where data are gathered by one company and then that company is taken-over by another company. In particular, if a UK company gathers data and then that UK company is bought by a US company, could the USA PATRIOT Act allow US authorities to access all the personal data owned by that UK company? At the very least, there is a

[41] For analyses of the reality of Section 215 of the USA PATRIOT ACT see the Friends Committee on National Legislation analysis at http://www.fcnl.org/issues/item.php?item_id=344&issue_id=68 or the EPIC analysis of the USA PATRIOT Act at http://epic.org/privacy/terrorism/usapatriot/

[42] As reported to the BBC. See http://news.bbc.co.uk/1/hi/uk_politics/7231186.stm

conflict of law – data protection law in Europe against the USA PATRIOT Act. Moreover, as the HMRC data loss example shows through the lack of clarity of the roles of the NAO in relation to HMRC, it could provide another potential area for uncertainty, and hence vulnerability. An individual data controller, faced by a request from the US authorities, might not be clear enough to known that they do not need to comply – indeed that they must not comply – and hence may allow access even when they should not.

2.5.3 Use of illegally acquired data

Another practice that has emerged in recent years is the use by governments of illegally acquired data. What appears to be happening is that governments, as their understanding of the potential use of data has grown, have also begun to be aware of the many sources of these data – and have few scruples about how that data might have been obtained. Two particular examples are notable from the last few years: the Chinese 'Google Hack' and the use by governments of stolen bank data from Lichtenstein and Switzerland. Though these two cases might seem at first to have little in common, when looked at more closely they have many similarities, particularly from the perspective of the governments concerned.

a) Google and China – the 'Google Hack'

In January 2010, Google reported that it had been the subject of a 'highly sophisticated and targeted attack on [its] corporate infrastructure'.[43] Google has 'evidence to suggest that a primary goal of the attackers was accessing the Gmail accounts of Chinese human rights activists' and that 'as part of their investigation, but independent of the attack on Google [they] have discovered that the accounts of dozens of U.S.-, China- and Europe-based Gmail users who are advocates of human rights in China appear to have been routinely accessed by third parties. These accounts have not been accessed

[43] Revealed in the Official Google Blog at http://googleblog.blogspot.com/2010/01/new-approach-to-china.html

through any security breach at Google, but most likely via phishing scams or malware placed on the users' computers."[44]

The revelation of this hacking attack started a row that brought the US and Chinese governments into direct conflict. Hillary Clinton, the US Secretary of State, called on Beijing to investigate the Google hack,[45] while the Chinese state-run news agency Xinhua attacked Google for having 'intricate ties' with the US government and of 'providing US intelligence agencies with a record of its search engine results',[46] which is interesting considering Google's resistance to the US DoJ's subpoena. Google used the event to trigger a withdrawal of cooperation with China's requirement for a strictly censored search engine and providing instead a simplified, uncensored search facility based in Hong Kong.[47]

It may never become clear whether the Chinese government was directly involved in the hacking of the data. Google themselves stated that they did not believe that the hack succeeded in accessing the Gmail accounts of the Chinese human rights activists that it was targeting, so it cannot be shown that the results of the hack have been used by the Chinese government. What is clear is that someone has been trying to access this kind of data, and presumably because they believe they can find some kind of use for this data – and the most obvious potential users of this kind of data are the Chinese government. There are a number of possibilities, the most direct of which are that the hackers were working directly for the Chinese government, that they were independent but commissioned by the government, or that they were independent but believed that they could sell the data they hacked to the Chinese government.

[44] Again, see the Official Google Blog at http://googleblog.blogspot.com/2010/01/new-approach-to-china.html
[45] See http://news.bbc.co.uk/1/hi/8472683.stm
[46] See http://news.bbc.co.uk/1/hi/8578968.stm
[47] See http://googleblog.blogspot.com/2010/03/new-approach-to-china-update.html

b) Lichtenstein and Swiss banking data

A seemingly very different example is the use of hacked or otherwise illegally obtained data to find and convict tax evaders. The practice emerged in Germany in 2008, when the German government purchased information taken from the Lichtenstein bank LGT,[48] in a complex story that eventually effectively forced Lichtenstein to drop most of its banking secrecy laws.[49] The Sunday Times reported that the German foreign intelligence service had paid 5 million Euros for the data[50] – and that HMRC in the UK had acquired similar information for £100,000.[51]

Then in February 2010, the German state of North Rhine-Westphalia announced that it had bought Swiss bank data as a part of its drive to deal with tax evaders, after having been given the go-ahead by the German federal government to buy the information, even if it had been obtained illegally.[52] The German authorities did not just use the information they had acquired to catch tax evaders, but by publicising the fact that had acquired this data – and would do their best to acquire similar data in the future – to convince other tax evaders to come forward voluntarily. This kind of 'persuasion' was successful: Reuters reported that 5,900 German citizens owing around 500 million Euros came forward in the months following the initial revelation of the data acquisition.[53]

In March 2010 the Sunday Times revealed that the authorities in Germany had been joined by their equivalents in France in the acquisition of this kind of data, and that those in the UK were also considering following the same path. According to the *Sunday Times* report, Hervé Falciani, a French software engineer, had illegally obtained the details of 24,000 customers with

[48] See for example http://www.reuters.com/article/idUSLDE6160KO20100207
[49] See for example http://online.wsj.com/article/SB124727784329626615.html
[50] See http://business.timesonline.co.uk/tol/business/industry_sectors/banking_and_finance/article3399526.ece
[51] See http://www.timesonline.co.uk/tol/money/tax/article3423428.ece
[52] See http://www.reuters.com/article/idUSLDE61P1FN20100226
[53] Also reported in http://www.reuters.com/article/idUSLDE61P1FN20100226

accounts at HSBC's private bank in Switzerland while working for the company in Geneva.[54] Falciani had already sold some of the information to the French government and HM Revenue & Customs in the UK was apparently about to receive some of that information from their French equivalents as a part of an information exchange. A 'senior tax official' was quoted as saying:

> "It's fair to say that the prospect of getting hold of this information has generated some excitement here."[55]

The Swiss authorities asked the French to have the data returned, but though the French authorities agreed, they kept copies of the files and used them to try to root out tax evaders – and France's tax office is thought to have subsequently recovered around half a billion Euros.

c) An emerging practice?

The media responses to the two examples were very different. The suggestion that the Chinese government might have been behind the Google hack provoked a mixture of anger and fear, links to the ideas of cyberwarfare and a general sense of paranoia about the way that the Chinese government seeks to suppress any kind of dissent, and the hackers as stooges in the hands of a totalitarian state, while the banking data hackers/whistleblowers have been portrayed as heroes, helping to bring tax-evaders to justice.

From the perspective of the governments concerned, however, the two scenarios are actually very similar. In both cases the governments feel they have an overwhelming need (and duty) to locate and catch people perpetrating serious crimes – for the German, French and UK governments the people evading taxes, for the Chinese, people putting the stability and security of their state at jeopardy. In both cases the governments are using

[54] See http://www.timesonline.co.uk/tol/news/politics/article7061114.ece
[55] See again http://www.timesonline.co.uk/tol/news/politics/article7061114.ece

whatever means they can find to perform that duty. Further, governments believe that this kind of a move will be popular. As German Finance Minister Wolfgang Schäuble put it:

> "In view of growing social tensions caused by globalization, the financial market crisis and the ludicrous bonus payments on the one hand, along with growing unemployment on the other, it would be completely unbearable if the state didn't do everything possible to ensure taxes were collected fairly."[56]

The Chinese government has not claimed responsibility for the Google hack, but it would be easy to frame a similar argument in support of such actions – that to respond to growing social tensions caused by globalisation (of communications) it would be completely unbearable if the state didn't do everything possible to prevent subversives from destroying the Chinese state and way of life. To take it another step further, consider the case of terrorism – a significant part of the so-called 'war on terror' has been based on the premise that almost everything should be allowed that can help is stop the terrorists from destroying the West's way of life, including all kinds of surveillance.

For the Chinese Government, subversion and dissent would certainly be defined as serious crime under its national law. There are strong parallels, therefore, between the use of data for the 'war on terror', for the catching of tax-evaders and for the suppression of dissent and subversion. In all these cases the importance of the objective seems to have overridden the need for an 'appropriate' method of obtaining the data.

The precise legal status of data acquired in this kind of a way is not entirely clear. There are conflicting and competing imperatives. In the case of the stolen bank data the German government has a duty to do what it can to collect tax fairly, but also has obligations under the Convention on

[56] http://www.reuters.com/article/idUSLDE6160KO20100207

Cybercrime, which it has both signed and ratified.[57] Those obligations include a requirement for international cooperation, and the extradition of those who commit computer crimes including the theft of data[58] – so the Swiss authorities would have expected the German government to arrest and hand over to them the hackers/whistleblowers who stole the data from the Swiss banks, together with the data that had been stolen. No such actions have taken place – the German government has effectively deemed its duties under domestic tax law to override its obligations under the Convention on Cybercrime. Indeed, it has gone further, for it can also be argued that by buying the data and by signalling that it would buy similarly acquired data in the future, the German government has effectively encouraged the commission of further data thefts – or the sale of the data to other countries. The further sale of data to the French and UK government adds weight to this suggestion – and demonstrates the way in which the existing international data security framework, of which the Convention on Cybercrime is a key pillar, is being denied effectiveness by Signatory States.

Acquiring the data through this kind of a process might also give governments more freedom in terms of how it is used than they would have if they acquired them in a more conventional way. As discussed in Chapter Three, data protection law requires specification of purpose – so when personal data are gathered the data subjects must be informed as to the purpose for which those data have been gathered. If data is gathered in this 'indirect' way, how is that purpose specified? Governments might use the same definition as set out above from the Data Retention Directive – that the purpose is the 'investigation, detection and prosecution of serious crime', which would then give them freedom to use the data exactly as they want in accordance with this goal. All that might even be sidestepped depending on the particular implementation of data protection law. The Federal Data

[57] The Convention on Cybercrime is downloadable from http://conventions.coe.int/Treaty/EN/Treaties/html/185.htm. Germany ratified the Convention on 9th March 2009. See https://wcd.coe.int/ViewDoc.jsp?id=1416299&Site=DC

[58] Convention of Cybercrime, Chapter III

Protection Act in Germany, for example, says that data may be collected without the participation of the data subject if:

"…a legal provision prescribes or peremptorily presupposes such collection…"[59]

German tax evasion laws would provide precisely that kind of a legal provision. Either way, the purpose specification and consent requirements of data protection law can be avoided.

2.5.4 Vulnerability to Governments – complications:

a) Easy targets or slippery slopes?

In most of the examples above, the 'targets' chosen by the governments are what might be described as 'easy targets', in the sense that they are exactly the kind of 'offenders' that the public would want to get 'caught'. The people concerned, whether they are tax evaders or terrorists (and from time to time creators or consumers of child pornography) might even be described as 'enemies of society' – though that emphasises again the parallels with the Chinese government's drive to root out subversives and dissidents.

It is not unusual for these kinds of offenders to be used to bring in laws that have far wider implications and cover far more people than their significance would suggest – and for the rights of ordinary individuals to be restricted as a consequence of the desire to apprehend them. The struggle to maintain civil liberties in the face of the demands of security is echoed strongly in this field. The practices and policies that have emerged to catch tax evaders, terrorists and murderers could equally be used to catch more minor 'offenders' – in the same way that anti-terrorism laws have been used in the UK to prevent

[59] German Federal Data Protection Act, Section 4 (1). Downloadable including English translation from http://www.bdd.de/Download/bdsg_eng.pdf

protests against the arms trade,[60] and as noted above the powers of RIPA used to catch the owners of fouling dogs.

This pattern can be viewed in a number of ways. Some privacy advocates see it as a kind of 'Trojan Horse' phenomenon – the 'easy targets' are used as an excuse to bring in powers or systems that governments would have liked to bring in anyway. Looked at another way, it is simply a matter of efficiency – if the systems, laws or powers exist, why not use them for other things? The local councils who have used their CCTV cameras to deal with dog fouling would certainly have seen it that way, and if a vast system of cameras like those used to enforce the London Congestion Charge exists, why not use those cameras for other, important purposes like the prevention of terrorism?[61] Either way, the phenomenon does appear to exist[62] – and the implications in relation to data could be even greater than those in the material world, if the possibilities for aggregation, analysis and data mining are considered.

b) Sensitive and 'less sensitive' data?

It might be thought that governments would only be interested in data that is of direct use, data which might immediately be thought of as significant and sensitive (following the rules of data protection law, as discussed in Chapters Three and Four) such as the financial data used to catch tax-evaders, politically sensitive information such as that sought by the Chinese hackers of

[60] In the case of *R. (on the application of Gillan) v Commissioner of Police of the Metropolis* [2003] EWHC 2545 (Admin), Gillan and Quinton challenged the use of anti-terrorism laws to stop and search them at an arms trade protest. They lost, lost again on appeal and again in the Lords, but the European Court of Human Rights overturned that decision, in *Gillan v United Kingdom* (4158/05)
(2010) 50 E.H.R.R. 45; 28 B.H.R.C. 420

[61] As reported in The Register: see http://www.theregister.co.uk/2007/07/18/smith_n_mcnulty_surrender_to_jihadi_bunglers/

[62] Function creep has been recognised and studied in various forms relating to this and associated fields – for example ID cards, DNA databases and Sex Offenders Registers. Examples of academic work in this field include GREENLEAF, G. 2008. Function Creep - Defined and still dangerous in Australia's revised ID Card Bill. *Computer Law & Security Report,* 24**,** 56-65., DAHL, J. Y. & SÆTNAN, A. R. 2009. "It all happened so slowly" - On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice,* 37**,** 83-103. and THOMAS, T. 2008. The Sex Offender 'Register': A Case Study in Function Creep. *The Howard Journal,* 47**,** 227-237. respectively.

Google, medical data and so forth. That, indeed, may generally be the case, but as discussed in Chapter Four, almost all data gathered and held may potentially be used to derive sensitive data – and governments everywhere are beginning to realise it.

There are a number of different examples of this, some of which have been touched on earlier in this chapter. One kind of use is for establishing patterns, and then using them to identify potential suspects/targets – through 'data mining', as discussed in Chapter Two. The way that MI5 was attempting to use Oyster Card data to monitor people's travel in London is one example. What applies for travel data could equally apply for many other kinds of data – web-browsing behaviour for example, or shopping behaviour like that collected through supermarket loyalty cards. There are echoes of this way of using data in the service provided by Google to predict regional flu epidemics on the basis of people's searching activities.[63]

What does this suggest? The most important thing is that it means that it should not just be the security and vulnerability of the most sensitive of data that needs to be considered, but that of all data – including all commercial data.

2.6 Hacking and technological vulnerability

There is another area of crucial importance when considering the vulnerability of data – vulnerability to hacking or other forms of technological attack. Hacking can and does occur, and there is a constant 'battle' between hackers and those employed to make systems secure. The Google Hack discussed above is one example. Though there has been considerable debate about who was responsible, no one has attempted to deny that the hacking attack took place. Taking it a step further, the whole concept of 'cyberwarfare' is predicated by the understanding that systematic attempts to hack into systems not only can happen but are happening.

[63] The service is called Google Flu Trends, found at: http://www.google.org/flutrends/

Dealing with the details of the technological issues is beyond the scope of this thesis – what needs to be understood at this stage is that data in almost any situation could potentially be vulnerable to some kind of technological attack or intrusion.

A particularly graphic example of this kind of vulnerability happened in April 2011, when hackers went into the Sony Playstation Network and stole the personal details of more than 100 million users.[64] These details included names, home addresses, email addresses, dates of birth, phone numbers, gender information and 'hashed password' – and in some cases direct debit details, credit card numbers and expiry dates. The direct debit and credit card details came from what Sony described as an 'outdated database'[65] – which in itself raised a lot of questions, most directly why that database even existed, let alone was accessible online for hackers. In terms of data minimisation – a common thread for many of the cases here – it is hard to see any justification for it. When the nature of Sony is considered, the hack is very revealing. Sony should be amongst the most technologically advanced and sophisticated organisations, with access to the best experts in security and in particular network security – and yet they were hacked, and hacked with great success. If Sony can be hacked, is anyone secure?

Sometimes it does not take particularly great technical expertise to access information on the internet. In May 2011, Matthijs R. Koot, a PhD student in the Netherlands, used simple techniques to mine Google's databases and put together a database of 35 million Google users including names, email addresses and biographical details. As Koot put it, it was 'completely trivial for a single individual to do this,'[66] and the process was entirely within Google's rules, as they allow indexing of their public user information.

---

[64] Sony has acknowledged that 77 million users of Playstations and 25 million users who access the Playstation Network through PCs or Facebook may have had their data stolen. See http://www.soe.com/securityupdate/pressrelease.vm and http://www.soe.com/securityupdate/index.vm

[65] http://www.soe.com/securityupdate/pressrelease.vm

[66] See for example http://www.theregister.co.uk/2011/05/25/google_profiles_database_dump/

2.7 The Reality of Data Vulnerability – the Big Picture

The organisations which have been included in the analysis here include those which 'ought' to have the strongest security. The government departments examined in the first two examples, HMRC and the MOD, are respectively the department with some of the most sensitive personal data (specifically the financial data) and the department that should have had the best understanding of security. And yet both showed a vast range of weaknesses at almost every level: strategic, managerial, individual and technological. These kinds of weaknesses are not limited to governmental departments – indeed something very similar happened to the HSBC bank, two of whose departments lost CDs with 180,000 and 2,000 customers' data respectively, putting those customers at significant risk.[67]

The Swiss bank scenario demonstrates a number of important things – not just the fact that governments seem willing to condone or even promote the illegal acquisition of data so long as it is in a 'good cause'. Perhaps even more important than that is the simple fact that the data thieves were able to steal the data in the first place. Further, the interaction between the vulnerability *of* governments with the vulnerability of information *to* governments needs to be considered. As has been shown, government bodies have a wide range of ways to acquire data, with various levels of legality. Once they have acquired this data, that data becomes potentially even more vulnerable. Moreover, the more information is copied and transferred, the more opportunities for vulnerability appear.

[67] HSBC was fined £3.2 million by the Financial Services Authority for these data losses. See for example http://www.telegraph.co.uk/finance/newsbysector/banksandfinance/5886419/HSBC-fined-3.2m-for-losing-customers-details.html

3 Data Vulnerability – Solutions?

What, if anything, can be done to address the problems of data vulnerability? There are wide ranges of tools that can be used to help improve the situation, many of which have been identified by the various reviews into the various incidents outlined above.

3.1 Changes in existing law and practice

The first and most direct way to deal with data vulnerability is through changes in existing law and practice. In the UK, this could begin with the suggestion by the ICO of increased fines and harsher sentences to deal with data losses and failures of data security, and in particular, the possibility of custodial sentences – something that arose from the T-Mobile data selling scandal discussed above. The possible penalties have recently been increased – from April 2010, fines could be as great as £500,000[68] – but to date custodial sentences have not been introduced. It seems quite possible, however, that in the current climate and with a new government emphasising these issues, that these kinds of penalties could be brought in. The Information Commissioner also suggested that possibility of extradition in appropriate cases should be opened up – given the nature of the internet that would again seem logical and appropriate.

Whilst there are benefits to these ideas in terms of deterrence, there are also significant weaknesses. The whole question of whether deterrence really 'works' is not within the scope of this thesis, but it is at least fair to say that, like the better use of encryption and other technological security measures, it can only offer part of a solution to the problem. Recent experience also suggests that even in the most obvious cases, and even after stronger powers

[68] Details of the new penalties, and guidelines from the ICO as to how they are intended to be applied are available at: http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/ico_guidance_monetary_penalties.pdf

have been granted, the ICO may not apply them.[69] So long as this is the case, the chance of deterrence is even less likely. Further to that, deterrence can only have a chance of functioning if the potential offenders believe that there is a significant likelihood of their being caught – and that is only likely if the enforcement arms of data protection authorities are substantially strengthened.

That too is something that could potentially make a difference, improving data security and reducing data vulnerability. Even so, the nature of current law and practice, and the principle of proportionality mean that this kind of law – and in particular the idea of penalties harsh enough to act as a deterrent – could only apply to significant breaches and clearly 'sensitive' data. As seen above the problems relating to data vulnerability do not just apply to large scale events or to directly sensitive data – the vulnerability of seemingly innocuous data is also important, and the accumulation and aggregation of individually insignificant pieces of data can also have a significant impact. These kinds of breaches are not only less likely to be detected but even if they are detected are highly unlikely to incur substantial penalties. More to the point, it would not be appropriate for them to do so.

3.2 Better use of technology

There are technological tools that can help with data security – the most obvious being the use of encryption. A proper discussion of the use of encryption is beyond the scope of this thesis, but it is clear that encryption is a powerful tool in the practice of data security. However, it is also important to understand that the real experts do not believe that encryption is anything more than a tool in the overall scheme of things. Ross Anderson, Professor of Computer Security Engineering at Cambridge University, and one of the leading experts in cryptography in particular and computer security in

[69] In the 2011 example of ACS:Law, where the ICO initially threatened the maximum fine, the eventual fine for Andrew Crossley, the sole-trader solicitor behind ACS:Law, was just £1,000. Crossley had wound up ACS:Law prior to the completion of the ICO investigation, so the ICO deemed that he had 'limited means' and though if ACS:Law was still trading the fine would have been £200,000

general, when asked 'How well-encrypted must data be, in order to be safe?' replied:

> "You are in a state of sin. This is a wrong question to ask, for many reasons. `Whoever thinks his problem is solved by encryption, doesn't understand his problem and doesn't understand encryption' (Needham and Lampson)"[70]

Quite how true this is can be seen by the examples of both HMRC and the MOD in this chapter. In the HMRC case, the policy was in place for encryption, the technology for encryption existed, but it simply was not applied. In the MOD case, though the encryption technology had been specified, its application had been delayed a number of times, inconsistently and in some cases incorrectly installed, and rarely actually used by the personnel involved.

What is more, even encrypted data is potentially vulnerable in two different ways. Firstly, the encryption itself can potentially be hacked or broken – there is an ongoing battle between the developers of encryption technology and the hackers trying to break it. Any code can and will eventually be broken – the question is whether those who are attempting to keep the data secure stay ahead of those who are attempting to break it. A further implication of this, and a further potential weakness, is that it requires those who use encryption to keep that encryption up to date – which leaves further scope for human error. That leads to the second weakness – that the use of encryption requires human interaction, and even if the encryption cannot be 'broken', sometimes the human can. As Ross Anderson puts it:

[70] In an interview for simple-talk.com, at http://www.simple-talk.com/opinion/geek-of-the-week/ross-anderson-geek-of-the-week/

> "As designers learn how to forestall the easier techie attacks, psychological manipulation of system users or operators becomes ever more attractive"[71]

Most directly, people might be persuaded either to release the keys to their encryption or even not to use the encryption properly at all. As discussed in many places in this thesis, the use of psychological or emotional manipulation, particularly on the internet, is a developing science. This is Ross Anderson again:

> "Deception, of various kinds, is now the greatest threat to online security. It can be used to get passwords, or to compromise confidential information or manipulate financial transactions directly."[72]

So what does all of this mean? Simply that though technological tools are a crucial part of the process of improving data security and reducing data vulnerability they are far from being the whole solution.

3.3 Changes in the community and culture

An overriding requirement, emphasised in all the reports that have been made into the data losses, is that all the issues concerning information vulnerability and security need to be taken more seriously at every level. That must start from the very top. The ICO's position paper, 'Taking Stock, Taking Action', for example, suggests that a 'role should be created at board level in larger organisations to deal specifically with information risk', and that '[a] post at senior executive level should oversee information security'.[73] The changes must be reflected throughout the organisation, and include proper and professional information risk management policies, periodic

[71] Ibid.
[72] Ibid.
[73] ICO 2008b. 'Taking stock, taking action'. London: Information Commissioner's Office. p6

reporting of information risk at board level, clear lines of accountability and so forth, together with proper staff training and support. This is clearly of great importance, and a crucial first step towards an environment where data vulnerability can begin to be reduced.

The possibility of culture change can be taken to another level. The Poynter Report and Burton Review – and indeed the ICO position paper following all the reviews – focus on that awareness simply in terms of the individuals' roles as employees of their organisations, but the real problem and indeed the potential solution runs deeper. If people were more aware of the issues of data security and vulnerability – and indeed of data privacy – in their ordinary personal and social lives, then it would be far easier for them to understand the importance of data security in their professional lives. They would find it easier to understand and implement information security policies, they would care more if the data encryption systems on their computers weren't functioning properly, and they would be less likely to fall for the kind of deceptive practices used by sophisticated cyber criminals.

This culture change is perhaps the single most important factor – but it is also a factor that is very difficult to change, and something likely to take a considerable amount of time. There are signs that it may be happening, but at the same time there are suggestions of precisely the opposite – Facebook founder Mark Zuckerberg's suggestion at the Crunchie Awards that 'privacy was no longer a social norm'[74] is just one of many who have followed Scott McNealy's famous quote that "You have zero privacy anyway. Get over it." This is a topic of fundamental importance, and one that will be discussed in more depth in Chapter Seven.

[74] Reported for example in http://www.guardian.co.uk/technology/2010/jan/11/facebook-privacy

3.4 Taking data minimisation seriously

Even if the law is both improved and better enforced, if the culture of organisations is more 'data-conscious', and if security technology is used appropriately and effectively – there will still be problems, and risks that cannot ever be completely eliminated. Human errors, human nature, human malice, technological error and technological developments, and community pressures such as the demands to fight terrorism or catch child abusers or murderers are just some of the possibilities. Ultimately, wherever data exist, they are vulnerable – so the only way that data can really not be vulnerable is for it not to exist. Blogger Harry McCracken, when talking about the vulnerability of data held on Facebook said:

> "Facebook has a history of asking for forgiveness rather than permission, and now says the default for everything is "social"–so the best way to keep things private is to keep them off the service, period."[75]

McCracken's argument can be extended not just to cover Facebook, but the whole of the internet. The default for the whole of the internet is that everything is "public": the best way to keep things private is to keep them off the internet completely. Taking it one step further, the best way to keep things private is not to keep them in a digital – and hence vulnerable – form. The ultimate weapon in the fight against data vulnerability is to eliminate the very existence of data wherever possible.

The starting point for this is stronger, better-understood and better-implemented data minimisation. The concept of data minimisation is built into data protection law. It combines the third and fifth principles: that data should be 'adequate, relevant and not excessive' and 'not kept for longer than is necessary'. It is, however, a concept that seems to be paid far less attention too than it should, partly, perhaps, because the terms are very difficult to

[75] See http://technologizer.com/2010/05/11/facebook-privacy-fodder/

define. What is 'excessive' and how long is 'necessary'? In specific cases the point has been argued at length by European regulators – most notably the search engine case that forms the central case study in Chapter Three of this thesis – but in general the answers to the questions have been left to the discretion of the businesses concerned. Unless specifically challenged, those businesses can choose how much data to hold and how long to hold it for – and as things stand, it appears that most businesses choose to hold more data than they need and for longer than they need to.

As noted throughout this thesis, the best way – perhaps the only way – for things to change positively in this field is for a new business model to develop. The key is to find a way to encourage the development of these new models in a way that get closer to a real sense of data minimisation.

There is another way to look at it. One of the keys to the issue of data vulnerability, particularly from the perspective of those interested in autonomy, is the question of who controls the data – and therein may lie a possible solution. If a way can be found to put the data subjects more in control of the data minimisation process, then not only will people be more in control of their own data but businesses would be put in a position where they have to develop business models that do not depend on their ability to gather whatever data they choose and hold it as long as they would like. How can that be done?

4 A change in assumptions – and the right to delete

What is really needed is a paradigm shift – a change in assumptions. The assumption should be that unless you have a strong reason to hold it, data should not be held.[76] Data holders should need to justify their holding, rather than the other way around. There are parallels between this argument and

[76] As discussed earlier in this chapter, this assumption does theoretically exist in data protection law – but as the case studies have demonstrated, the reality is very different, and data minimisation is neither effectively followed or enforced

the arguments in favour of opt-in rather than opt-out discussed in relation to consent in Chapter Four. It can be looked at as an extension to the logic behind the concept of Collaborative Consent introduced in that chapter. Collaborative Consent needs to be continually renewed – but this time it is not just for the gathering of data but also for holding data. The first implication of this is direct – that unless otherwise permitted to keep data, following strict rules which will be set out in a moment, holders of data must directly and expressly receive consent in order to be able to keep holding that data.

The shift of assumptions in favour of deletion rather than retention can be taken a step further – by establishing a general 'right to delete' data. That is, in general a data subject should be see as having the right to have any data held relating to them deleted, and that those holding that data must put into place systems that allow that right to be enforced at any time. This right to delete should not be seen as akin to a 'right to be forgotten' – as indicated at the start of this chapter what is being suggested is not about rewriting history or about censorship so much as it is about placing more effective, better controlled and realisable limitations on government or commercial holding of more data than is needed. Indeed, to describe it as a 'right to be forgotten' could even be seen as misleading or disingenuous – it is not about forgetting, but about control, and about autonomy. Talking about a right to be forgotten is attractive in some ways – but it could also distract from the real point. There are reasons why people might not have a right to be forgotten – rewriting history is something that is rightfully considered contentious or even dangerous – but they are not reasons not to have a right to delete. A right to be forgotten can be seen as a rejection of society and something ultimately undemocratic[77] – whereas a right to delete, properly set out and balanced with other rights, can be something precisely the opposite, and act as a support and protection for individuals in their interactions with society.

[77] Suggested for example by Tessa Mayes at the Westminster Media Forum 22 March 2011, and in her article in the Guardian http://www.guardian.co.uk/commentisfree/libertycentral/2011/mar/18/forgotten-online-european-union-law-internet

The right to delete is a step towards putting data minimisation in the hands of the individual, at least to a certain extent. It is a shift in defaults – rather than asking when people should have the right to delete personal data, the opposite should be asked: in what circumstances and what kind of data should people *not* have the right to delete. The default should be that an individual *can* delete data – an extension of the 'rights of data subjects' set out in the Data Protection Directive.

It is important to ensure and remember that this should not be considered to be an alternative to the current ideas of data minimisation; nor does it remove any aspect of responsibility for data minimisation from the data processor or data controller. Rather it should act as an additional safeguard, another level of protection for the individual.

4.1 When would the data subject not have the right to delete?

It is suggested that there are five principal categories of reason for which data might need to be preserved regardless of an individual's wishes to delete it – where the presumption in favour of deletion can be rebutted.

1 **Paternalistic reasons** – where it is in the individual's interest that the data be kept, and society can override the individual's desire. The primary example of this is medical data;

2 **Communitarian reasons** – where it is in the community's interest that the data be kept. This might include criminal records, for example;

3 **Administrative or economic reasons** – where the economic or administrative needs of society require records to be kept. This could include tax records, electoral rolls and so forth;

4 **Archival reasons** – for keeping a good, accurate and useful historical record of events. This might include newspaper archives, blogs and so forth. This category is very important, but could easily be governed through a system by which a particular database is agreed to be 'archival' in nature, and thus not covered by the right to delete – but also restricted in the uses to which it can be put and so forth. This is in itself another contentious issue. The British Library, for example, is campaigning for a 'right to archive', effectively asking for the right to archive web pages without needing to get permission from the website owners.[78] At first sight this might appear to be precisely the opposite of the shift of assumptions being suggested in this chapter, but in reality the two rights are quite compatible: both require close scrutiny and regulation of an archive. The British Library could be included on a 'register of archivists' that are permitted to keep such an archive – but required to control and report on that archive.

5 **Security reasons** – where the data are deemed to be needed for security purposes. This might include records of criminal investigations, or such communications records as are set out in data retention laws. This category is by its nature highly contentious, and should be subject to close scrutiny – including political scrutiny – and regularly reassessed, applying a principle of proportionality to assess whether the impact on privacy is so sever that it is *not* overridden by the interests of security.

Effectively, these are limitations on autonomy – but limitations that are understood and reasonable in society. They can be compared with the data protection principle of 'fair and lawful processing' concepts (consent, vital interests, administration of justice, functions of crown and public interest), 'processing exemptions' (research, history and statistics, and the special purposes exemptions: journalism, artistic use and literary use), and the exemptions to access rights set out in Schedule Seven of the Data Protection

[78] See for example http://news.bbc.co.uk/1/hi/technology/8535384.stm

Act 1998.[79] All of these cover similar kinds of ground – so the concept of such limitations should be familiar and acceptable. Indeed, setting these terms out from a rights perspective could be part of a harmonisation process, making all these areas consistent and coherent.

It should be specifically stated that 'supporting your business model' should not be a sufficient reason to deny data deletion – this could be viewed simply as taking data minimisation seriously, but needs to be spelled out. One of the purposes of rights, after all, is to spell things out so that people understand the principles, and might even begin to understand the reasons behind those principles.

4.2 Highlighting of profiling and other derived data

The exemptions set out above cover the kinds of data for which deletion should not be possible – but there is another end of the spectrum: a category of data that would need to be specially highlighted as 'available for deletion'. That is, not just that the data subject has the right to delete them, but that attention must be drawn to them and it must be made simple, direct and clear how to delete them. The most direct example of this would be 'profiling' or 'channelling' data – so that an individual would be able to delete information derived about them from their behaviour in one form or another. The reasons for highlighting this kind of data are twofold: firstly, because this kind of data can represent the most direct threat to autonomy, and secondly because profiling or derived data could be a way that data gatherers attempt to avoid or circumvent data minimisation rules in relation to the time that data is held. To give a simplified example, if someone searched for and looked at a particular website in January 2009, then if the periods of data retention suggested by the Article 29 Working Party in their recommendations to Google as discussed in Chapter Three are reasonable, the fact that they performed that search could only be retained for six months, until June 2009.

[79] Data Protection Act 1998, s28, 29, 33, Sch. 7(9), Sch. 7(1), Sch. 7 (8) respectively

If at that point, however, whilst deleting that search log data the search engine provider creates some new 'profiling' data, categorising the person as a 'visitor of websites of that kind in early 2009', that profiling data could be classified as 'new' data in June 2009, and then kept for a further six months, before being incorporated into some new form of profiling data, and kept for another six months. Intelligent use of profiles can effectively extend data retention for unlimited periods – and hence special provision needs to be made to cover it.

4.3 Deletion and anonymisation

There is another key issue in relation to the deletion of data – the issue of anonymisation. There is a close relationship between the two, and as and where it is technically possible the right of data deletion could be augmented with a form of subsidiary right – the right to have data anonymised. The primary right would be to delete data – but in certain circumstances a data controller could offer the option to anonymise the data instead, if the data subject would be willing to let that happen.

The relationship between deletion and anonymisation is not a simple one. For one thing, it should be noted that if the right to delete is brought in, a data controller could avoid the possibility of that data being deleted by prior anonymisation – as the data would no longer be linked to an individual, no individual would have the right to delete it. Moreover, data is not always related to just one person – one clear example of this would be a group photograph in which a number of the people pictured are 'tagged'. That could bring a conflict of rights – if one person wants the data deleted but the other does not, whose rights have priority? Anonymity could apply here as well – in the photo example, it would be the tag that could be deleted rather than the photograph itself, effectively using the subsidiary right of anonymisation.

Even more importantly it must be remembered that anonymisation is far from a reliable process. Indeed, there is evidence to suggest that much

supposedly 'anonymised' data can be 'de-anonymised', by combining it with other, often public, data sources. In 1997 Latanya Sweeney demonstrated that by combining an anonymised hospital discharge database with public voting records a range of identifiable health data could be produced.[80] Computer scientists have continued to work on de-anonymisation – their models are getting substantially stronger and more applicable to the kind of data now being generated on the internet. In a 2008 paper, Narayanan and Shmatikov of the University of Texas demonstrated by combining the databases of Netflix and the online movie database IMDB that if you knew the county someone lived in and one movie that they had rented in the last three years, they could be uniquely identified 84% of the time. Moreover, they suggested that their results could be generalised – and applied to most other similar databases.[81] Work in this field has continued – and its implications are significant. At worst, it can be argued that anonymisation is simply an illusion[82] – and even at best it means that it needs to be considered very carefully and its weaknesses taken seriously.[83]

4.4 The virtue of forgetting

As noted above, the idea of a 'right to delete' should not be seen as a 'right to be forgotten' – there are technical, practical and emotional reasons to differentiate between the two. Nonetheless, there are still aspects of forgetting that are closely related and both important and beneficial. Viktor Mayer-Schönberger has written compellingly about the virtues of forgetting

[80] SWEENEY, L. 1997. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics,* 25**,** 98-110.
[81] NARAYANAN, A. & SHMATIKOV, V. 2008. Robust De-anonymization of Large Sparse Datasets. *IEEE Symposium on Security and Privacy.* 2008 ed.Available online at http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf
[82] As suggested, for example, by Michael Colao at a meeting of the Society for Computers and Law in March 2011. See http://www.scl.org/site.aspx?i=ne19845
[83] See for example the work of Paul Ohm, in OHM, P. 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review,* 57**,** 1701-1778. Ohm analyses the work of computer scientists from Sweeney onwards and suggests that a full understanding of the weaknesses of the anonymisation process is required if methods to protect privacy are to be effective.

in 'Delete'.[84] Perhaps most importantly in the context of this thesis, Mayer-Schönberger analyses how the developing 'default' of perfect digital memory takes control out of the hands of the individual, as their information and history becomes an indelible part of an mass of information usable and controllable by others. Moreover, it removes some of the positive effects of the passing of time. Digital memory can bring back information that has been forgotten for a reason, as part of the brain's method of navigating through life. As Mayer-Schönberger puts it:

> "…forgetting is not an annoying flaw but a life-saving advantage. As we forget, we regain the freedom to generalize, conceptualize, and most importantly to act."[85]

Mayer-Schönberger's analysis is deep and detailed, providing strong arguments in favour of forgetting, and against the ideas presented by Bell and others[86] that digital memory is a purely beneficial development. Furthermore, Mayer-Schönberger has suggested a solution to the problem of 'excessive remembering' by digital systems, the idea of expiration dates on information – as he puts it, 'reviving forgetting'. His suggestion is an ingenious and interesting way to find solutions to the problem, and in practice could have many benefits.

Establishing and implementing a right to delete could take it a step further, particularly from the perspective of autonomy, as it would put more control in the hands of the individual. The two could and should work together – implementation of expiry dates on certain forms of data would provide a kind of overarching control over data, while the specific right to delete would provide further autonomy and put further pressure on businesses to develop better, faster acting and more flexible business models. This suggestion will also be examined in more depth in Chapter Six.

---

[84] MAYER-SCHÖNBERGER, V. 2009. *Delete : the virtue of forgetting in the digital age,* Princeton, N.J., Princeton University Press.
[85] Ibid., p118
[86] In BELL, C. G. & GEMMELL, J. 2009. *Total recall : how the E-memory revolution will change everything,* New York, N.Y., Dutton. for example

4.5 The implications of the right to delete

The right to data deletion can be seen as a change in the focus of data protection – making it more about rights and principles of data subjects and less about a legal framework for businesses to work around, as it currently often appears to be in practice. It would work as an extension and better implementation of data protection principles: extending data access rights, and giving users a way to 'enforce' data minimisation.

Establishing a right of data deletion would have many direct implications:

1) It would give individuals the possibility of more control over their data and hence more autonomy
2) It could directly reduce the amount of data that is held – and hence that is vulnerable
3) It could force those holding data to justify why they're holding it – in such a way that the data subjects understand, for if data subjects cannot understand why the data are wanted, they might simply delete it. This, once again, would reinforce the positive aspects of the symbiosis on the web. If there is mutual benefit, and that mutual benefit is made clear, why would an individual wish to delete that data?
4) It would encourage the development of business models that do not rely on the holding of so much personal data.

This last point is the most important – particularly when considering the vulnerability of data, and when the right to delete data is considered together with the rights suggested in the previous two chapters, the right to roam with privacy and the rights associated with Collaborative Consent. If data gatherers have to explain the benefits of their gathering of data before those data are gathered, and data holders have to explain the benefits of their continuing to hold data if the data are not to be deleted, then a great deal of that data gathering will simply no longer happen. And, perhaps most

importantly, the pressure will be on the businesses to develop new models that no longer rely on the holding of so much personal data.

The ultimate weapon against data vulnerability is the elimination of that data. Where data do not exist, they are not vulnerable. If businesses no longer create so much data, or no longer hold so much data, there is less data to be vulnerable – and hence less vulnerability for individuals. How these business models might look is examined in Chapter Six – but it could represent a profound change in the nature of the commercial internet.

Chapter 6: A rights-based approach

The previous three chapters have examined particular issues concerning personal data, through looking at the way that these issues have played out in reality, using case studies and examples. Specifically, they have looked at how data are gathered and used, and the vulnerabilities of data when and where held – and a particular rights analysis has been suggested as a way of dealing with the issues and problems that arise. This chapter will now develop this approach further. First of all, it will examine how various rights work together to form the basis of an integrated right to use the internet with freedom. Then it will look at how these rights might be brought into play in reality. It will introduce a new concept, 'Autonomy by Design': a development of the idea of 'privacy by design', suggesting a route by which organisations gathering, using or holding personal data on the internet should approach their engagement with personal data. This will lead on to a look at how business models might function while following the new rights approach suggested here, something that is of great importance if the web symbiosis is to continue to be essentially beneficial. The chapter concludes by looking at the idea of a rights-based approach as a whole: why it is particularly suited to the online environment, and how, using the theory of symbiotic regulation, it might be able to produce real results.

1 Putting the rights together

Three rights have been proposed in the course of this thesis: a right to roam the internet with privacy in Chapter Three; a right to monitor the monitors in Chapter Four; and a right to delete personal data in Chapter Five. Though these rights are important individually, they really come into effect better when considered together. To understand why this is, and how the relationships between the three rights operate and are important, it is useful to divide engagement with personal data into three related processes: the gathering, utilisation and holding of data. These three phases are related and intertwined, as are the related rights: for data gathering, the right to roam

with privacy; for data utilisation the right to monitor the monitors; and for data holding, the right to delete. The three combine to provide a coherent rights approach to people's engagement with the internet.

1.1 Data gathering – and the right to roam the internet with privacy

The first, most direct, and potentially most important of the phases is the data-gathering phase. If the data gathering does not take place, there are no data to be utilised and no data to be held – so the issues relating to the utilising or holding of data do not come into play. For that reason, the right to roam the internet with privacy is not only the first of the rights to consider, but the most important, particularly from the perspective of autonomy, in the broader sense of the concept as discussed in detail in Chapter One.

The right, as set out in Chapter Three, would be a right to roam the internet, and use its fundamental tools without being monitored and without personal data being gathered. The key point of this right is that following the change of paradigm suggested in Chapter One and throughout the thesis, the default position is that data are not gathered for storage and use unless an express choice has been made, for example by logging in to a premium, specified service. That then leaves the questions of what 'roaming' the internet means, and what can be considered the internet's fundamental tools.

So far as what constitutes a 'fundamental tool of the internet' is concerned, the starting point is to look at current habits and market shares – at present, for example, search engines are amongst the fundamental tools of the internet, and ISPs are required in order to roam the internet. This right, therefore, would require that search engines would effectively be required to offer their basic services in an alternative, 'privacy-friendly' form. For Google, for example, rather than just the one 'search box', there could be two: Google might call them 'basic search' and 'super search' – super search allowing

Google to gather data but also allowing them to tailor results and to offer 'better' services.[1]

With other tools, there are more questions to ask – to start with, how many people are using the services (both in absolute and market-share terms) and for what purpose. Where use of a particular service becomes 'the norm' rather than the exception, that service could be said to be fundamental – and when there is doubt, the benefit of that doubt should lie with the individual. For some websites, it would depend on whether the user was 'signed in' or not, and the kind of data that could be acceptably gathered without explicit consent would vary. One example of this would be online retailers such as Amazon, who could (and would) gather actual transaction data, which they need for financial records, but would not be permitted to gather clickstream or 'browsing data' linked to the user from those visiting their websites unless the user had signed in. For other not so directly commercial websites – news sites such as the BBC, for example – unless some kind of sign-in process happens, clickstream or browsing data could only happen in a strongly anonymised form. When the users sign in, that begins the consent process – and the second of the rights discussed in this paper kicks in: the right to monitor the monitors.

The paradigm shift for which this thesis is arguing would mean that unless there is a good reason for tracking – that is, unless tracking is a fundamental part of how the service functions (as opposed to how the *business model* for the service functions) then it should be possible to use the service without tracking. A news or information website, for example, has no 'need' to track or to gather information, while a social networking service works on the basis of membership and linking between members, so would require that information to be gathered and used.

Moreover, as discussed in Chapter Two, as systems for profiling develop, tracking has the potential to become more pernicious – the extreme

[1] As discussed in Chapter Three

examples of things like 'whites-only websites' give some idea of why, at least in general, the idea of a right to roam with privacy should have general rather than specific application, and why tracking should be the exception rather than the rule, particularly if the internet is viewed as a 'public' resource. The idea of a 'no blacks' sign in a shop window is universally condemned in modern societies – a right to roam with privacy would provide protection against similar, hidden but equally pernicious practices online. More directly, policies like price setting based on profiles could be curtailed.

The way this might function in practice will be looked at later in this chapter – but it is important to understand that the right to roam with privacy is intended as a principle and a guide: how it will apply in practice is something that will develop over time as habits and expectations develop. As shall be shown later in this chapter, there are a number of possible business models that could function effectively whilst still satisfying the new rights. It is also important to understand that it is possible to have 'membership' or 'paid for' services (like, for example, the new subscription version of the websites of the Times newspapers[2]) that generate income without tracking or gathering information about the user. Once entry to the 'private' section of the website has been made, there needn't be anything that would require the website owner to track the user – so the right to roam with privacy can still function in that kind of a scenario.

A right to roam the internet with privacy is a radical suggestion, and would require a significant change in the approaches of search engines, ISPs and other web-providers – and in governments' approaches to security, as it would be directly contradictory to the current policy of data retention (as described in Chapter Three) and all that surrounds it, amongst other things. It is, however, something that could ultimately be of benefit to all. It is also a part of the process of changing the paradigm of internet use so that privacy is the default, and that surveillance and data gathering is the exception – either opted into or brought into play when needed for security or other similarly

[2] http://www.timesplus.co.uk/welcome/index.htm

important reasons. Those exceptions could include security exceptions – security services could for example monitor particular users (even particular ISPs) or particular websites or services, when they have reason believe that monitoring those users, sites or services would be useful and proportionate. Similarly, commercial web services would need to record certain transactions for legal purposes, and record certain data for reasonable economic purposes – but again, those purposes would need to be justified and specified.

1.2 Data utilisation – Collaborative Consent and the right to monitor the monitors.

As and when data gathering is to happen – when it is consented to – more control and understanding of the process needs to be provided to those who are being monitored, those about whom data is being gathered. This control needs to cover not simply the data gathering process but the data utilisation process: people have rights not only over when, how and where data is gathered about them but when, how and where it is used. The right to monitor the monitors – and the concept that accompanies it, Collaborative Consent – is the start of a process through which that control can be made possible. The key to the right and its underpinning concept is another aspect of the paradigm shift suggested throughout this thesis. In this case, it translates to an understanding that the consent that is given to a data gatherer is a privilege – and that it should be taken seriously, both in terms of ensuring that the initial consent is one that is informed and understood, but also that it is a consent that can be modified or withdrawn at the will of the individual concerned, as their understanding develops, as their knowledge increases and as their views and opinions change.

As set out in Chapter Four, Collaborative Consent has two key aspects. Firstly, it treats consent not as a discrete, one-off decision but as a process, and secondly it looks at consent as a two-way agreement – so that the consenter is allowed and enabled to see to what they have consented; to monitor,

modify or withdraw that consent in real time, and where the enterprise seeking the consent must communicate and collaborate with the consenter not just at the start of the process but throughout. That means not just when the data is gathered, for example while browsing the web, but also when it is used. That might mean when an advertisement targeted based on gathered data appears, or when tailored content or recommended links appear, or when the price offered for goods or services is similarly tailored for an individual. Collaborative Consent requires a two-way process, a form of dialogue between the enterprise and the individual. The internet provides the kind of medium for immediate and interactive communication that allows such a process to be possible.

From the perspective of the Symbiotic Web, as developed in this thesis, Collaborative Consent has a crucial advantage: if those who are monitoring and targeting people require continued consent from those being monitored and targeted, they will need to communicate the benefits that those being monitored and targeted are getting. That in turn means that they first need to ensure that a benefit really exists, not just in the minds of the providers, but also a benefit that the user can understand and appreciate.

Once the data has been gathered, the third of the rights set out in this thesis comes into play. Rights are needed not only over what happens to personal data as they are gathered and processed, but also over what happens to that data after they are gathered. For that, one further fundamental right is required, the right to delete data.

1.3 Holding data – and the right to delete

As shown in detail in Chapter Five, it is axiomatic that wherever and however data are held, they can be vulnerable, whether to technological failure, human error, human malice, business pressures, political pressures, hacking, leaking, theft, legal loopholes or a wide variety of other risks. Ultimately, the only way to ensure that data is not vulnerable is for it not to exist at all. That

is the starting point when considering the holding of data – and the first part of the reasoning behind the suggestion of a right to delete.

The second part is one of the key questions asked throughout this thesis: is 'personal' data in any sense 'ours'. If the answer to this question is positive, as this thesis suggests, then what needs to be considered is what kind of rights individuals have over data that can be related to them. As for the issues of data gathering and processing, what is really needed is a paradigm shift – a change in assumptions. The assumption should be that unless there is a strong reason for data to be held, data shouldn't be held. Data holders should need to justify their holding, rather than the other way around. It can be looked at as an extension to the logic behind the concept of Collaborative Consent. Consent needs to be continually renewed – but this time it is not just for the gathering of data but also for holding data.

This shift of assumptions is taken a step further by the establishment of the general right to delete data. It is a shift in defaults: rather than asking when people should have the right to delete personal data, the opposite should be asked: in what circumstances and with what kind of data should people not have the right to delete? This follow a similar logic to the exceptions to the right to roam the internet with privacy – the five principle categories of reasons are: paternalistic, communitarian, administrative, archival and security-based. As noted in Chapter Five, it should be specifically stated that 'supporting your business model' should not be a sufficient reason to deny data deletion. The exceptions that are suggested should ensure that the right remains in balance – indeed that it supports the positive balance that underlies the concept of the Symbiotic Web.

In the end, this idea boils down to ensuring that those holding data understand that holding data connected to an individual is a privilege, and one that can in most circumstance be revoked at any time. The organisations holding data should be asking the individual 'can we keep your data please?'

as a real question, asked in a real way, and with the answer given being properly respected.

1.4 Three rights together – protecting autonomy, balanced with the needs and rights of others

In summary, then, for the three phases – the gathering, utilising and holding of data – three respective rights are proposed: the right to roam with privacy, the right to monitor the monitors, and the right to delete. The right to roam with privacy is the overriding right. The right to monitor the monitors kicks in when that initial right is waived. The right to delete applies to the data once they have been gathered. The three are intended to work as a coordinated set, with the intention of moving to a situation where the gathering, using and holding of data is seen as a privilege rather than the default.

The crucial point here is that the rights are intended to protect autonomy. They are intended to preserve such autonomy as already exists in the internet, protecting it from the emerging threats discussed in the thesis – both the theoretical threats suggested by the model of the Symbiotic Web in Chapter 2 and the practical threats discussed in the case studies in Chapters 3 to 5.

These rights are not what might be called 'fundamental' rights, but rather are 'mechanistic', designed for the world as it now is, and for the relationship between the online and 'real' worlds as they are currently developing. They can be seen as both positive and negative rights – freedom *to* roam and use the internet and at the same time freedom *from* surveillance and from manipulation. They are alienable rights – the alienability of the right to roam with privacy is critical to the way that it works, and to the way that the internet works in its symbiotic form.

They are rights held in balance with the rights and needs of other individuals, of governments, and of businesses. It is not in any way the intention for the rights to stop the gathering, using or holding of data, for that gathering, using and holding of data is what fuels the current, beneficent symbiosis and provides individuals with so many useful, entertaining and enlightening tools and services. Governments have the right to do what is needed to preserve security, to encourage and support business and so forth. Businesses need the freedom to do what they need to do in order to function. The needs and rights of all must be balanced. The rights suggested in this thesis are not intended to override the rights of others. Rather, they are intended to shift the balance of control more in favour of the individuals – to give them more autonomy within the process – and thus give the symbiosis the best chance of continuing to grow and develop in positive ways. The rights are intended to put some of the most important aspects of the data protection regime – most directly access and minimisation rights – more into the hands of the individuals.

That it is the function of these kinds of rights regimes to maintain a balance has been played out in recent cases. In first Scarlet v SABAM[3] and then SABAM v Netlog[4] the European Court of Justice ruled against the imposition of general filtering mechanisms to screen for copyright infringing materials, first on an ISP and then on a social media provider. The balance between individual rights to privacy and freedom of expression and intellectual property rights must be maintained: as the press release from the European Court of Justice on SABAM vs Netlog put it:

"Such an obligation would not be respecting… …the requirement that a fair balance be struck between the protection of copyright, on the one hand, and the freedom to conduct business, the right to protection of personal data and the freedom to receive or impart information, on the other."

[3] See http://curia.europa.eu/jcms/upload/docs/application/pdf/2011-11/cp110126en.pdf
[4] See http://curia.europa.eu/jcms/upload/docs/application/pdf/2012-02/cp120011en.pdf

1.5 How the rights might apply

How might these rights change current services on the internet? Looking first at a search engine, the right to roam with privacy would mean that search engines' 'basic' services should be available in a form that does not track or record search terms linked or linkable to the searcher. A search engine provider could offer an alternative service – they could label it a 'super search' and explain that data would be gathered but that the service would be 'better' in the sense of more tailored. At that point, the right to monitor the monitors would kick in, and they would have to provide regular notifications as to when data was gathered, label any advertisements that were targeted or any content that was tailored and so forth – and also provide real-time access to all data held, including, following the right to delete, an easy and simple way to delete those data.

Let us consider the example of a shopping website. The right to roam with privacy would mean that it should always be possible to 'browse' a shopping website without data about that browsing being gathered. Individuals should know that their browsing is private unless they have signed in and committed themselves to having that browsing monitored. Real transaction data would still be gathered, and the services would be free to make recommendations and so forth based on those real transactions – which is effectively what Amazon.com does when it makes personal recommendations. The right to monitor the monitors would only apply when and if a customer signs in and agrees to have their browsing monitored – it would have no effect on the transaction data, which is required to be held for administrative reasons. Similarly, the right to delete would only apply to data like browsing or related data, not to transaction data. If the shopping site also uses advertisements, where those advertisements are targeted based on monitoring, just as with search engines, those advertisements would need to be clearly labelled.

Looking next at a news or information website, the right to roam with privacy would apply in that once again it should be possible to surf without data being gathered. That should mean that even for 'paid' services like the Times+ subscription service, it should be possible to surf with privacy. Once again, it might be possible to have 'super' services that allow browsing data to be gathered – perhaps with fewer ads on the page, taking a leaf out of the book of the 'apps' market on smartphones. Here, apps are often available in two forms, one free but where there are advertisements cluttering the small screen of the smartphone, the other costing money but 'ad-free'. The right to monitor the monitors would apply throughout – and again, the right to delete should be available simply and easily.

The area where the rights would be at the same time the simplest and the most controversial is when applied to ISPs. In its essential form, the right to roam with privacy should apply to ISPs across the board, preventing the kind of universal data gathering and use suggested by Phorm's Webwise model described in Chapter Four. That would directly contradict the essence of the kind of general data retention regime currently popular with governments – but that is unavoidable, and though a big political challenge, from the perspective of privacy and autonomy it would be highly desirable.

2 Autonomy by Design

The idea of 'privacy by design' was conceived and developed by Dr Ann Cavoukian, Ontario's Information and Privacy Commissioner, in the 1990s,[5] but has since become widespread, though generally as an aspirational tool rather than an idea with any legal force. In the UK, the Information Commissioner's Office has been pushing it as an approach to data protection since November 2008. As the ICO puts it:

> "Privacy by Design is an approach whereby privacy and data protection compliance is designed into systems holding information

[5] See http://www.privacybydesign.ca/about/history/

> right from the start, rather than being bolted on afterwards or ignored, as has too often been the case."[6]

The Privacy by Design programme is intended to 'encourage public authorities and private organisations to ensure that as information systems that hold personal information and accompanying procedures are developed, privacy concerns are identified and addressed from first principles.'[7] As set out in the report, the concept of Privacy by Design is a good one, though its strengths and weaknesses reflect those of the ICO in general, and indeed those of the data protection regime as a whole. The principles are essentially good, but it suffers from working largely through 'recommendation' rather than having much real force, it places more emphasis on encouraging the flow of data than it does on individual rights, and it appears not yet to have as much effect in reality as it does in theory.[8] The concept, however, is a good one – the idea that privacy should be planned for from the very first, and that it should be taken into account through all phases of the data cycle.

2.1 Autonomy by Design: the change of paradigm

Looking at the three rights set out in this thesis, and taking the idea of privacy by design as a starting point, a new concept can be set out: 'Autonomy by Design'. Privacy by design was all about data management under what could be described as the 'old' paradigm: Autonomy by Design is about proper 'data management' under the 'new' paradigm. The focus should be on the individual – and on managing the data from the point of view of that individual and that individual's rights. As data are gathered in they should be put into a format that helps the individual at the same time as they are put into a format that suits the business. What suits the individual?

---

[6] In http://www.ico.gov.uk/news/current_topics/privacy_by_design.aspx

[7] ICO 2008a. Privacy by Design Report. p2. The full report is downloadable from http://www.ico.gov.uk/upload/documents/pdb_report_html/privacy_by_design_report_v2.pdf

[8] The title of a March 2011 event organised by the Society for Computers and Law, 'Privacy by Design: 'Grand Design' or 'Pipe Dream'?' gives some indication as to the doubts that exist about the effectiveness of Privacy by Design in reality. See http://www.scl.org/site.aspx?i=ev18730

Knowledge, understanding and control – which means that the individual has to be able to know what data are held, understand why they are held and how they are used, and then be able to control all of those things: control what is gathered (via the right to roam with privacy), control how it is used (via the right to monitor the monitors and Collaborative Consent) and control what is held (via the right to delete).

If data are put into this kind of a format from the moment they are gathered, and mechanisms are put into place to allow both access and control over that data in real time, then all the three rights can be promulgated/brought into play together. The same systems can allow the user to turn the data gathering off, to look at and modify what data are being gathered, and what data have already been gathered, and to delete the data that are being held. This is taking the concept of 'privacy by design' to a new level: it can be described here as 'Autonomy by Design'.

Phorm's Webwise – and all behavioural targeting systems – show that it is possible (and part of the point) to make instantaneous use of data, if the software that does the work is designed accordingly – why not use that kind of approach to analyse and prepare the data for presentation and use by the individual at the same time? If it can be done for the benefit of the business, it can be done for the benefit of the individual – and using the logic of the Symbiotic Web, doing so for both is more likely to produce a positive outcome for both. In the ICO's 'Privacy by Design Report' it was suggested that organisations should 'incorporate subject access request (SAR) functionality at the design stage… [it is] in the interests of both organisations and individuals to ensure that systems are designed to automatically service SARs.'[9] Autonomy by Design would take that to another level. Both the business and the individual would benefit in the medium to long term if systems are designed from the outset not only to service the relatively bureaucratic and essentially legalistic process of a SAR, but to give the individual real and practical access to their data, in real time, together with

[9] ICO 2008a. Privacy by Design Report. p24.

the ability to correct, control and delete that data, and indeed to manage the data gathering process (including turning it off). As noted in Chapter Four, Google's developing 'dashboard' system shows that this would be possible, and hints at how it might work in practice.[10]

2.2 Autonomy by Design: user-friendly rights

A crucial requirement of all these rights is that they would need to be rights that can genuinely be exercisable – that the rights are both practical and user-friendly. That is one of the key features of Autonomy by Design. At present, it is generally far from the case, whether it is access to data (as in the use of the Google dashboard or Google ad preferences) or the deletion of data, as those who try to delete their accounts from Facebook, for example, will have experienced. There is no reason for this not to be possible – organisations like Google and Facebook have the highest levels of expertise in making all kinds of services accessible, user-friendly and attractive. If the will was there, they could do the same with the deletion of data. Combining the kinds of systems required for 'monitoring the monitors' discussed above, with an ability to immediately delete data would be a simple way to make this possible.

What's more, with these rights in place, it would put pressure on enterprises to apply their considerable skills to find ways to present the data that they gather in user-friendly ways. If the user has the option to prevent data gathering (as the right to roam with privacy would require) and to delete any data that has been gathered, then if they are presented with something confusing or over-complex they are likely to exercise those negative options. It would be in the interests of the business to make the process as user-friendly and easily understood as possible – and indeed, as quick and slick as possible, so that the user moves quickly through the processes and on to the real uses of the systems and services, the uses that make money for the enterprises. Autonomy by Design, in a user friendly form, would satisfy the

[10] https://www.google.com/dashboard/?pli=1

requirements of both businesses and individuals, and could begin to provide the kind of systematic underpinning of a future, autonomy- and privacy-friendly internet.

For the rights to have this kind of pressurising effect, they would have to be backed by appropriate legal tools, enforcement mechanisms and penalties – the kind of substantial fines that are now exercisable by the Information Commissioner in the UK, or have been applied by the European Commission against Microsoft for anti-competitive practices. It is likely, for example, that the substantial fines applied against Microsoft played at least some part in persuading Google to give way (insofar as it did) to the Article 29 Working Party over data retention periods, as examined in Chapter Three. A general understanding and acceptance of these rights, however, makes the establishment of those legal mechanisms more possible. That is part of the point of the rights in the first place.

Just as for Collaborative Consent, Autonomy by Design is ultimately about harnessing the strengths of the internet – its immediacy, its interactivity, its powers of analysis and so forth. Businesses are continually striving to harness these strengths for their own benefit – what Autonomy by Design would do is harness them for everyone's benefit, and though that may not immediately appear to be as attractive to businesses as something more focussed on their short-term bottom line, in the end they should benefit. Once more, this is the logic of the Symbiotic Web coming into play: for when businesses and individuals both benefit, that benefit is more sustainable.

Autonomy by Design can be seen as a form of multifaceted regulation – in Lessig's terms, it works by using code, law, norms and markets at the same time. More accurately, as shall be set out at the end of the chapter, if it should come to pass it would be best achieved through symbiotic regulation. In the end, it will work as and when the different pressures in all those fields make it desirable. Here there are lessons to be learned from the field of online banking. In the early days of the internet, the idea of making payments online

was considered highly risky and was discouraged. Now it is mainstream. This transformation has taken place as a result of a combination of the needs and desires of people to do business online, of banks to take advantage of these desires, and of the abilities of the legal and technical communities to find functional solutions to the problems. Making secure payments online works because of the technology, the legal infrastructure, and the way that people have been persuaded that it is safe – the code is there, the law is there, the market is there and the norm has been established.

The banking example is just one example of 'security by design'. The ICO Privacy report analyses how security by design was effectively introduced and how the lessons from that introduction can be applied to privacy by design. It suggests five key factors: understanding the threat, management standards, executive awareness, language and frameworks, and organisation and responsibilities.[11] Once again, the report is good as far as it goes, indicating the technical and organisational challenges – some of which, the organisational ones in particular, would apply in an even more significant way to the introduction of Autonomy by Design. These key challenges could be better met with the introduction and acceptance of rights such as those suggested in this thesis. Well-communicated rights would help the management of businesses to understand some of the threat, and to improve executive awareness. Well-written rights could help provide the language and framework, and help frame the management standards, as well as provide motivation and direction towards setting up appropriate organisational structures and setting appropriate responsibilities. Autonomy by Design and the rights set out here work hand in hand: Autonomy by Design would be a way to enable organisations to meet the challenges of the rights, while the rights would also help support the establishment and acceptance of Autonomy by Design.

[11] ICO 2008a. Privacy by Design Report. p23

2.3 Business in the new internet

Establishing and instituting rights like these would place burdens on businesses. They would be required to create systems to allow individuals knowledge of the monitoring and access to their data, in an understandable and user-friendly form. They would have to allow individuals to control and even delete data that they, the businesses, could in the past have held and used as they wished, and benefited from both financially and strategically. Are these unnecessary or disproportionate burdens? On the surface, and from the businesses' point of view, it might seem so – but if the gathering, processing and holding of data about individuals is seen as a privilege rather than a right, and a privilege that must be earned, it is neither disproportionate nor inappropriate. Putting these kinds of systems in place is part of the price that must be paid to earn that privilege – and any business model that relies on deceiving or taking advantage of individuals is not a model that a good society and good legal system should support. It should be entirely possible to develop business models that respect these rights – and respect both customers and others from whom data might be gathered. There are many possible approaches, some of which are set out below.

Model 1 – the anonymous, non-data gathering model

In this model, data are not gathered either about the person browsing or their activities. At first glance this could be viewed as representing the "old style" internet, as it existed before the Google/Facebook business models transformed the internet, but it is still a model that could function effectively. Moreover, not all of the advances made in the development of business models for the internet rely on tracking or gathering of data. Advertisements, for example, can be 'tailored' or targeted in some ways without collecting data or monitoring the individual. Tailoring based on the route from which the individual arrived, for example, would not breach privacy, and maintain all the suggested rights.

Examples of how this might work for search engines were set out in Chapter Three – but the application is broader, and there is still potential for further development. There is a significant amount of additional information available for tailoring without having to infringe on the browser's privacy or record any data – and for much of the internet (for example news and information sites, blogs and wikis for reading rather than contributing and so forth) this is the model that should be prevalent.

Model 2 - the instant tailoring model

In this kind of model, data would be gathered "on the fly", results tailoring and displayed, and all data deleted as they are used – learning from those positive features of Phorm's Webwise that led to 80/20 Thinking to produce an essentially positive privacy impact assessment of their business.[12] Systems using instant tailoring would have to be clearly signalled, and would have to find positive ways to deal with the consent issue – for example by using Collaborative Consent in a simple form, notifying and allowing opting out of targeting. It should, however, be possible for them to function, and the failure of Phorm should not be seen to be the death knell of instant tailoring. As shown in Chapter Four, the failure of Phorm was brought about by many factors, and the technological elements were far from decisive.

This kind of a model would be suitable for more commercially focussed websites than those using the anonymous, non-data gathering model. It should also be remembered that in this model there are other tools available: anonymisation and aggregation. Once again learning from some of the positive elements of the Phorm model, it is possible to gather data and use it for analysis without

[12] See Chapter 4, Section 4.3

identifying the individual involved, either as part of a quantitative analysis of behaviour of visitors to a website or users of an internet service (via a form of aggregation) or as part of a profiling system that identifies a type of user without linking that type to a specific individual (a form of anonymisation). As noted in Chapter Five, anonymisation is a tool to be used with caution, as unless sufficient care is taken it can be possible to 'de-anonymise' records.[13]

Model 3 - the "instant collaboration" model

Instant collaboration would mean a kind of 'mini-membership' system, whereby a user could enter into a short and most importantly very temporary relationship with a service, via something like a simple online form. Once again this would be learning from the positive aspects of Phorm, but this time suggesting the establishment of an immediate relationship with the surfer, a relationship based on collaboration and understanding. The collaboration would be immediately terminated if such a service was just browsed away from or ignored, and in those cases data would need to be immediately and automatically deleted.

One key point here is that the idea of instant collaboration should be one that grows as people begin to understand the way that the web symbiosis works – if people become more aware of the way that data are gathered and used on the internet, they should become more aware of the reasons for this kind of collaboration, and the kinds of benefits that it can provide. They will also become more accustomed to the practice, and more willing to participate – and the providers should become better at implementing the systems, making them more user-friendly and attractive. If the web symbiosis is to continue to grow and develop positively, that overall level of awareness and understanding must increase – that is part of the key to a more

[13] See Chapter 5, Section 4.3

autonomous internet. An example of the kind of website that might use this model would be a 'casual' shopping site – one which relies principally on one-off or irregular customers. As noted for the 'instant tailoring model' above, the tools of aggregation and anonymisation, would be possible in this model, and with similar caveats.

Model 4 - the long-term collaboration model.

Here, a user would establish some kind of permanent membership with a service or site. This is essentially another 'traditional' model, but in a much more collaborative and controllable form. Access to data would have to be provided, as well as choices about how it is used, simplification and so forth - and immediate ability to delete all the data, really, and effectively. Social networking services are perhaps the best current examples of the kinds of systems that would use this sort of a model – though the simplicity and clarity of how they work in relation to surveillance and personal data would have to be significantly changed. This is an area where Autonomy by Design should come into its own.

2.4 Leaner and more efficient businesses

Embracing these rights would encourage efficiency and 'leaner' business models. That leanness and efficiency would cover not only the data – for as noted particularly in Chapter Five, the idea of taking data minimisation seriously should be fundamental to the future internet – but the systems and interfaces around data. The volume of data would need to be controlled, the forms in which they are stored would have to be more systematic and efficient, and access, manipulation and control over the data would need to be faster, cleaner and clearer. It should encourage more efficiency, which should help and support businesses that do it well. If done without the negative sides demonstrated by examples like Phorm and Beacon, and with more transparency, consultation and coherence, it could provide something

new and different. The innovative skills of those who have developed businesses like Google and Facebook should be able to create similarly innovative and imaginative ideas that could function in this new environment. Indeed, ideas and businesses that could thrive – and take the internet onto a new, far more autonomous level.

3 A rights-based approach?

One question underlying all this work is whether, in the first place, a rights-based approach is appropriate or likely to be effective. Further to that, is it not true to say that the current European data protection regime is 'rights-based' anyway? Data protection has been discussed in some depth in Chapter Two. In principle it is 'rights-based', at least in the sense that its origins include Article 8 of the European Convention on Human Rights, the right to respect for privacy (embracing a right to a private life), which is mentioned in the preamble to the Data Protection Directive.[14] In practice data protection is more about the regulation of data flow than the protection of individuals' privacy, and it is treated to a great extent as a piece of technical legislation to be complied with rather than as a statement of rights and principles. Whether that changes as the Directive is reviewed has yet to be seen – the review process is currently under way, but is far from complete. The current timetable suggests that the revised Directive will be published late in 2011.[15] It appears unlikely, however, that fundamental changes will be suggested, and so it remains probable that the focus of data protection will remain, as its name suggests, on the data, rather than on the individual.

In the UK, the Information Commissioner's Office has produced a comprehensive report to the Ministry of Justice about the current state of the data protection regime, and how it sees the future of that regime, particularly

[14] Paragraph 10 of the Preamble to the Data Protection Directive – Directive 95/46/EC, available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML
[15] See for example http://www.out-law.com/page-11292

in terms of the current review of the Directive.[16] While in most ways it is a good report, highlighting the successes and failures of the regime and providing positive recommendations throughout, it still places the emphasis on the data flow, with the protection of individual rights simply something to take into account when encouraging data flow. In the online world as it exists now, it does not seem that data flow needs much encouragement – the opposite, as the various case studies in this thesis have shown, and as the imperatives of the drives within the Symbiotic Web would suggest. Though the data protection regime has some of its roots in rights, its realities are more focussed on economic drives – rights are treated more as a qualifier, an influence, but not as the backbone of the regime.

So long as the primary focus remains on economic success, privacy and autonomy are likely to be squeezed, as so much of this thesis has shown, from the Google data retention story in Chapter Three to the numerous examples of data vulnerability examined in the Chapter Five. A more direct and genuinely rights-based approach would put that focus back on the rights of the individual. It would look at the issues from the perspective of the individual, and how the individual experiences things, how the individual is affected by events, and how that individual can understand those events, rather than the precise and technical details of what may or may not be happening to particular pieces of data.

What is more, one of the implications of the current approaches to data, privacy and related areas is that the law (and in particular data protection law) tends to be very technical and opaque. This is another reason for a rights-based approach. Where there is confusion, conflict or lack of clarity in law then what is needed is something to fall back on, to guide, and to set principles – and confusion and lack of clarity has been central to many of the case-studies set out in this thesis.

[16] ICO 2010. Response to the Ministry of Justice's Call for Evidence on the current data protection legislative framework.

3.1 Technological neutrality

There are further reasons why a rights based approach is particularly appropriate in this field. The first is that of technological neutrality. A rights-based approach can get closer to being 'technologically neutral' than a more legalistic approach. Looking from the perspective of the individual and their experiences rather than in detail at a particular form of technology gives more chance of setting principles that can be applied when technologies develop. When working solely with law, approaching technological neutrality is possible, but has pitfalls: if the law is too specific, it can be sidestepped, whilst if it is too general, it is hard to apply. As new technologies emerge, or new ways of using data are developed, well-expressed rights are not as easily sidestepped, neutered or avoided as are more specific laws – so the combination of rights setting the principles and specific laws used for enforcement is one that has a better chance of success.

The problems surrounding technological neutrality can be further examined by looking at some particular examples:

a) Phorm's Webwise

Phorm's 'Webwise', as discussed in Chapter Four, was at least arguably compliant with data protection legislation in the UK because of the way that it used its UID system to avoid storing what could technically be described as 'personal data', though in the eyes of the public it seemed to infringe on privacy, and 'should' have been illegal in terms of the effect of what it did rather than the technological specific. It could be said that Phorm's Webwise breached peoples' rights, even if it might not have broken the law – and as noted in Chapter Four, the EU is still pursuing the UK government over its implementation of the Data Protection and e-Privacy Directives in relation to Phorm.

b) Flash Cookies

'Flash cookies' provide another example of the importance of technological neutrality, and the advantages of a rights-based approach. Effectively, flash cookies could be used to sidestep legislation that targets 'traditional' HTML cookies – though flash cookies might end up performing even more intrusive functions, amongst other things having no set expiry date and being able to regenerate previously deleted HTML cookies. As security expert Bruce Schneier put it:

> "Unlike traditional browser cookies, Flash cookies are relatively unknown to web users, and they are not controlled through the cookie privacy controls in a browser. That means even if a user thinks they have cleared their computer of tracking objects, they most likely have not.
>
> What's even sneakier?
>
> Several services even use the surreptitious data storage to reinstate traditional cookies that a user deleted, which is called 're-spawning' in homage to video games where zombies come back to life even after being "killed," the report found. So even if a user gets rid of a website's tracking cookie, that cookie's unique ID will be assigned back to a new cookie again using the Flash data as the "backup.""[17]

There are three linked points to consider here, all of which have more general application than just for flash cookies. Firstly, that legislation specifically covering HTML cookies would not cover flash cookies, so flash cookies could be used in the place of HTML cookies and avoid that legislation. Secondly, that technology itself can be used to avoid specific legislation. In this case, if legislation required an HTML cookie to expire after a certain time

[17] See Schneier's blog at http://www.schneier.com/blog/archives/2009/08/flash_cookies.html

or after a particular event (such as logging out of a service), a flash cookie could be used to reinstate that cookie, again avoiding the legislation. Thirdly, that technology can also be used to avoid technological controls – in this case, that the flash cookie avoids the privacy controls (and in particular the privacy settings on browsers) used to govern HTML cookies.

The starting point for dealing with all three of these points could be looking at the process from a rights-based perspective. The rights establish the principles, and the legislation could (and should) follow the principles that have been established. As technology develops, specific pieces of legislation will require amendment, updating or replacing, but if the principles are set out in terms of rights, those amendments will be easier to establish and to understand, and, potentially at least, easier to put through legal systems. Flash cookies are only one example – but they do demonstrate the problem that emerging and developing technologies can provide for the law.

c) HTML5 – and future web technology

Another pertinent example is the ongoing development of the next generation of HTML (HyperText Markup Language) – the technical language in which web pages are written. The newest version, HTML5, is currently under development, and is expected to improve many aspects of the way that the web works, integrating multimedia more effectively. It may also make it easier for individuals surfing the web to be tracked. As reported in the *New York Times* in October 2010:

> "The new Web language and its additional features present more tracking opportunities because the technology uses a process in which large amounts of data can be collected and stored on the user's hard drive while online. Because of that process, advertisers and others could, experts say, see weeks or even months of personal data. That could include a user's location, time zone, photographs, text from

> blogs, shopping cart contents, e-mails and a history of the Web pages visited."[18]

As commentators have pointed out,[19] and as the examples described elsewhere in this thesis have shown, much of this has been possible with existing technology – indeed, much of it has already been done with existing technology. What is important about the way that it may be incorporated directly into HTML5 is that it could make it both easier to do and more 'mainstream'. Indeed, if it is built into the standards through which websites are created, full-scale detailed tracking could become the default position.

The development of HTML5 is close to completion, though the precise date when it becomes 'the standard' is something that is far from certain. Parts of it have already been implemented, but full implementation is likely to take a number of years. The W3C, who have been developing the specification, issued their 'Last Call document', summarising the final issues remaining to be resolved, in May 2011,[20] with a ten week initial review to follow.[21]

How long it takes to become a de facto standard is yet to be seen. What is clear is that web technology is moving forward, and part of that move forward may be significant impacts on privacy and autonomy – and that those who are concerned about privacy and autonomy need to find a way to have their voices heard, and taken account of, in the nature of the developments of this technology. This is where rights can make a difference. They can add to the strength and in particular the coherence of the voices making claims about the need for privacy and autonomy. Rights in place (using the Autonomy by Design approach) could help find more appropriate uses for technologies like flash cookies – and could, if accepted and

---

[18] http://www.nytimes.com/2010/10/11/business/media/11privacy.html?_r=1&adxnnl=1&adxnnlx=1286794927-imUlp74yLiTLxFEIpNQWmA

[19] For example, see blogger Space Ninja's comments at http://spaceninja.com/2010/10/html5-nyt-privacy/

[20] See http://lists.w3.org/Archives/Public/public-html/2011May/0162.html

[21] A first draft edition of HTML5 for web authors was issued in August 2011, too late to be properly considered in this thesis. See http://www.w3.org/News/2011#entry-9169

understood, shape the development of technologies like HTML5 and in particular its successor technologies.

3.2 Rights, business models and jurisdictional issues

What is even more important is that if clear and well-understood rights can be established, they can play a key part in helping businesses to develop more positive business models. They can provide clear guidance to businesses that some ideas that they might be considering would be inappropriate and ultimately both illegal and unsustainable. The case of Phorm is one of the most direct – as noted in Chapter Four, if clearer, more explicit rights had been established and set out beforehand, the people who supported Phorm might have had a much better idea that Phorm would ultimately fail, and been able to avoid the whole farrago, with all its damaging implications.

As the concept of the Symbiotic Web suggests, the business model issue is crucial. To a great extent business models are what drive the development of the internet and all the commensurate benefits and technological advances that follow from it. Google and Facebook are perhaps the most obvious and dramatic recent examples – their ways of doing business have transformed the internet. Future developments are likely to be similarly inspired and shaped: if the future of the internet is to be more 'privacy-friendly' it will be through the development of privacy-friendly business models. Coherent, well-established rights could be a key tool to help businesses to develop those models.

Another advantage of this kind of an approach is that it has the potential to overcome, at least to an extent, the issue of jurisdiction. Chapter Three provided a prime example of how this works, with the way in which Google changed it policies on data retention in response to pressure applied by the

EU Article 29 Working Party.[22] Google applied those changes worldwide as it prefers to have global policies and practices as it considers itself a global organisation. The impact of this is that though US users (for example) of Google do not have the benefit of European data protection, they have benefited from much reduced retention of their search data. Google is just one example, albeit probably the largest. In the internet as it is currently developing global businesses are increasingly prevalent, from fields like search (Google) or social networking (Facebook) to those who create the software with which people browse the internet or even create the web pages that make up the internet. Influencing the practices of those global businesses is perhaps the most effective way to influence the environment of the internet itself.

Moreover, developments at most levels of the internet are essentially global: the standards set for the protocols and languages used in creating web-sites are global, the hardware used for PCs at least to some extent follow global standards (as do the operating systems) or are even made by companies marketing the same products worldwide, the software used to browse the web is generally the same the world over. If those products are built or written to work in places where there are high standards and requirements for privacy and autonomy (for example in the EU) then those high standards for privacy and autonomy will effectively benefit the users of the relevant software and hardware worldwide.

3.3 Real rights – and underlying issues

Ultimately, however, a rights-based approach is appropriate because the kind of rights that are set out in this thesis really do constitute 'rights'. They are rights that people consider to be appropriate in the world, as it has developed, if public actions over the kinds of events discussed above are considered. The facts that more than 30,000 German citizens signed up for a

[22] See the detailed analysis in Chapter Three

legal action against the implementation of the Data Retention Directive,[23] and that the public uproar was critical in the collapses of Phorm's Webwise and Facebook's Beacon are just some of the examples of how the public feel about this.

Each of the rights set out in this thesis emerges from real wishes and needs. If the internet is a 'place' that is in effect an extension of the 'real' world, then the idea that people should have a right to roam that place with a reasonable expectation of privacy is sensible. Where surveillance occurs, it is reasonable that this surveillance is held in balance, and is accountable – so the right to monitor the monitors matches with real expectations. In the UK, for example, where CCTV operates there is an expectation that not only are there warnings about that CCTV, signs indicating where it is in place, and legal controls about how and when it is used[24] – the right to monitor the monitors takes the idea a step further, but a step that fits with the different (and in some ways even more intrusive and pervasive) nature of internet surveillance. Finally, the idea that people should have control over the data that others hold on them is inherent in the data protection regime – the right to delete that data is in a sense a logical extension of that concept, and as argued in Chapter Five, a necessary extension if control over personal data is to be properly effective.

Meeting the real needs and desires of individuals is the point – as are helping to establish the nature of those real needs and desires and finding ways to communicate them. Throughout the case studies, most notably Google in Chapter Three and Phorm in Chapter Four, the battle over 'hearts and minds' has been crucial. Here too, following on from Gearty's suggestion that it is "the intellectuals, the workers and the streets"[25] that matter, rights play a critical part. For the public to become engaged and to have their voices heard,

[23] See Chapter 3 Section 2.3.2
[24] See the ICO CCTV code of practice, online at http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/ico_cctvfinal_2301.pdf
[25] Quoted for his talk 'The Rights' Future', see http://www2.lse.ac.uk/publicEvents/events/2010/20101006t1830vSZT.aspx

they need a language to use, and they need the support of people who can help them to articulate their concerns, their needs and their desires. In this context that can be particularly important – when concepts like privacy are difficult to pin down, to define and to explain, and where technical and technological language is complex and opaque, it is hard for ordinary people to explain why they are concerned about something. Furthermore, as shall be set out at the end of this thesis, a basis of rights can underpin a symbiotic regulation approach to bringing about change – one that in this situation is the most likely to succeed.

The three rights should be seen as a whole, and as underpinning a more basic right – the right to use and enjoy the internet with freedom. That, in turn, as discussed from the start of this thesis, arises from the idea that rights should be in place to enable people to fully enjoy and fully participate in society – to flourish. If, therefore, people in our modern society have the right to use and enjoy the internet with the same kinds of freedom and the same kinds of rights as they have to enjoy the 'real' world, then these are exactly the kinds of rights that are required.

That, then, leads back to the paradigm shift referred to at the start of this chapter. It is this paradigm shift that is the basis of all the rights suggested. Freedom should be the assumption, the default, and surveillance and control the exception. The question should not be when should individuals be able to roam freely, unfettered and unmonitored, but when and why should they not be able to roam free. Similarly the question should not be when should individuals have access to (and more importantly control over) data about them, but when should they not have such access and control. Businesses and governments should need to justify the gathering, utilising and holding of data, rather than having individuals obliged to make claims in the other direction.

4 Rights and symbiotic regulation

In Chapter Four, the Phorm saga was looked at through the lens of symbiotic regulation, and it was argued that the fall of Phorm was to a large extent due to its failure to understand the complexity or the nature of the regulatory matrix in which it operated. Further, it was argued that coherent and comprehensible rights could have helped Phorm and its backers to understand that regulatory matrix better, and to realise earlier in the affair the problems that were inherent in its business plan and its approach. Specifically, if they had understood why people would be concerned about the way they were being monitored and tracked by Phorm, and that the people who were concerned would have the ability to make that concern known – through working with advocacy groups, through their influence on other related commercial interests, through political lobbying, through legal challenges and so forth – then they would have been in a position to modify their business model to better meet the concerns of the community.

4.1 Extending the argument

These arguments extend to the other examples and other scenarios examined in this thesis, and indeed to almost any related situation. In all cases the regulatory matrices are complex and multifaceted. In all cases there are many related and interested parties that have influence and effect. There are commercial organisations that both compete and cooperate with each other in different contexts – Google, for example, cooperate closely with Apple in the integration of Google applications and services onto Apple's iPhone and iPad, but compete with it in that Google's Android operating system is used on smartphones competing directly with Apple's iPhone. Similarly, government and other regulatory bodies both cooperate and compete for jurisdiction and influence – the relationships between the UK government and various European bodies show this on a regular basis, with the EC's current legal action against the UK about the implementation of its various directives in relation to Phorm a case in point.

Individuals have relationships with all those organisations and bodies, in various different ways – as customers of the commercial organisations, as voters for the various governments, as members of pressure groups or advocacy organisations and so forth. The analysis of the Phorm farrago shows this in detail, but the same is true for Google in relation to its struggle with the Working Party over data retention periods – and even about the UK government in terms of its adoption or rejection of an ID card database. None of these situations are straightforward, and in all cases negotiating the complex regulatory matrix has proved far from straightforward.

That is another key reason to push for the use of a rights-based approach. Rights, if appropriately expressed, can help to clarify the complex regulatory matrix and help those who wish to negotiate it to a particular end to find a better way to do so. Those who opposed Phorm's Webwise were better able than Phorm's supporters in doing this, as the eventual result of the saga demonstrated, but after a long and painful struggle, one that could have been shorter and less painful had the relevant rights been clearer and better understood. Similarly, the struggle between Google and the Article 29 Working party was long and drawn out – and to an extent continues to this day. The Article 29 Working Party have the backing (and are formed from) the various information commissioners and their equivalents throughout the EU, but from the outside their mandate, their backing and their effectiveness is not as clear as it might be if the rights which it is effectively attempting to enforce were better understood and accepted. Well-presented and supported rights could have assisted the Working Party and perhaps brought about a more positive resolution to their struggle faster and more clearly. Similarly, working within the UK, properly expressed rights could support the work of the ICO, which to a great extent at the moment works with one hand tied behind its back, having to function through a softly-softly approach of coaxing and suggestion, only occasionally resorting to big threats – as in the

ACS:Law example[26] – and even then such threats are rarely realised, as businesses know very well.

4.2 Empowering individuals

Most importantly, well-expressed rights could support and empower the individuals themselves – and help them to find ways to get their concerns across. It could help them to know which commercial services to choose, for example, if they are able to understand which are doing better in terms of its supporting and satisfying their rights. In the Phorm story, Tim Berners-Lee said that he would change his ISP if it introduced a system like Webwise:[27] if more people had known and appreciated their rights they could have been in a position to make similar decisions. What is more, the commercial organisations themselves would know that and would be more likely to set their policies accordingly. This would be symbiotic regulation in practice – rights having a ripple effect through the regulatory matrix, ultimately influencing the businesses that create the environment itself. If advocates wish for a privacy-friendly, or even an 'autonomy-friendly' internet, then rights can be the first step in bringing it about.

One further benefit of a rights-based, symbiotic regulatory approach in this field is that it has a chance of overcoming, at least to an extent, the differences of both political culture and legal systems between Europe and the United States. By working through rights to influence business models and to stimulate 'customer' attitudes and activity, it might be able to manoeuvre its way through the market-based systems that are both prevalent and preferred in the United States, whilst by providing back up and support for legislators like the Article 29 Working Party it can help bolster the positive aspects of the more interventionist and legalistic systems of the European Union. That, together with the cross-jurisdictional issues discussed earlier in the chapter – the way that if a global internet entity

[26] See Chapter 5, footnote 69
[27] See Chapter 4, Section 4.1

changes its policies in one jurisdiction the benefits of those changes can apply world wide – gives this kind of approach a chance of having more of a global impact.

4.3 A self-balancing system?

The struggle through which privacy is gaining prominence and perhaps more acceptance may well be one that is happening anyway, without the need for the assertion of rights. In some of the case studies – most notably that of Phorm (and of Beacon) and to an extent that of the Google data retention periods – the outcome has at least in some senses been a 'privacy friendly' one. Over the last few years the whole subject of privacy has gained prominence and more attention is paid to it by the internet industry than had been in the past.

That does not mean that there is no need for the assertion or establishment of rights. The purpose of the rights, and indeed of their assertion, is not to start or impose a solution, but to support and help shape a process that has some momentum of its own – for if the rights are real rights, they would not need to be imposed, but would instead fit with something that people already want or need. Moreover, the trend in this field is hardly one in a clearly – let alone purely – positive direction for those in favour of privacy and autonomy. Data retention is still very much in place. Behavioural targeting, despite the setbacks of Phorm and the growing willingness to legislate on both sides of the Atlantic,[28] is still flavour of the day in the internet advertising industry, used by most of the big players including Google, Microsoft and Yahoo!. News stories of data losses and debatable data privacy practices are arriving on a near daily basis and in a seemingly never-ending stream.

The trend in terms of privacy is far from clear – which makes the benefit of rights even more apparent. What is more, the rights proposed should also empower those who are losing (or have already lost) their autonomy – for

[28] See Chapter 4, Section 5

there have been many victims in this process, and there will almost certainly be many more. Indeed, part of the overall effect of the changes in the internet has been to further widen one aspect of the digital divide, the 'savvy' and the 'non-savvy' – 'savvy' people have a better idea of what the problems are, and how to avoid them. That is one of the most important aspects of the idea of promoting the kinds of rights suggested in this thesis – to try to address the issue of this digital divide. It is the 'non-savvy' whose autonomy is most under threat, and who are at the most risk.

It is the contention of the thesis that through these rights a better, more sustainable, more autonomous, more egalitarian and less divided and divisive internet could develop in the future – one in which the digital divide between those in the know and the rest has less impact. That future, how it might look, and how the changes in it might address some of the problems inherent in the current situation, is one of the principle subjects of Chapter Seven.

Chapter 7: A privacy-friendly future?

1 A threat to autonomy?

The research question for this thesis is: "Do deficiencies in data privacy threaten our autonomy and if so, can informational privacy rights meet this threat?" To see whether that question has been answered through the research set out in the thesis, the question needs first of all to be broken down into its three components:

1) Are there deficiencies in data privacy?
2) Do these deficiencies threaten our autonomy? and
3) Can these threats be addressed by informational privacy rights?

1.1 Are there deficiencies in data privacy?

Establishing that there are deficiencies in data privacy has been a central part of this research – and most of the case studies set out in Chapters Three to Five concern one kind of deficiency in data privacy or another. Some of these are systematic and what might be described as 'intentional' problems – problems where the privacy issue is intentionally avoided, subverted or confused. These issues include: the lack of transparency in the way that search engines function, set out in Chapter Three; the way that the issue of consent is generally dealt with at best superficially, set out in Chapter Four; and the basic nature of monitoring systems such as the behavioural targeting systems, also described in Chapter Four. These systematic and intentional deficiencies in data privacy represent at least to an extent the essential reality of the internet as it currently exists: the basic paradigm for much of the internet is that surveillance, monitoring and data gathering is the norm, privacy the exception. It is that paradigm that this thesis suggests needs changing.

These systematic and intentional deficiencies in data privacy are built upon and compounded by the various forms of data vulnerability set out in Chapter Five. These vulnerabilities in themselves represent deficiencies in data privacy, but the combination of the intentional and systematic deficiencies and the various vulnerabilities is perhaps even more significant – data gathered systematically then becomes vulnerable, and vulnerable in ways not initially envisaged by those gathering the data. Those gathering the data and invading people's privacy are often not doing so for any kind of malignant intent, or even particularly manipulative purpose. Even if they work without real consent and without the intention of gaining consent, their motivations are frequently purely pecuniary or even based on the desire to develop interesting ideas or projects. Many things are just developed because 'they're cool' – the people creating and developing ideas may do so out of sheer enjoyment in the creative process. The implications of their ideas – and the ways in which their ideas may be used by others – can be more significant than they envisage. Once data gathered for benign purposes gets into the hands of others with less positive motivations, whether they be criminal, governmental or simply less than scrupulous businesses, that data may find very different uses than the neutral or benign ones initially envisaged. The Wikileaks phenomenon, which has exploded onto the public consciousness in the last year, is just one example of how the data genie once released is very hard to get back into the bottle.

From the perspective of law, the data privacy deficiencies described throughout the thesis show many different characteristics. Some appear to be currently legal in most jurisdictions, while some are clearly illegal – including many of the vulnerabilities outlined in Chapter Five. Many, however, are seemingly somewhere between the two – Phorm is perhaps the most direct example of this, though the ongoing discussions between Google and the European authorities suggest that the legality of the more mainstream activities of the largest operators on the internet may not be as

clear-cut as they seem.[1] Moreover, as the law is currently complex and to a great extent opaque, there is scope for doubt not only as to what is acceptable but also as to what is actually legal – and again the Phorm saga provides a good example of how this kind of problem may resolve itself. There are more problems too, as revealed by the case studies: issues of jurisdiction, of enforcement, conflicts between different areas of law and conflicts of interest within governments. Governments, charged with the enforcement of privacy law, also have an interest in encouraging the gathering of personal data: the Google case study demonstrated this tension through the conflict between data retention and data protection law. The Phorm case study showed the conflict of interest for governments between supporting businesses and supporting privacy. All these issues and conflicts can and do result in further deficiencies in privacy.

In conclusion, therefore, the research in this thesis suggests that deficiencies in data privacy do exist, both at a systematic level and otherwise, and that those deficiencies are not being currently addressed sufficiently by the law – or by those currently operating the primary systems of the internet.

1.2 Do these deficiencies threaten our autonomy?

As discussed in Chapter One, Section 2 autonomy in this context is considered in a broad sense – Raz's concept of an autonomous person being the (part) author of their own life. It includes the freedom to be irrational and the idea of autonomy in a social context. This broad definition of autonomy is used for a number of reasons. First of all it serves to reflect the real issues in relation to informational privacy on the internet, arising as they do to an extent from the advertising industry which uses emotional and subliminal rather than purely rational methods. Secondly it addresses the increasingly social nature of the internet – social networking services like Facebook and Twitter are just one aspect of this social nature. Thirdly, it is useful because,

[1] Phorm is discussed in depth in Chapter Four, the Google vs EC dispute in Chapter Three

as noted briefly in Chapter One and more extensively below, a broad definition of autonomy addresses some of the key criticisms of a more traditional privacy/autonomy approach – particularly the feminist and communitarian critiques, but also certain key aspects of the crucial transparency critiques and challenges, addressed in section 4 below.

The key case studies of this thesis – in particular Google in Chapter Three and Phorm in Chapter Four – addressed this kind of autonomy directly and demonstrate that threats to autonomy do indeed arise through the deficiencies in privacy revealed. Some of these threats to autonomy are inherent in the current commercial set-up of the internet – the conceptual model, the Symbiotic Web, described in Chapter Two, sets out the nature and impact of these inherent threats to autonomy, arising amongst other things from the commercial pressures building as a result of the symbiotic dependence of enterprises on the gathering and utilisation of personal data.

1.2.1 Specific threats from case studies

The case studies have also revealed specific threats to autonomy arising in particular situations. Search engines, the centrepieces of Chapter Three, are the key current tools for those using the internet to find what they want and to do what they want – so the ways in which search results are determined are fundamental to internet autonomy. That search engines do not currently operate with much transparency, that they have the potential to be 'tailored' without the knowledge or understanding of the user, that they could conceivably be set out to favour particular sites, services or providers all restrict the autonomy of the user.

Behavioural advertising – and in particular Phorm's 'Webwise', the central case study of Chapter Four – uses private and personal data and surveillance in order to impact upon individuals' autonomy. What is more, though it was initially intended to work in an advertising context, Phorm explicitly intended to extend to the tailoring of content as well as advertising – which

could potentially have a significant impact upon autonomy, particularly when done without transparency or consent. Furthermore, as for search engines, this tailoring could (and is perhaps likely to be) set out in favour of the commercial interests of those doing the tailoring rather than in the interests of the user – and whilst there is often a considerable overlap between the two, as most people may well prefer relevant advertising and content, that overlap is *not* complete.

1.2.2 Threats from profiling and related processes

In both the search engine and the behavioural advertising context, a key element of the processes used is profiling or its equivalents, and that in itself has a very direct impact upon autonomy. What is more, profiling means even apparently inconsequential data can become highly sensitive – and the wide variety of vulnerabilities shown in Chapter Five demonstrates how those data in particular can be vulnerable. There is attention to security – and an addressing of security by better management, technological security methods etc – but that attention is focussed almost entirely on the most directly and obviously sensitive data.

Autonomy is threatened both by 'true' data and by false, by both accurate suppositions and predictions and by inaccurate suppositions and predictions, and by both accurate and inaccurate profiling. A targeted advertisement, for example, based upon an assumption that an individual is gay could cause different kinds of damage to an individual if it was true or if it was false – but the damage could be there either way. Autonomy and privacy includes the right to keep certain things secret and private, even if they are true – whilst having decisions made about oneself based on false assumptions derived from infringements of privacy is self-evidently damaging and potentially unfair.

1.2.3 Profiling, the internet and politics

The question of whether any of this matters if it just concerns advertising has been addressed in Chapter Four, but there is a further angle to consider. Imagine for example tailored advertisements created for individual 'swing voters' (selected automatically through profiling), pointing out a party's positive steps in the policy areas that are most likely to interest them (also selected automatically), omitting those areas where party policy doesn't fit, and couching it in a language appropriate to the individual's ethnic, educational, cultural and linguistic background, illustrated with a few appropriate news TV clips, and playing background music exactly to the individual's taste and voiced over by an actor that profiling reveals that individual likes? The reverse, of course, about the political party's opponents – negative campaigning and personal attacks taken to an extreme level. This could be extended from tailored advertisements to whole 'news' pages where the 'news' provider has a particular political agenda, and also (and more simply) to individual automated emails.

The nature of the internet and the data gathering and utilisation processes described throughout this thesis suggest that a system that is already contentious could become worse and more persuasive than ever before. Much of this activity will by its very nature remain hidden. The extent of monitoring and the technologies used and developed for this purpose are naturally difficult to assess – and it is important neither to overestimate what is happening nor to ignore either the possibilities or realities of what is going on. With every new example it becomes clearer that those with a desire to profile and control, and indeed to take action, are aware of the potential. Recent developments – for example the 2011 protest-led regime-changes in Tunisia and Egypt – have shown that communications technology in general, and the internet in particular, are having a growing political impact.

In summary, through the case studies and analysis – and in particular the symbiotic nature of the current model of the internet, as set out in the vision

of the Symbiotic Web described in Chapter Two – it appears clear that autonomy is threatened through the kinds of deficiencies in privacy examined here. There are potential reductions in choices, selections of choices and inappropriate prioritisation of choices. What is more, autonomy is affected both by actual threats and by perceived threats – choices made depend on both – and one trend that also appears clear is that people are becoming more aware of the potential threats to their privacy and hence their autonomy.

1.3 Can these threats be addressed by informational privacy rights?

This issue was looked at in principle in Chapter Six – not only what rights are being suggested, but how they interact and how they could work in practice. That they have already been seen to be effective is borne out by the realisation that rights played a key part in the outcomes to the two principle case studies in Chapters Three and Four – outcomes that were at least partially positive from the perspective of privacy. In the case of Google, it was the Article 29 Working Party's insistence that Google's data retention period breached the Working Party's understanding of the users' rights that helped bring about the changes in Google's policies, whilst in the case of Phorm, it can be argued that the whole conflict was brought about by a realisation and assertion that what Phorm were doing was in breach of users' rights to privacy.

This thesis does not suggest that without clearly expressed, coherent, understood and acknowledged rights that positive outcomes like these would not occur – rather, as has been argued in Chapter Six, it is the case that rights expressed and understood in this way would support the processes which bring about these outcomes. They could make those processes less painful for all concerned, and with less 'collateral damage', less loss of trust. They could help prevent future similar problems – to the benefit not only of the individuals but also to the businesses whose ideas would otherwise be likely

to fail, or at the very least be less successful and less profitable. In effect, they could support the beneficial symbiosis described in Chapter Two.

The rights would provide support and stimulus for the various mechanisms through which changes can and do actually take place. This is what is meant by the suggestion that the rights would work through symbiotic regulation. The rights could support the activities of bodies such as the Article 29 Working Party in their negotiations with Google – providing backup for their arguments and helping in their attempts to win the public battles for hearts and minds. The rights could help the advocacy groups and others make their points about business ideas like Phorm's Webwise – and they could help businesses themselves realise that ideas like Phorm are unlikely to succeed, either legally or as a business.

2 An internet with rights

The rights-based approach and the reasons for it have been discussed both above and in the previous chapters, most directly in Chapter Six, but there is another important angle to consider – what would the idea of a right-based approach mean for the internet as a whole? What is being suggested in this thesis is the idea that if human rights exist and are taken seriously in the real world, then they should be extended to activities in the online world. This has significance well beyond what has so far been considered.

2.1 Rights to support autonomy

The rights that have been discussed so far relate directly to personal autonomy, as these are the rights that have emerged through reflecting on the question upon which this research is based. They are, however, not the only rights in relation to autonomy that would arise once the general concept of an internet with rights is taken seriously.

2.2 Free expression – and the right to be found?

Perhaps the most important right to consider in the relation to the internet is a right to free expression – the internet is to a great extent a communications medium, and much of the current use of the internet relates to the expression of ideas, particularly in relation to what might loosely be described as the Web 2.0 applications: blogs, wikis, social networking and related services. Free expression can be considered another aspect of autonomy – indeed, the privacy-related threats to autonomy can have a significant impact on freedom of expression, not just directly (for example where a dissident blogger is tracked down and arrested as a result of breaches of privacy) but through the chilling effect – a kind of 'Internet Panopticon' effect – that the knowledge of the potential privacy-related risks can produce.

There are also aspects to the issue of free expression that go beyond issues of privacy. One particularly direct aspect relates to the functioning of search engines and other navigation methods through the internet. The idea of neutrality of search has been discussed from the perspective of the searcher in Chapter Three – but in terms of a right to free expression, the neutrality of search from the perspective of the site is what matters. Does the creator of a website have a 'right to be found'? To be more precise, a 'right to be found if you want to be found'. That kind of a right wouldn't mean that a site could demand special treatment from a search engine – but that the site should be able to be sure that it wouldn't receive especially unfair treatment, and that a search engine should treat it on its merits, according the principles that are known and understood. The implementation of a right like this would have particular difficulties in relation to the rights of the search engines themselves to trade secrets insofar as their search algorithms are concerned, but companies and the EC have already bitten the bullet sufficiently to take Google on in terms of possible biasing of search results in the 'Foundem' case[2] and this kind of right would relate directly to this kind of bias, as bias in

[2] The European Commission has opened an anti-trust investigation into Google which 'follows complaints by search service providers about unfavourable treatment of their

favour of something is by its very nature bias against something else. Google have responded to this accusation in relation to Foundem by saying (amongst other things) 'We built Google for users, not websites,'[3] but in an increasingly personal internet, where users are becoming publishers, is that a sufficiently strong argument? If free expression is to be taken seriously, it may not be.

A 'right to be found' would be intended to prevent both bias and censorship – of the kind exercised by authoritarian regimes, for example – and would also have direct implications on filtering or blocking mechanisms such as BT's Cleanfeed, requiring them to be properly transparent and accountable, something that currently is questionable.[4] A right to be found (if you want to be found) could change the way that this kind of mechanism is looked at. As for the other rights discussed in this thesis, the aim would be to change the paradigm, to make individual rights and freedoms the starting point from which things proceed. This subject is one of significance and debate – and warrants further research.

2.3 Rights to anonymity and identity?

It is important to be clear that the right to roam the internet with privacy introduced in Chapter Three does not amount to a right to anonymity. The concepts of privacy, anonymity and identity are closely connected, but those connections are complex and at times unclear. A right to anonymity would be a step further than a right to roam with privacy, but they are on the same continuum. This thesis does not propose a general right to anonymity – it is hard to reconcile universal and unequivocal anonymity with the kinds of

---

services in Google's unpaid and sponsored search results coupled with an alleged preferential placement of Google's own services' See EC press release at http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1624&format=HTML&aged=0&language=EN&guiLanguage=en

[3] Quoted for example in http://www.pcpro.co.uk/news/363244/google-faces-eu-probe-over-doped-search-results

[4] See for example MCINTYRE, T. J. & SCOTT, C. D. 2008. Internet Filtering: Rhetoric, Legitimacy, Accountability and Responsibility. *In:* BROWNSWORD, R. & YEUNG, K. (eds.) *Regulating Technologies.* Oxford: Hart Publishing. McIntyre and Scott argue that 'where it is not clear what is being blocked, why, or by whom, the operation of mechanisms of accountability – whether by way of judicial review, media scrutiny or otherwise – is greatly reduced'.

accountability needed in an internet where access is in general paid for, for example, and with even the most basic needs of security. Nonetheless, in certain circumstances there is a strong argument for a need for a right to anonymity. People as diverse as victims of domestic violence, whistleblowers, dissidents and others organising protest groups under oppressive regimes have a need for protection and anonymity – and if the valued freedoms of association and assembly (and other related freedoms) are extended to the online world, then anonymity can often be a key tool in their practical application.

Even in less extreme situations a right to anonymity may be argued for. In the UK, the resistance to the introduction of identity cards is at least to some degree based on the idea that a person's identity is there own business, and that in normal circumstances it should not be required to be proved. This follows a historical belief in this kind of privacy – Privacy International, in part of their campaign against the ID card, refer to the 1952 case of *Willcock v Muckle*, which signalled the end of the use of wartime ID cards in the UK.[5] In his judgment of that case, Lord Goddard summed up the attitude to the need to prove your identity:

> "…to demand production of the card from all and sundry, for instance, from a woman who has left her car outside a shop longer than she should, or on some trivial occasion of that sort, is wholly unreasonable."[6]

On this basis the default position should be one where people have a right to anonymity. Indeed, one of the hallmarks of a 'police state' in the minds of many Britons is that you have to have your papers with you at all times and that the police, without any reason, can demand that you prove your identity. This situation, and this standpoint, is not replicated in all countries, and for some the idea of carrying identity cards isn't seen as an imposition but just as

[5] See https://www.privacyinternational.org/article/id-card-frequently-asked-questions
[6] Willcock v. Muckle, [1951] 2 K.B. 844

something normal and acceptable. According to Privacy International, in 1996 around 100 countries had official, compulsory ID card systems used for a variety of purposes, including countries such as Germany, France, Belgium, Greece, Luxembourg, Portugal and Spain. Countries without cards include the United States, Canada, New Zealand, Australia, Ireland, and the Nordic countries as well as the UK.[7] Attitudes to anonymity are clearly contentious – and if a right to anonymity is developed and supported, it would need to be a right that came with well-developed and strong checks and balances.

The concept of a right to identity, though in some ways the converse to a right to anonymity can be seen more directly, as shall be shown below, as a complement to it. From a practical perspective, there are times and places both offline and online when the assertion (and certification) of identity can be and is required, and others where it can be of significant help – where finance is concerned, or when dealing with e-Government. What is more, the right to identity has already begun to take legal forms. The United Nations Convention of the Rights of the Child includes a right to identity[8] while jurisprudence from the European Court of Human Rights appears to have derived such a right through their interpretation of Article 8 of the ECHR.[9]

If this interpretation is followed to its logical conclusion in the online world, then some kind of right to an online identity is likely to arise. Looking at it from another angle, if people have a right to internet access, as discussed in Chapter One, then it can be argued that they must logically have a right to an

[7] See https://www.privacyinternational.org/article/id-card-frequently-asked-questions It should be noted that some of these countries (e.g. Sweden) have what almost amounts to a de facto compulsory ID card system: an ID card that is required to be produced when a Swedish national pays for something using a credit card.

[8] Article 8.1 of the UNCRC states "States Parties undertake to respect the right of the child to preserve his or her identity, including nationality, name and family relations as recognized by law without unlawful interference." See http://www2.ohchr.org/english/law/crc.htm

[9] In *Tysiac v Poland* (Application no. 5410/03 Judgment 20 March 2007) for example, the European Court of Human Rights 'reiterates that 'private life' is a broad term, encompassing, inter alia, aspects of an individual's physical and social identity including the right to personal autonomy, personal development and to establish and develop relationships with other human beings and the outside world'. This case and others, and the related rights, are analysed in MARSHALL, J. 2009. *Personal freedom through human rights law? : autonomy, identity and integrity under the European Convention on Human Rights,* Leiden ; Boston, Martinus Nijhoff Publishers.

online identity. This takes the logic used earlier in this thesis a step further. If online life is now an intrinsic component of modern life, then to function fully in modern life – to flourish in modern society – then human rights must extend to our online life, and to flourish online an individual needs to be able to assert an online identity, because as shall be discussed below, a recognised online identity is already becoming important for much of what happens online, and is likely to become more significant in the future.

From a more philosophical perspective, identity can be viewed as a narrative. As Andrade puts it,

> "Identity is not looked at as a sum of different elements, representative of one's identity and subject of being misrepresented and falsified, but as a narrative, an individual inner story that each person needs to build, develop and rewrite over time in order to define the meaning of their lives.'[10]

This kind of approach is particularly relevant to online identities. The process of creating, establishing, asserting and protecting an online identity is something that happens in a similar way in many online communities. In these communities, which come in a wide variety of forms from message boards to virtual worlds like Second Life or games like World of Warcraft, identity is taken very seriously, and built up over time through a mix of creative and interactive processes. It can be argued that the personalities and identities created and used online have little connection with those of the people creating them, but the conception of identity put forward by Andrade and others suggests that the development of 'real' personalities and identities follows very similar patterns. The online identity could therefore be viewed more as an 'extension' of the offline identity and personality of the individual concerned.

[10] DE ANDRADE, N, N Gomes, *"Human Genetic Manipulation and the Right to Identity: The Contradictions of Human Rights Law in Regulating the Human Genome"*, (2010) 7:3 SCRIPTed 429, http://www.law.ed.ac.uk/ahrc/script-ed/vol7-3/andrade.asp, p432

Moreover, in these kinds of communities there is a clear and recognised interest in individuals protecting their online identities, and there are regular and specific threats and indeed attacks on those identities. Where communities are moderated or regulated there are generally rules against impersonating someone else's identity, whether for malicious purposes or otherwise – because such practices are relatively common.[11] Such issues have parallels in more 'conventional' online situations – the 'Blaney's Blarney' case, most often noted as the case in which an injunction was permitted to be served via Twitter, revolved around this kind of deception, where the defendant served was effectively impersonating blogger Donal Blaney's online identity on Twitter.[12] In the Blaney's Blarney case the link between the online and real world identities was direct, intentional and explicit – but it may not be too great a leap from this kind of a case to purely online impersonation.[13]

There are important complications when considering online identities. What happens if a real person creates more than one online identity, a far from rare occurrence in many online communities? What about identities created or maintained by more than one 'real' person? Can an online identity be passed from one real person to another? Even more to the point, what about 'fake' identities created for negative or damaging purposes – or even created and maintained automatically, using what have become known as 'persona management software' designed specifically to create and support false identities to be used for various reasons including the generating of

[11] These kinds of strictures exist in most varieties of online community, from the simple message board to full scale virtual worlds. The 'house rules' of the BBC's '606' message board, for example, say that they "…reserve the right to fail contributions which… … [a]re seen to impersonate someone else." See http://www.bbc.co.uk/dna/606/houserules (accessed February 16th 2011). At the other end of the scale, the terms and conditions of Second Life include requirement not to "(ii) Impersonate any person or entity without their consent, or otherwise misrepresent your affiliation…" see http://secondlife.com/corporate/tos.php#tos8

[12] See http://www.griffinlaw.co.uk/2009/10/01/griffin-law-makes-law-by-serving-via-twitter/

[13] Another relevant case is Applause Store Productions Limited, Matthew Firsht v Grant Raphael [2008] EWHC 1781 (QB), relating to a form of impersonation on Facebook, resulting in defamation.

apparently popular support for products or opinions?[14] These kinds of issues and the questions that surround them are complex and demanding, and the answers to them may be problematic and changeable. As technologies and habits develop, they are likely to become even more challenging.

Despite these complications and difficulties, it appears that a prima facie case for some kind of right to online identity does seem to exist. The basic right could be set out as a right to create, assert and protect an online identity. Following and supporting this, but much more qualified, there could be a right to privacy over the link between that online identity and the creator's 'real' identity. Similarly, there could be controls over not only impersonation of another's real or online identity, but over the deliberate creation of false online identities for malicious or misleading purposes. This area is one that needs more work and thought than is appropriate for this thesis, but it is an area that suggests scope for significant future research.

2.4 Privacy, identity and anonymity

With a general assumption of privacy in place, the ideas of when and where identities need to be verified can become clearer and more appropriate. Having an internet where the paradigm is one of privacy – and indeed sometimes anonymity – does not preclude the development and use of more sophisticated and reliable systems for the verification of identity in appropriate circumstances. Indeed, it may even hasten the development of such systems.

This new privacy-friendly internet can be seen as a sea of privacy with islands where identity is required and others where 'true' anonymity is possible. The important thing is that the default position is privacy, and that

[14] For a description of some of these kinds of practices, known as 'astroturfing', http://www.guardian.co.uk/environment/georgemonbiot/2011/feb/23/need-to-protect-internet-from-astroturfing. The name astroturfing is derived from the idea of creating fake 'grassroots' movements, something apparently practiced by lobby groups from the tobacco and fossil fuels industries. See for example http://www.guardian.co.uk/environment/2006/sep/19/ethicalliving.g2?INTCMP=SRCH

there must be good reasons to require identity or to allow full anonymity. In this 'sea of privacy', where needed and appropriate, identity could be requested and asserted. In online shopping, for example, browsing could be private, but once the decision is made to make a purchase, the online identity could be asserted. If a user wants to use their identity to browse, to receive tailored content, tailored search results, links and so forth, he or she could chose to do so.

This idea of a default position of privacy also provides part of an answer to the question of how 'public' the internet is. The 'sea of privacy' can be viewed as the 'public' internet – an internet in which people are free to roam, but with an underlying level of responsibility. Rules about what is and is not acceptable in this public space would be relatively clear – general and universal monitoring would not be expected, but the possibility of individual and targeted monitoring or localised surveillance would be known and understood. The islands of anonymity or identity could be seen as the 'private spaces' on the internet, where different rules apply – in either direction.

As with many ideas in this area there are risks – risks of function creep, risks that islands demanding identification could grow in inappropriate or disproportionate circumstances and merge into great continents that could fill the 'sea of privacy'. Coherent and understood rights can help, underlining the shift of paradigm. By establishing that the default position should be – and is – privacy, the onus is put upon those who want to breach that privacy to establish systems that are appropriate, proportionate and consensual – and follow the kinds of principles set out in Chapter Five looking at exceptions to the right to delete personal data.

2.5 A declaration of rights?

As discussed in Chapter Six, the kind of rights envisaged in this thesis are not intended as simple legal rights, implemented through legal means, but

something more fundamental, having effect through a process of symbiotic regulation. Communication and dissemination of the rights is a key part of this kind of process – and looking at it from both a historical and a practical perspective, a 'declaration of rights' would be a particularly appropriate method of such communication.

As Gearty puts it in his web project 'The Rights' Future' 'Proclaiming something is not enough to make it true, but it is a necessary preliminary in the struggle for its realisation.'[15] This could be particularly true in relation to online rights. There are many different groups (and kinds of groups) with both an interest and an influence over rights on the internet: states, corporations, programmers and hackers, community groups both online and offline, NGOs, pressure groups and lobbyists as well as all the individuals who inhabit or could in future inhabit the online world. The huge differences between these groups in form, practices and histories makes communication and mutual understanding both difficult and complex – something like a declaration of rights can help bridge these gaps and underpin more specific and precise agreements, rules and laws that are needed to enforce and protect these rights.

Such a declaration would not follow the lines of John Perry Barlow's famous 'Declaration of Independence of Cyberspace' in 1996,[16] or its 2010 YouTube successor by the now notorious hacker group Anonymous.[17] The suggestion here is to develop such a declaration by working with governments rather than confronting them as the Barlow and Anonymous initiatives did – and it would not attempt to deny their jurisdiction or ability to regulate, but instead aim to guide, influence and shape the regulation that they can and do provide. The declaration would recognise that the essence of a successful internet is one of mutual support and cooperation. It would in effect be an

[15] In his online web project, 'The Rights' Future', Track 3, at http://therightsfuture.com/t3-making-truth/
[16] https://projects.eff.org/~barlow/Declaration-Final.html
[17] On YouTube as a video at http://www.youtube.com/watch?v=gbqC8BnvVHQ

acknowledgment of the symbiotic nature of the web, with the benefits and dependencies that result from it.

As awareness of the issues has spread, the idea of a 'Bill of Rights for the internet' has begun to be talked and written about recently. Andrew Murray suggested something along these lines in his blog in October 2010, [18] analysing and developing two working suggestions, one based in Brazil by the country's Ministry of Justice, in partnership with the Centre for Technology and Society from Fundação Getúlio Vargas,[19] the other under the auspices of the Internet Rights and Principles Coalition.[20] Also in October 2010, Conservative MP, Robert Halfon made the suggestion of an Internet Bill of Rights to the Backbench Business Committee of the House of Commons.[21]

"*It should be up to the internet companies to respect the rights of the individual, not the other way around. I am calling for an internet bill of rights, a proper inquiry and an Information Commissioner who genuinely acts to safeguard our liberties. I hope that hon. Members and the Government will be able to support that.*"

Even more recently, the Obama administration in the US has begun to suggest something similar, confirming that it would support any congress proposal for a 'privacy bill of rights',[22] though what is being suggested or

[18] See http://theitlawyer.blogspot.com/2010/10/bill-of-rights-for-internet.html
[19] See http://portal.fgv.br/ Not that this page is in Portuguese. Murray's blog referred to above describes and analyses the key elements of this page in English.
[20] See http://internetrightsandprinciples.org/ and for the proposed bill of rights itself, http://internetrightsandprinciples.org/node/367 . The Internet Rights and Principles Coalition brings together people from academia, civil society, governmental institutions and the private sector that has 'set out to make Rights on the Internet and their related duties, specified from the point of view of individual users, a central theme of the Internet Governance debate held in the IGF context.'
[21] See http://www.theyworkforyou.com/whall/?id=2010-10-28a.143.0
[22] White House spokesman, Commerce Department Assistant Secretary Lawrence Strickling, announced this in testimony before the Senate Committee on Commerce, Science & Transportation. See http://commerce.senate.gov/public/index.cfm?p=Hearings&ContentRecord_id=e018f33b-d047-4fba-b727-5513c66a6887&ContentType_id=14f995b9-dfa5-407a-9d35-56cc7152a7ed&Group_id=b06c39af-e033-4cba-9221-de668ca1978a For further discussion see for example in the Wall Street Journal http://blogs.wsj.com/law/2011/03/16/obama-to-push-privacy-bill-of-rights/

envisaged is a very limited and focuses entirely on the commercial use of data (and in particular behavioural advertising), rather than on any broader issues. The research in this thesis highlights the limitations of such a limited approach – as the case studies throughout the thesis, from Google in Chapter Three and Phorm in Chapter Four to many of the examples discussed in Chapter Five have demonstrated, the actions, tactics and approaches of governments and businesses are inextricably intertwined and effectively inseparable. A declaration or bill of rights or similar would need to cover both – and, indeed, that should be one of the primary strengths of a rights-based approach, for by focussing on the rights of the individual, the actions of all those who infringe those rights, be they governmental agents or commercial actors, can be scrutinised, assessed and where appropriate opposed both politically and legally.

The Global Network Initiative (GNI) is another initiative working towards somewhat similar goals. The GNI is a non-profit organisation bringing together ICT companies with civil society groups, academics and investors with the principle aim of helping companies find ways to 'respect and protect the freedom of expression and privacy rights of users in responding to government demands, laws and regulations.'[23] Though the GNI has strengths, it also has weaknesses and has not yet been able to produce much in the way of results. Important players in the field signed up to it from the start: some of the biggest ICT companies (specifically Google, Microsoft and Yahoo!), the best of civil society (including the Center for Democracy & Technology, the Electronic Frontier Foundation and Human Rights Watch) and some highly respected academic organisations (including the Berkman Center for Internet and Society at Harvard University). This breadth of membership may be the key to its problems – it appears to have moved slowly since its foundation in 2008. It has gained no additional corporate technology members since that foundation, and has not actually performed any of the reviews that were stated as one of its principle activities.[24] It has

[23] See http://www.globalnetworkinitiative.org/cms/uploads/1/GNI_WhoWhatWhere.pdf
[24] See for example http://www.nytimes.com/2011/03/07/technology/07rights.html?_r=1

also faced criticism – from groups like Amnesty International, who withdrew cooperation prior to the initial launch, citing weakness in responding to human rights concerns,[25] and from bloggers like Larry Downes in *Forbes*.[26]

An analysis of the problems faced by the GNI is beyond the scope of this thesis, but though it shows some promise, it does not appear that it has the kind of strength to influence either corporations or governments sufficiently to make a real difference. It may be that in time it will be in a position to do so and that the work that it has done will go on to lay the groundwork for real progress – but so long as it is driven to a great degree by the industry, rather than informed by a properly supported, popular vision of rights, it is hard to see that its impact will be significant.

That, in many ways, is the key. It is the groundswell of community understanding that must underpin change: the rights suggested in this thesis are intended to represent and support *real* rights as needed and desired by people. Those must come first: declarations of rights are statements of what is already there in hearts and minds, not something imposed upon people from above.

2.6 The role of law

Rights understood and 'declared' can be seen as effectively arising from the community – but to make rights meaningful more is needed than community agitation and international declarations. As can be seen from the case studies throughout this thesis, the law itself, on many different levels, has a significant part to play.

In the Google vs. EU case study in Chapter Three, the lobbying of the Article 29 Working Party needed the backing of European law. The combination of

[25] See http://www.guardian.co.uk/technology/2008/oct/30/amnesty-global-network-initiative
[26] See http://blogs.forbes.com/larrydownes/2011/03/30/why-no-one-will-join-the-global-network-initiative/

the existence of the Data Protection Directive and the history of European action and substantial fines to the likes of Microsoft supported the Working Party's intervention. In the Phorm case study in Chapter Four, the law was invoked both at a national level in relation to the Data Protection Act, RIPA and the Fraud Act and by bringing in the European angle in terms of the UK's implementation of the European Directives on Data Protection and ePrivacy. Again, the law was needed to back up the actions and pressure brought to bear by the privacy advocacy groups and others in conflict with Phorm. In a slightly different way, the two recent SABAM cases[27] have emphasised the role that European Law can play in supporting individual privacy rights against the competing right of intellectual property – law can and does help keep the balance between the various competing rights.

That is the next key stage of the process: putting the laws into place that support and enforce these rights. The existence and coherent expression of rights can help in the processes through which the laws emerge. Declarations of rights can lead to international conventions, which can lead to regional and then local laws – but laws that are more coherent and more in harmony with one another as they have a common basis. As Reed's concept of the 'cyberspace fallacy' made clear, human actors in cyberspace are accountable to local laws as they have a physical existence in a particular locale.[28] However, as he has subsequently argued, excessive use of local laws in cyberspace can lead to many problems, from conflict of law to the law itself having less effect – and thus being less respected.[29] Having more harmonious laws governing activities in cyberspace could ameliorate those problems. Coherent rights could help speed and simplify the kind of legal convergence

[27] Scarlet v SABAM in Novermber 2011 and SABAM v Netlog in February 2012, discussed in Chapter 6, section 1.4

[28] See REED, C. 2004. *Internet law : text and materials,* Cambridge, Cambridge University Press., Chapter 7.

[29] See REED, C. 2010. Think Global, Act Local: Extraterritoriality in Cyberspace. *Working Paper Series, Queen Mary University of London School of Law.* Reed argues that as states attempt to apply their own national laws to 'foreign' cyberspace actors, they can 'reduce the normative force of law as a whole and create the risk that otherwise respectable cyberspace actors become deliberate lawbreakers.'

that Reed argues can be beneficial – that is part of the strength of the rights-based approach.

Through a rights-based approach existing local laws can be brought more into line with common standards, and new laws brought in where required. With convergent laws, there would be less need for supra-national courts or dispute resolution systems: rather, competent local courts in relevant jurisdictions should be able to deal with the key issues. There will continue to be issues and complications that need to be dealt with: to an extent it is likely to be through the resolution of these issues and complications that appropriate systems will arise. Through these processes appropriate laws will be developed, and greater levels of expertise and understanding of the complexities of both the technologies and how they are used will be built up. Expertise can be developed and communicated, and with the backing of coherent rights can give assistance to both legislatures and courts in navigating their way through the new and unfamiliar territory.

Law is not the starting point for the process of making rights real – compared to the development of privacy- and autonomy-friendly business models it plays a relatively small part, and in addition is likely to be difficult and lengthy. It is, however, a key element of the regulatory matrix and a vital part of the process.

3 The internet of the future – and addressing critiques

The key test of the effectiveness of a rights-based approach to this issue is how it would pan out in reality. Rights in theory are only really worthwhile if they produce positive results. A sketch of how the internet might look if these rights were introduced – and how that form of internet might meet the criticisms that the ideas of privacy and autonomy face – was begun in the previous section, looking at the interaction of privacy, anonymity and identity. Filling in this sketch now can serve to demonstrate how the rights

suggested in this thesis might address the critiques to the suggested approach discussed in Chapter One.

3.1 Communitarian Critiques

From a communitarian perspective, as discussed in Chapter One, it can be argued that an emphasis on the importance of privacy and individual autonomy tends to prioritize the individual over the community and/or misunderstands the essentially social nature of humanity. As also discussed in Chapter One, the broader definition used for the purposes of this thesis begins the process of addressing this, but the way that an internet with these rights in place would support the building and functioning of safe and effective communities plays a much more important part in meeting the critique.

The rights suggested in this thesis would specifically support those human rights that are far from individualist in character: freedom of association and freedom of assembly. The right to roam the internet with privacy combined with the right to monitor the monitors are particularly crucial in this, allowing online assembly and association with far less likelihood of that assembly and association being monitored, disrupted or prevented, as those who oppose that assembly and association would have their abilities to monitor and control curtailed. Similarly, the use of communications technologies to organise and support 'real world' assembly and association, the importance of which was graphically demonstrated in the uprisings in Egypt and Tunisia in January and February 2011,[30] would be harder to disrupt.

[30] Whilst headlines like 'Tweeting Tyrants out of Tunisia' in Wired (http://www.wired.com/threatlevel/2011/01/tunisia/) were somewhat hyperbolic, it does appear that the internet in general and social media in particular played a role in the organisation of the uprisings. At the very least, those whose power was under threat seemed to believe so, doing their best to either hack into those systems or, in Egypt's case to attempt to shut off the internet in total in their country in order to deny their opponents the opportunity to use those systems.

The model suggested above, combining a default of privacy with enclaves of anonymity and identity, would directly support these kinds of rights, and indeed the communities that the rights themselves support. Privacy for most activities, anonymity in times and places of trouble, and the technology and ability to verify identity when needed – for example to root out spies or infiltrators working for those wishing to disrupt the communities – is the kind of combination that is needed.

The risks of communities fracturing through 'back-door Balkanisation', as discussed in Chapter Two, would be reduced by this kind of a rights model – though the risks of the more deliberate form of fracturing, as set out in Sunstein's Republic 2.0,[31] remains. Indeed, if there is excessive use of anonymity, it might be increased – a private, unmonitored internet could allow further polarisation without the means to govern or control. Privacy, rather than anonymity, is the key here – privacy that can be overturned where absolutely necessary. Perhaps more importantly, one of the purposes of the rights suggested here is that they are intended to help build trust – a trust that the current deficiencies in privacy is putting at risk. Trust is crucial for community building – trust of business, trust of government, trust of other individuals. The rights suggested in this thesis are intended to help to rebuild that trust.

3.2 Feminist Critiques

From a feminist perspective, the first and perhaps most important thing to note is that the internet has huge potential for women. It has the capacity to reduce both deliberate and 'automatic' discrimination – removing both the face-to-face problems and those that are triggered by indicators such as names. It can be a key enabler for women in many ways – supporting working from home and other forms of flexible working. It can allow access to information and services, to democracy, communication and more to those

[31] SUNSTEIN, C. R. 2007. *Republic.com 2.0,* Princeton, Princeton University Press.

previously isolated in homes. That potential needs to be acknowledged and supported – the internet can be a key tool for liberation and empowerment.

The privacy/identity/anonymity model suggested above could provide even more potential for women – it would allow people to disclose their sex only when they want to, and might even allow some kind of voluntary 'proof' of sex in certain circumstances. This is delicate and contentious ground, but something that might be effective if the rights and controls are in place. It takes some of the developments in terms of monitoring and profiling that might have been used in a negative way and turns them into something positive – from the instant tracking/tailoring of Phorm to the detailed profiling systems being developed throughout the advertising industry. As noted above when looking at freedoms of association and assembly, spotting 'imposters' has some valid applications – in particular areas and as the exception rather than the rule. The existence of the rule – in this case the default of privacy – can help enable the exception.

There are also specific risks associated with the Symbiotic Web that have a particular relevance to feminist critiques – the potential for automated discrimination, including price, service and access discrimination, is something that should be prevented. The 'whites-only websites' noted as a nightmare vision in Chapter Two could just as easily be 'men-only' or 'women-only' websites – and whilst there are sometimes positive justifications for such kinds of exclusivity (at least on a gender basis) to be able to do so surreptitiously and automatically is something that needs to be done with care and control. Again, the privacy/identity/anonymity model could be the key here – if a website is intended to filter entry in ways other than by membership, it would have to demonstrate the need for that kind of filtration.

Having said all of this, there are still strong arguments against the excessive use of privacy and anonymity from a feminist perspective. Privacy has historically often been used to protect the powerful; from perpetrators of

domestic violence to Catholic priests accused of child abuse, and that kind of privacy should not be prioritised or protected. As Allen puts it, it should be possible to:

> 'Rip down the doors of "private" citizens in "private" homes and "private" institutions as needed to protect the vital interests of vulnerable people'[32]

This is one of the reasons why this thesis focuses on privacy as a protector of a balanced, broadened form of autonomy, rather than privacy per se, and also why this thesis does not suggest an unlimited right to anonymity. As noted in Section 3.1, the 'sea of privacy' is one in which people have general privacy, but that privacy can be 'overturned' when needed. Indeed, that is one of the principles behind the paradigm shift that underlies the rights set out in this thesis. The shift is a shift in defaults, but not an absolute shift. It is a shift from the general to the specific, from the all-encompassing to the targeted.

Taking this further, general surveillance and monitoring is unlikely in practice to provide the kind of protections 'in the private sphere' that the feminist critique would require. Those engaged in counter-terrorism have huge resources and often think on huge scales – hence the approach of data retention. They have access to databases, the cooperation (at least to a degree) of similar organisations around the world. From the perspective of those wishing to, in Allen's terms, rip down the doors of private citizens in private homes, focussed, intelligent, targeted approaches are more appropriate and more likely to be effective. Once again, having a default of privacy could encourage the development of the technologies and systems to provide those focussed, targeted systems.

[32] From ALLEN, A. L. 2003. *Why Privacy Isn't Everything: Feminist Reflections on Personal Accountability,* Lanham, Maryland, Rowman & Littlefield., p195-196

3.3 The Security Challenge

The security challenge, as set out in Chapter One, is one of the hardest challenges to engage with in a detailed and specific fashion. The very nature of the challenge makes it hard to examine – those engaged in counter-terrorism are loath to reveal their tactics and techniques, for understandable reasons. Nonetheless, the universalist approach that has led to concepts such as data retention, as discussed in depth in Chapter Three, has been coming under increasing pressure in recent years. As noted in that chapter, in December 2010 Peter Hustinx, the European Data Protection Supervision, called for a European Commission review of the Data Retention Directive to prove that it had achieved results or repeal it.[33] Further, Ian Brown argued in 2010 that the Directive is not proportionate to the harm it seeks to remedy.[34] Both Hustinx and Brown stress the serious level of intrusiveness of data retention – such intrusion needs a great deal to justify it. Is the threat to security posed by terrorism sufficient, and even if it is, does data retention really work?

On the other hand, Walker, in 2009, wrote that 'Communications data retention and interception have become a non-negotiable fact of modern life.'[35] Indeed calls in the US for the institution of mandatory data retention have been growing. In January 2011 the US Department of Justice put their position to the house crime subcommittee: 'Data retention is fundamental to the department's work in investigating and prosecuting almost every type of crime.'[36] The issue remains highly contentious and, as suggested in Chapter Three, it is likely to remain so in the short term at least.

---

[33] Hustinx's speech can be found at http://www.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/EDPS/Publications/Speeches/2010/10-12-03_Data_retention_speech_PH_EN.pdf

[34] In BROWN, I. 2010. Communications Data Retention in an Evolving Internet. *International Journal of Law and Information Technology,* 19.

[35] WALKER, C. 2009. Data retention in the UK: Pragmatic and proportionate, or a step too far? *Computer Law & Security Review*, 325-334., p333

[36] See for example http://news.cnet.com/8301-31921_3-20029423-281.html#ixzz1C4sUeyy4%3Cbr%20/%3E

There are, however, pragmatic arguments against current practices in relation to security – to start with because what can be used for 'good' security can be used for 'bad' security. Technologies and practices developed by governments or other authorities that might be viewed as 'benevolent' or 'in favour' of human rights (which in itself is a complex and contentious question) can equally be employed by the malevolent or oppressive. Susan Landau, an expert in encryption and surveillance, has suggested that embedding eavesdropping mechanisms into communication technology itself builds tools that can be turned against those who wish to be protected.[37] A version of this may have already happened in the case of the Chinese hack of Google discussed in Chapter Five – where it has been suggested that the hackers took advantage of a 'backdoor' created on the effective instructions of the American intelligence services. This case has not been proven – but even the existence of the rumour suggests that this kind of thing is being considered. It also highlights a fundamental problem with this kind of approach – as Schneier puts it:

> "It's bad civic hygiene to build technologies that could someday be used to facilitate a police state. No matter what the eavesdroppers say, these systems cost too much and put us all at greater risk."[38]

In a similar way, when considering particularly the nature of data retention, data vulnerability in itself undermines the security challenge – and as the numerous cases discussed in Chapter Five demonstrated, even those who might be expected to keep their data most securely, from the MOD and HMRC to the banks in Switzerland and Lichtenstein have found their data vulnerable in one way or another. The operations of Wikileaks have been shown to take these vulnerabilities to another level again and highlighted the fundamental problem with the gathering and accumulation of data. As Roger Smith, director of Justice, put it in the Law Society Gazette, writing about

[37] This is one of the principle messages of her most recent book, LANDAU, S. 2011. *Surveillance or Security? The Real Risks Posed by New Wiretapping Technologies*, The MIT Press.
[38] In his blog at http://www.schneier.com/blog/archives/2010/09/wiretapping_the.html

'Cablegate': '[c]reate a database of fascinating information that is accessible to over three million people and the only issue is how long you wait for the first leak'.[39] The argument can be widened to cover the whole of the internet. Governments have been made acutely aware of their own need for privacy – in the end, they may perhaps realise that a more privacy-friendly internet could be in their interest too. Murray has asked whether states should have a right to privacy[40] – this is something that would become much more possible in an internet where privacy is the default rather than the exception.

The most important argument against the security challenge, however, is one that has run throughout this thesis. It is an argument of principle, and one that echoes Gearty's call in The Rights' Future: 'In taming counter-terrorism law, human rights can forge a soul.'[41] The underlying question here is what kind of a vision for the internet is to prevail – and beyond that, what kind of vision for society as a whole is to prevail, for as discussed at the start of this thesis the internet is a reflection and an extension of the 'real' world, intrinsically intertwined with it. As Benjamin Franklin wrote:

"They who can give up essential liberty to obtain a little temporary safety, deserve neither liberty nor safety."[42]

This thesis has attempted to show how essential the liberties that are at stake here can be: as Hustinx, Brown[43] and others have suggested, that the safety being obtained is anything but little and temporary has yet to be shown.

---

[39] In *WikiLeaks take us into a legal – and moral – maze*, Law Society Gazette 16 December 2010: http://www.lawgazette.co.uk/opinion/rights-and-wrongs/wikileaks-take-us-a-legal-and-moral-maze

[40] In MURRAY, A. 2004. Should States have a Right to Informational Privacy. *In:* MURRAY, A. & KLANG, M. (eds.) *Human Rights in the Digital Age.* London: The Glasshouse Press.

[41] See http://therightsfuture.com/t14-triumph-through-adversity/

[42] This much quoted (and varied) phrase can be found in *FRANKLIN, B. & FRANKLIN, W. T. 1818. Memoirs of the life and writings of Benjamin Franklin, London, Henry Colburn.* p270.

[43] See BROWN, I. 2010. Communications Data Retention in an Evolving Internet. *International Journal of Law and Information Technology,* 19.

4 A transparent society?

In many ways what was labelled in Chapter One as the 'transparency critique' is the most fundamental critique not only of this thesis but of the whole idea of privacy not only on the internet but in society as a whole as it is currently developing. It represents a challenge of critical importance, a challenge that must be met if privacy is to be taken at all seriously.

As noted in Chapter One, there are three principle variants to the transparency critique:

1) That the struggle for privacy is already lost – as epitomised by McNeally's suggestion that 'You have zero privacy anyway, get over it'[44]
2) That the struggle for privacy is outdated – as implied for example by Facebook founder Mark Zuckerberg[45]
3) That the struggle for privacy is 'wrong'. The virtues of a 'transparent society' have been written about by Brin[46] and more recently by Bell & Gemmell[47] amongst others.

The case studies and analysis throughout this thesis provide strong responses to all three of these variants. The principle case studies in Chapters Three to Five all argue against them. The strength of the public responses to Phorm and Beacon, and the massive outcry and concern over the various data leaks from the HMRC disk loss onwards indicate that the argument that people don't care about privacy is far from proven. This suggestion has been emphasised by the continual struggles that Facebook has had in terms of its privacy policies – there have been at least three episodes in this saga over the last few years. Each time Facebook has introduced something reductive of

[44] Quoted for example in Wired, at http://www.wired.com/politics/law/news/1999/01/17538
[45] As noted in Chapter One, Zuckerberg made related statements to various elements of the media during 2010. See for example Chris Matyszczyk's blog on CNET, at http://news.cnet.com/8301-17852_3-10431741-71.html.
[46] In BRIN, D. 1998. *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?,* Reading, Mass., Addison-Wesley.
[47] In BELL, C. G. & GEMMELL, J. 2009. *Total recall : how the E-memory revolution will change everything,* New York, N.Y., Dutton.

privacy then after an outcry it has had to change them – the most recent reincarnation of Facebook's privacy policy, introduced in December 2010, has been in direct response to public concern.[48]

There are more arguments against the idea that privacy is an outdated concept. First and foremost, there is little other than anecdotal evidence in its support – more research is needed into the subject, as there is little empirical evidence either way. Research such as the Turow report on behavioural advertising[49] discussed in Chapter Four has been limited in scope – the evidence that it provides suggests that people do care about privacy, but it is far from sufficient for strong conclusions to be drawn. Secondly, the focus of Zuckerberg and others on young people's attitudes is limited in many ways: not only young people matter, and young people grow up, while lack of privacy, particularly on the internet, can last forever.[50] Solove's strong arguments about common misconceptions of privacy and its importance undermine the idea further.[51] The relationship between privacy and autonomy discussed throughout this thesis adds to these arguments: and even if the suggestion that people don't care about privacy were true, it would be hard to argue that people don't care about autonomy either.

As we have already briefly discussed,[52] in *Delete*,[53] Mayer-Schönberger has argued strongly against many aspects of the transparent society, suggesting not only that it is unrealisable but also that in many ways it goes against human nature – effectively subverting the crucial human ability to forget. As Mayer-Schönberger suggests, it would be evolutionarily very difficult, even if

[48] There have been outcries against Facebook's privacy policies and practices at various times since Facebook was founded. The Electronic Privacy Information Centre, for example keeps a documented list of issues at http://epic.org/privacy/facebook/, while the Electronic Frontier Foundation keeps a record of the developments on Facebook's privacy policies at http://www.eff.org/deeplinks/2010/04/facebook-timeline/

[49] TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania,

[50] Mayer-Schönberger's ideas on data having limited lifespans looks at exactly this: see MAYER-SCHÖNBERGER, V. 2009. *Delete : the virtue of forgetting in the digital age,* Princeton, N.J., Princeton University Press.

[51] Most directly in SOLOVE, D. J. 2007. "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review,* 44.

[52] In Chapter Five, Section 4

[53] MAYER-SCHÖNBERGER, V. 2009. *Delete : the virtue of forgetting in the digital age,* Princeton, N.J., Princeton University Press.

it were desirable, for humans to make the shift to living without forgetting. Some of his arguments go directly towards autonomy. As he puts it '[w]ithout our ability to forget, whenever faced with a decision we would always recall all our past decisions, resulting in potential indecision'.[54] The arguments set out in *Delete* are complex and compelling – and multifaceted.

The suggestion that the struggle for privacy is already lost is also challenged by the case studies. In many of the cases, it can be argued that privacy 'won', from Google's decision to reduce in its data retention periods to the failure of Phorm and Facebook's abandonment of Beacon. The toughening up of the Information Commissioner's Office, whether it ends up in being effective or not, suggests that the UK government has not given up the idea of privacy. Neither have the judiciary – a notable example being the Ryan Giggs case where judges would not accept that either Twitter or Parliamentary privilege should be allowed to override privacy.[55] Other examples such as the fact that in Germany, when given the option, almost 250,000 citizens opted to have the pictures of their houses 'blurred out' of Google Street View,[56] and the strength of the current movement in the US for 'Do Not Track' options in internet browsing (as discussed in Chapter Four) suggest once more that people do care about privacy and that the delivery of options like these can achieve results.

Even when and where transparency is desired or desirable, should that transparency not be two-way? At the moment the transparency is all from the individuals: those gathering data are conspicuously opaque both in their practices and policies – as Mayer-Schönberger suggests, this may be an inevitable result of the relative imbalance in power between individuals and those gathering their data.[57] This suggestion is borne out in the case studies throughout the thesis. The nature of Google search, as discussed in Chapter Three, makes their data gathering activities far from clear, while the whole

[54] Ibid. p117
[55] The case is *CTB v News Group Newspapers* [2011] EWHC 1232 (QB)
[56] See http://www.bbc.co.uk/news/technology-11595495
[57] See MAYER-SCHÖNBERGER, V. 2009. *Delete : the virtue of forgetting in the digital age*, Princeton, N.J., Princeton University Press., particularly pp107-108

approach of Phorm as detailed in Chapter Four was one where transparency was conspicuous by its absence. Facebook's privacy policies have been vehemently criticised for their complexity and opaqueness – their most recent change in December 2010 was intended to address this issue,[58] though its success in achieving a great degree of clarity is still under question,[59] and a new proposal was on the table only a few months later.[60] Taking this a step further, the strong reactions by governments and others to the WikiLeaks saga suggests that the idea of a fully transparent society is something that governments and businesses are far from enthusiastic about – and, as Murray has argued,[61] they may have good reasons for this attitude. Transparency needs limits.

The need for limited, two-way transparency rather than an open 'free-for-all' can be seen as another example of the web symbiosis, and of how it can function at its best. Individuals, businesses and governments all have needs for privacy and duties for transparency – and can get benefits from both. What is more, both privacy and transparency are in themselves two-way issues: privacy is privacy *from*, while transparency is transparency *to*. The rights presented here should be a positive step towards reaching the appropriate levels of transparency – two-way transparency, limited in both directions at appropriate levels. Ultimately they could be key tools in developing what is more like a positive version of transparency, as they could encourage more trust, and from trust you can get transparency.

---

[58] See for example the MSNBC blog on the subject 'Facebook Dumbs Down Privacy Policy (In A Good Way) http://technolog.msnbc.msn.com/_news/2011/02/25/6133609-facebook-dumbs-down-privacy-policy-in-a-good-way

[59] See for example ZDNet blogger Adrian Kingsley-Hughes, commenting in 2011 that 'Facebook Privacy Settings are Garbage' http://www.zdnet.com/blog/hardware/facebook-privacy-settings-are-garbage/11029

[60] In February 2011, Facebook put forward a proposal for 'A Privacy Policy Re-imagined For Users Like You', opening a consultation process with users in order to make their privacy policy more accessible and comprehensible. See: http://www.facebook.com/note.php?note_id=10150434660350301&id=69178204322and www.facebook.com/note.php?note_id=10150434652940301

[61] In his blog at http://theitlawyer.blogspot.com/2011/02/freedom-of-information-in-wikileaks-era.html . Murray has taken this argument a stage further, noting that not only governments and businesses but Wikileaks themselves understand that transparency needs limits – for Wikileaks are far from transparent about their own operations. See his article MURRAY, A. 2011. Transparency, Scrutiny and Responsiveness: Fashioning a Private Space. *Political Quarterly,* 83.

5 A privacy-friendly future?

When this research began in 2007 the internet was a very different place. Much of what is now considered central to it was either very new or had not even been launched. Twitter was barely a year old,[62] the iPhone had been launched a matter of months before in the US and had not been launched in the rest of the world, the Kindle had not yet been released at all.[63] The iPad was not even available until 2010.

Facebook, a relative veteran at 3 years old, had what looked an impressive 50 million users – but by August 2011 it had 750 million.[64] By the end of 2010, Amazon was reporting that Kindle e-books had outsold both hardbacks and paperbacks.[65] In the quarter ending 25th December 2010 alone, Apple sold more than 7 million iPads and 16 million iPhones[66] - the ways that the internet is being accessed, and the way that it is being used, have changed dramatically. The Apple 'App Store', through which applications for iPhones and iPads are sold, recently passed 10 billion app sales, and claims to have more than 160 million customers[67] – and the global market for smartphone apps is predicted to hit $17.5 billion by 2012.[68]

The growth in quantity of the personal data that has been gathered and generated as part and parcel of these changes has been equally profound – and the changes in the nature of the data perhaps even more so, as new data types have emerged or grown in significance. One prominent new form is social data – who someone's 'friends' are and how they interact with each other – from social networking sites. Habitual data – which books people read or movies they watch, what sports they follow and so forth – has

[62] Twitter was launched in July 2006
[63] The Kindle was released in November 2007
[64] See http://www.facebook.com/press/info.php?statistics
[65] See http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=1521090&highlight=
[66] See http://www.apple.com/pr/library/2011/01/18results.html
[67] See http://www.apple.com/pr/library/2011/01/22appstore.html
[68] See for example http://news.bbc.co.uk/1/hi/technology/8571210.stm

become far more significant as more people spend more of their time on the internet. Even more recently, information with deep significance for privacy such as the geolocation data generated by smartphones has emerged. Combining all of these different forms of data – and adding to it – is the profiling data that can ultimately be the most potent form of data of all in terms of both commercial opportunities and impact on autonomy.

These changes have been accompanied, as has been seen throughout this thesis, by problems and issues relating to how these data are gathered, processed, used and held. What was already an important issue when the research for this thesis began has become significantly more important. The scale and nature of these events and the media coverage and public reactions to them suggests three things: that significant events in the field are occurring with increasing regularity, that public interest in those events is increasing, and that the idea that people have rights concerning them is increasing in prevalence.

This is what makes the subject matter of this thesis important. The shape and form of the internet both mirrors and impacts upon the shape and form of society as a whole – and not only in rich Western countries but throughout the world. The role that the internet has played in the uprisings in the Arab world in 2011 has been much debated – but the fact that it did play a role, and that the challenged leaders of the countries concerned believed that this role mattered is clear. The internet is no longer a luxury, it is no longer either just about information or just about leisure – it is a key part of the lives of a significant and growing proportion of the people of the world.

Given that this is the case, it is crucial that the nature of the internet is given sufficient thought. If human rights and freedoms – and autonomy – are to be taken seriously in the world, they need also to be taken seriously in the online world. For that to happen, a lot of changes will need to take place. That, ultimately, is the aim of the paradigm shift and of the rights suggested in this thesis. If the human rights that are considered important are to be

protected – more, to be fostered and supported – then privacy and autonomy needs to become the default. Surveillance and breaches in privacy need to be the exception, and exist only when truly justified.

Is this kind of real change possible? The case studies suggest that it might be – there have been some very positive and quite radical changes in the policies and practices of some of the biggest players in the internet world. Where technology is concerned, changes are possible faster than in many fields, and in more radical ways. Whether or not it will happen in practice is another question. There are signs both ways, and pressures both ways.

Governments, and the US government in particular, seem conflicted or confused – or at very least capable of demonstrating distinct double standards or exceptionalism. On the same day in February 2011, for example, Hillary Clinton made a key speech on internet freedom[69] and the US Justice Department tried to subpoena the personal records on Twitter of some of those associated with Twitter[70] after having failed to prevent the disclosure of this fact via a gagging order. Just two days later the FBI requested more backdoors into social networking services.[71] In Europe, there are similar tensions between the needs of privacy and security, as played out in the conflicts between advocates of data protection and data retention detailed in Chapter Three. Governments, however, as noted throughout this thesis, are not the key to the development of the internet: to a great extent they are peripheral, piggybacking on the developments made by and the practices of businesses. A privacy-friendly internet, if it is to come to pass, is far more likely to be driven by the business sector than by government.

That is where hope appears to lie – and where the rights suggested in this thesis could play their most important role. What is more, there are signs that businesses are starting to understand the importance of privacy – and in

[69] On February 15th 2011, see http://www.bbc.co.uk/news/world-us-canada-12475829
[70] Also on February 15th 2011, see http://blogs.abcnews.com/politicalpunch/2011/02/doj-seeks-twitter-records-in-wikileaks-probe.html
[71] See for example http://www.wired.com/epicenter/2011/02/fbi-backdoors/

particular how much people are beginning to show that they care about it. Some of the biggest players on the net are starting, at least on the surface, to embrace the idea of privacy. Microsoft, Mozilla and Google are all engaged in the 'do not track' initiative for their respective browsers. Facebook has recently opened up their privacy policies for consultation, with an avowed aim of making privacy simpler and more user-friendly, and more in the hands of its users. Twitter demonstrated commendable courage in challenging the gag order placed upon it in the US government's attempt to subpoena personal data from particular individuals associated with WikiLeaks. Google has been taking steps in the direction of both privacy and autonomy: Alma Whitten, the company's Director of Privacy, Product and Engineering, wrote a blog in February 2011 entitled 'The freedom to be who you want to be…' [72] embracing the ideas of 'unidentified' and 'pseudonymous' uses of their services and introducing ways to tell when and how Google is monitoring your activities – at least beginning to take on some of the concepts introduced in this thesis from the right to roam with privacy to the right to monitor the monitors.

There are, however, distinct issues with these proposals – Google's idea of 'unidentified' is far from real anonymity, still retaining such information as IP addresses, and still posing risks to privacy, while Facebook's regular amendments to its privacy policies have often promised far more than they have delivered. Even so there are reasons for optimism. At the very least, Google, Facebook and Twitter have understood that there is public desire for privacy and autonomy.

The work of privacy advocates, of the Article 29 Working Party, and most importantly of the online community has played a key part in helping them to start along the path. If a privacy-friendly internet is to become reality then they need be guided, supported and assisted along the way. That, ultimately, is the part that rights such as those put forward in this thesis can play. They

[72] http://googlepublicpolicy.blogspot.com/2011/02/freedom-to-be-who-you-want-to-be.html

can support businesses so that businesses can help to provide people with the kind of internet that they want and need. The symbiotic relationship between businesses and individuals has been hugely productive and substantially beneficial over the last few years – with the rights in place it can continue to be so into the future.

BIBLIOGRAPHY

Academic sources, reports and related documents

AGRE, P. & ROTENBERG, M. 1997. *Technology and privacy : the new landscape,* Cambridge, Mass., MIT Press.

ALLEN, A. L. 2003. *Why Privacy Isn't Everything: Feminist Reflections on Personal Accountability,* Lanham, Maryland, Rowman & Littlefield.

ALLEN, A. L. & MACK, E. 1991. How Privacy Got Its Gender. *Northern Illinois Law Review,* 10**,** 441-471.

AYRES, I. 2007. *Super Crunchers: How Anything Can Be Predicted,* London, John Murray.

BEALES, H. 2009. The Value of Behavioral Targeting. Network Advertising Initiative.

BELL, C. G. & GEMMELL, J. 2009. *Total recall : how the E-memory revolution will change everything,* New York, N.Y., Dutton.

BERNERS-LEE, T. & FISCHETTI, M. 2000. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor,* New York, HarperCollins Publishers.

BOHM, N. 2008. The Phorm "Webwise" System - a Legal Analysis. Foundation for Information Policy Reseach.

BRIN, D. 1998. *The Transparent Society: Will Technology Force Us to Choose Between Privacy and Freedom?,* Reading, Mass., Addison-Wesley.

BROWN, I. 2010. Communications Data Retention in an Evolving Internet. *International Journal of Law and Information Technology,* 19.

BURTON, E. 2008. Report into the Loss of MOD Personal Data. MOD.

CATE, F. H. 1997. *Privacy in the Information Age,* Washington, D.C., Brookings Institution Press.

CLARK, A. & CHALMERS, D. J. 1998. The Extended Mind. *Analysis,* 58**,** 10-23.

CLAYTON, R. 2008. The Phorm "Webwise" System. http://www.cl.cam.ac.uk/~rnc1/080518-phorm.pdf: Online.

COCHRANE, A. 2007. Animal Rights and Animal Experiments: An Interest-Based Approach. *Res Publica,* 13**,** 26.

CRISP, R. 1987. Persuasive Advertising, Autonomy, and the Creation of Desire. *Journal of Business Eithics,* 6**,** 413-418.

DAHL, J. Y. & SÆTNAN, A. R. 2009. "It all happened so slowly" - On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice,* 37**,** 83-103.

DE ANDRADE, N, N Gomes, *"Human Genetic Manipulation and the Right to Identity: The Contradictions of Human Rights Law in Regulating the Human Genome"*, (2010) 7:3 SCRIPTed 429, http://www.law.ed.ac.uk/ahrc/script-ed/vol7-3/andrade.asp,

EDWARDS, L. & WAELDE, C. 2000. *Law and the Internet : a framework for electronic commerce,* Oxford, Hart.

FABRE, C. 2000. A Philosophical Argument for a Bill of Rights. *British Journal of Political Science,* 30**,** 77-98.

FRANKLIN, B. & FRANKLIN, W. T. 1818. *Memoirs of the life and writings of Benjamin Franklin,* London, Henry Colburn.

GANDY, O. H. 1993. *The panoptic sort : a political economy of personal information,* Boulder, Colo, Westview.

GEARTY, C. A. 2006. *Can Human Rights Survive?,* Cambridge, Cambridge University Press.

GEWIRTH, A. 1982. *Human Rights: Essays on Justification and Applications,* London, University of Chicago Press.

GREENLEAF, G. 2008. Function Creep - Defined and still dangerous in Australia's revised ID Card Bill. *Computer Law & Security Report,* 24**,** 56-65.

HART, H. L. A. 1955. Are There Any Natural Rights? *Philosophical Review,* 64**,** 175-191.

HUNTER, D. 2001. Philippic.com. *California Law Review,* 90**,** 70.

ICO 2008a. Privacy by Design Report.

ICO 2008b. 'Taking stock, taking action'. London: Information Commissioner's Office.

ICO 2010. Response to the Ministry of Justice's Call for Evidence on the current data protection legislative framework.

JACKSON, E. 2001. *Regulating Reproduction: Law, Technology and Autonomy,* Oxford, Hart.

KELLY, P. J. 1990. *Utilitarianism and Distributive Justice: Jeremy Bentham and the Civil Law,* Oxford, Clarendon.

KLUG, F. 2000. *Values for a Godless Age: The Story of the UK's New Bill of Rights,* London, Penguin.

LANDAU, S. 2011. *Surveillance or Security? The Real Risks Posed by New Wiretapping Technologies*, The MIT Press.

LESSIG, L. 2006. *Code: Version 2.0,* New York, Basic Books.

MACINTYRE, A. 1981. *After Virtue: A Study in Moral Theory,* London, Duckworth.

MARSHALL, J. 2009. *Personal freedom through human rights law? : autonomy, identity and integrity under the European Convention on Human Rights,* Leiden ; Boston, Martinus Nijhoff Publishers.

MAYER-SCHÖNBERGER, V. 2009. *Delete : the virtue of forgetting in the digital age,* Princeton, N.J., Princeton University Press.

MCINTYRE, T. J. & SCOTT, C. D. 2008. Internet Filtering: Rhetoric, Legitimacy, Accountability and Responsibility. *In:* BROWNSWORD, R. & YEUNG, K. (eds.) *Regulating Technologies.* Oxford: Hart Publishing.

MILL, J. S. & HIMMELFARB, G. 1982. *On Liberty,* Harmondsworth, Penguin.

MOSSBERGER, K., TOLBERT, C. J. & STANSBURY, M. 2003. *Virtual inequality : beyond the digital divide,* Washington, D.C., Georgetown University Press.

MURRAY, A. 2004. Should States have a Right to Informational Privacy. *In:* MURRAY, A. & KLANG, M. (eds.) *Human Rights in the Digital Age.* London: The Glasshouse Press.

MURRAY, A. 2011. Transparency, Scrutiny and Responsiveness: Fashioning a Private Space. *Political Quarterly,* 83.

MURRAY, A. D. 2006. *The Regulation of Cyberspace: Control in the Online Environment,* Milton Park, Abingdon, UK ; New York, NY, Routledge-Cavendish.

NARAYANAN, A. & SHMATIKOV, V. 2008. Robust De-anonymization of Large Sparse Datasets. *IEEE Symposium on Security and Privacy.* 2008 ed.

NISSENBAUM, H. F. 2010. *Privacy in context : technology, policy, and the integrity of social life,* Stanford, Calif., Stanford Law Books.

NORRIS, P. 2001. *Digital divide : civic engagement, information poverty, and the Internet worldwide,* Cambridge, Cambridge University Press.

OHM, P. 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review,* 57**,** 1701-1778.

PAINE, T. & BURKE, E. 1791. *Rights of Man: Being an Answer to Mr. Burke's Attack on the French Revolution,* Dublin, [s.n.].

PARISER, E. 2011. *The filter bubble : what the Internet is hiding from you,* London, Viking.

POSNER, R. A. 1998. *Economic analysis of law,* Boston, Aspen Law & Business.

POYNTER, K. 2008. Review of information security at HM Revenue and Customs. *In:* TREASURY, H. (ed.). London: HMSO.

RAWLS, J. 1999. *A Theory of Justice,* Cambridge, Mass., Belknap Press of Harvard University Press.

RAWLS, J. & FREEMAN, S. R. 1999. *Collected papers,* Cambridge, Mass., Harvard University Press.

RAZ, J. 1986. *The Morality of Freedom,* Oxford, Clarendon.

REED, C. 2004. *Internet law : text and materials,* Cambridge, Cambridge University Press.

REED, C. 2010. Think Global, Act Local: Extraterritoriality in Cyberspace. *Working Paper Series, Queen Mary University of London School of Law.*

RORTY, R. 1993. Human Rights, Rationality and Sentimentality. *In:* SHUTE, S. & HURLEY, S. L. (eds.) *On Human Rights: Oxford Amnesty Lectures.* Oxford: BasicBooks.

SAMUELSON, P. 2000. Privacy as Intellectual Property *Stanford Law Review,* 52**,** 1125-1175.

SOLOVE, D. J. 2007. "I've Got Nothing to Hide" and Other Misunderstandings of Privacy. *San Diego Law Review,* 44.

SUNSTEIN, C. R. 2007. *Republic.com 2.0,* Princeton, Princeton University Press.

SWEENEY, L. 1997. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics,* 25**,** 98-110.

TAYLOR, C. 1992. *The ethics of authenticity,* Cambridge, MA, Harvard University Press.

TEFF, H. 1994. *Reasonable care : legal perspectives on the doctor-patient relationship,* Oxford, New York, Clarendon Press ; Oxford University Press.

THOMAS, R. & WALPORT, M. 2008. Data Sharing Review Report. London: Ministry of Justice.

THOMAS, T. 2008. The Sex Offender 'Register': A Case Study in Function Creep. *The Howard Journal,* 47**,** 227-237.

TUROW, J., KING, J., HOOFNAGLE, C. J., BLEAKLEY, A. & HENNESSY, M. 2009. Americans Reject Tailored Advertising. Annenberg: University of Pennsylvania.

VAN DEN POEL, D. & BUCKINX, W. 2005. Predicting online-purchasing behaviour. *European Journal of Operational Research,* 166**,** 557-575.

WALKER, C. 2009. Data retention in the UK: Pragmatic and proportionate, or a step too far? *Computer Law & Security Review***,** 325-334.

OTHER WEB-BASED SOURCES AND REPORTS:

All Party Parliamentary Communications Group report into internet traffic, downloadable from http://www.apcomms.org.uk/uploads/apComms_Final_Report.pdf

Amnesty International: Undermining Freedom of Expression in China. The role of Yahoo!, Microsoft and Google. Downloadable from: http://www.amnesty.org/en/library/info/POL30/026/2006

'Anonymous' declaration of intent 2010. See http://www.youtube.com/watch?v=gbqC8BnvVHQ

Attorney General press release re identity fraud. http://www.attorneygeneral.gov.uk/nfa/whatarewesaying/newsrelease/pages/identity-fraud-costs-27billion.aspx

Barlow: 'Declaration of Independence for Cyberspace', found at https://projects.eff.org/~barlow/Declaration-Final.html

Chinese Government white paper 'The Internet in China' (available online at http://www.china.org.cn/government/whitepaper/node_7093508.htm)

'Data Protection in the European Union', http://ec.europa.eu/justice_home/fsj/privacy/guide/index_en.htm

Data Protection Directive – first implementation report: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0265:FIN:EN:PDF

'Digital Britain', the interim report from January 2009: http://www.culture.gov.uk/images/publications/digital_britain_interimreportjan09.pdf

'Dutch Data Protection Agency' opinion on data retention: http://www.dutchdpa.nl/downloads_adv/z2006-01542.pdf?refer=true&theme=purple

E-money evaluation report: Available online at http://ec.europa.eu/internal_market/bank/docs/e-money/evaluation_en.pdf

Fleischer, Peter, letter to Peter Schlaar, 10 June 2007, downloadable from http://64.233.179.110/blog_resources/Google_response_Working_Party_06_2007.pdf

Gearty, Conor: The Rights' Future: http://therightsfuture.com

ICO CCTV code of practice, online at http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/ico_cctvfinal_2301.pdf

ICO code of practice for Privacy Notices: http://www.ico.gov.uk/for_organisations/topic_specific_guides/privacy_notices.aspx

ICO guidance on compliance with 'Cookies Directive': http://www.ico.gov.uk/for_organisations/privacy_and_electronic_communications/~/media/documents/library/Privacy_and_electronic/Practical_application/advice_on_the_new_cookies_regulations.ashx

ICO press release on their investigation into Street View: http://www.ico.gov.uk/~/media/documents/pressreleases/2010/google_inc_street_view_press_release_03112010.ashx

ICO 'specialist guide' as to what constitutes personal data: http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/personal_data_flowchart_v1_with_preface001.pdf

ICO 'specialist guide' to new penalties for data breeches 2010: available from http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/ico_guidance_monetary_penalties.pdf

IPCC report on HMRC data loss 2008: downloadable from http://www.ipcc.gov.uk/final_hmrc_report_25062008.pdf

Pew Internet Research report on 2008 election campaign: http://people-press.org/report/384/internets-broader-role-in-campaign-2008

Steiner, Peter, cartoon, 'On the internet nobody knows you're a dog', (1993). In 2000, the New York Times published a piece entitled 'Cartoon Catches the Spirit of the Internet'. http://www.nytimes.com/2000/12/14/technology/14DOGG.html?pagewanted=1&ei=5070&en=f0518aafeccf36fd&ex=1183089600

Article 29 Data Protection Working Party Opinions:

Working Party Opinion 4/2005 (WP 113) on the Proposal for a Directive of the European Parliament and of the Council on the Retention of Data Processed in Connection with the Provision of Public Electronic Communication Services and Amending Directive: Available from: http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/2005_en.htm

Working Party Opinion 3/2006 (WP 119) on the Directive 2006/24/EC of the European Parliament and of the Council on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC: Available from: http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/2006_en.htm

Working Party Opinion 1/2008 (WP 148) on data protection issues related to search engines. Available from: http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/2008_en.htm

Working Party Opinion 2/2010 (WP 171) on online behavioural advertising http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf

A note on web sources

Unless otherwise noted, all web links were last accessed 1st September 2011. There are extensive additional links to web sites are included within the footnotes – these sources are not individually listed here as they relate principally to news stories or companies or services available over the web rather than the subjects of academic analysis.

LAWS AND OTHER STATUTORY SOURCES:

UK Laws:

Data Protection Act 1998 (available online at http://www.opsi.gov.uk/Acts/Acts1998/ukpga_19980029_en_1)

Data Retention (EC Directive) Regulations 2009 (available online at http://www.legislation.gov.uk/uksi/2009/859/contents/made)

Interception of Communications Act 1985 (available online at http://origin-www.legislation.gov.uk/ukpga/1985/56/pdfs/ukpga_19850056_en.pdf

Privacy and Electronic Communications (EC Directive) (Amendment) Regulations 2011 (available online at http://www.legislation.gov.uk/uksi/2011/1208/contents/made)

Regulation of Investigatory Powers Act (RIPA) 2000 (available online at http://www.legislation.gov.uk/ukpga/2000/23/contents)

Unfair Contract Terms Act 1977 (available online at http://www.legislation.gov.uk/ukpga/1977/50)

International declarations and conventions:

Universal Declaration of Human Rights (UDHR) (available online at http://www.un.org/en/documents/udhr/)

United Nations Convention on the Rights of the Child (UNCRC) (available online at http://www2.ohchr.org/english/law/crc.htm)

International Covenant on Economic, Social and Cultural Rights, 1966, downloadable from http://www2.ohchr.org/english/law/cescr.htm

European Convention for the Protection of Human Rights and Fundamental Freedoms (CPHRFF), commonly known as the European Convention on Human Rights (ECHR) (available online at http://www.echr.coe.int/NR/rdonlyres/D5CC24A7-DC13-4318-B457-5C9014916D7A/0/ENG_CONV.pdf)

Convention on Cybercrime (downloadable from http://conventions.coe.int/Treaty/EN/Treaties/html/185.htm)

European conventions and directives:

Treat Establishing the European Community, downloadable from: http://eur-lex.europa.eu/en/treaties/dat/12002E/pdf/12002E_EN.pdf

Data Protection Directive (95/46/EC): available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

Data Retention Directive (2006/24/EC) available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32006L0024:EN:HTML

E-Commerce Directive (2000/31/EC), available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0031:EN:HTML

E-Money Directive (2000/46/EC), available online at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0046:EN:HTML

E-Money Directive (2009/110/EC), which came into force in April 2011 and is available online at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:267:0007:0017:EN:PDF

E-Privacy Directive (2002/58/EC) available online at http://eur-lex.europa.eu/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf replacing Directive 97/66/EC, http://eur-lex.europa.eu/pri/en/oj/dat/1998/l_024/l_02419980130en00010008.pdf

'Cookies Directive' labelled PE-CONS 3674/09, modifying Directive 2002/22/EC *on universal service and users' rights relating to electronic communications networks and services*, Directive 2002/58/EC (the ePrivacy Directive) and Regulation (EC) No 2006/2004 *on cooperation between national authorities responsible for the enforcement of consumer protection laws*. Available at: http://register.consilium.europa.eu/pdf/en/09/st03/st03674.en09.pdf

Other European Law:

German Federal Data Protection Act. Downloadable including English translation from http://www.bdd.de/Download/bdsg_eng.pdf

U.S. statutory sources:

'Building Effective Strategies To Promote Responsibility Accountability Choice Transparency Innovation Consumer Expectations and Safeguards Act' or 'BEST PRACTICES Act', available online at http://www.house.gov/apps/list/press/il01_rush/h_r_5777_the_best_practices_act_2010.pdf

Child Online Protection Act of 1998 (COPA)

Communications Decency Act of 1996 (CDA)

'Do-Not-Track Online Act of 2011', available online at http://commerce.senate.gov/public/?a=Files.Serve&File_id=85b45cce-63b3-4241-99f1-0bc57c5c1cff

Foreign Intelligence Surveillance Act of 1978 (FISA)

The Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT ACT) Act of 2001, downloadable from http://thomas.loc.gov/cgi-bin/bdquery/z?d107:hr03162:%5D

CASES:

UK Cases:

Applause Store Productions Limited, Matthew Firsht v Grant Raphael [2008] EWHC 1781 (QB)

Campbell v Mirror Group Newspapers Ltd [2004] UKHL 22

CTB v News Group Newspapers [2011] EWHC 1232 (QB) (the Ryan Giggs super-injunction case)

Durant v FSA [2003] EWCA Civ 1746

Liberty v UK [2008] 48 EHRR 1

Mosley v News Group Newspapers [2008] EWHC 1777 (QB)

R. (on the application of Gillan) v Commissioner of Police of the Metropolis

[2003] EWHC 2545 (Admin)

Willcock v. Muckle [1951] 2 K.B. 844

European Cases:

Ireland v European Parliament and Council of the European Union, Case C-301/06

European Parliament v. Council and Commission, Joined cases C-317/04 and C-318/04 (passenger numbers case)

Gillan v United Kingdom (4158/05) (2010) 50 E.H.R.R. 45; 28 B.H.R.C. 420

SABAM v Netlog NV [2012] CJEU C 360/10

Scarlet Extended v SABAM [2012] ECDR. 4

Tysiac v Poland (Application no. 5410/03 Judgment 20 March 2007)

Von Hannover v Germany [2004] ECHR 294

U.S. Cases:

ACLU v. Gonzales, No. 98-CV-5591, pending in the Eastern District of Pennsylvania.

Cairo, Inc. v. Crossmedia Services, Inc., 2005 WL 756610 (N.D. Cal. Apr. 1, 2005)

Gonzales v. Google, Inc., No. CV 06-8006MISC JW (Mar. 17, 2006)

Hotmail Corp. v. Van$ Money Pie Inc., 1998 WL 388389, *6 (N.D. Cal. 1998)

ProCD v. Zeidenberg, 86 F.3d 1447 (7th Cir. 1996)

Specht v. Netscape Communications Corp., 150 F. Supp. 2d 585 (S.D.N.Y.2001)

Ticketmaster L.L.C. v. RMG Tech., Inc., 507 F.Supp.2d 1096, 1102-1103 (C.D. Cal., 2007)