Dataveillance and Terrorism: Swamps, Haystacks and the Eye of Providence

Stuart Macdonald[*]

Introduction

In today's 'pre-crime'[1] society it is easy to understand the appeal of a technology that promises to identify terrorist plots and stop attacks before they happen – even when the would-be perpetrators have never previously attracted the attention of the authorities or aroused any suspicion (unknown unknowns, in Rumsfeldian terminology). Such claims are not merely the stuff of Hollywood movies and TV shows like *Person of Interest*. Mass dataveillance programmes – which have been the subject of much discussion and controversy since the Snowden revelations of 2013 – have been claimed to offer this degree of predictive potential.

Existing literature on dataveillance programmes and their regulation commonly urges the importance of striking a balance between the competing demands of protecting national security and respecting individuals' privacy, notwithstanding the fact that the flaws of this approach have been well-documented.[2] For a start, when the issue is approached in this way it tends to end up being viewed as *our* security versus *their* (suspected terrorists') liberty.[3] This is certainly true in discussions of dataveillance, where it is common to assume that the privacy interests of law-abiding citizens are not at stake – since the authorities will not be interested in their data – while suspected terrorists have no legitimate privacy interests.[4] When the liberty side of the equation is perceived in this way, and contrasted with bold predictions of the potential security gains dataveillance has to offer, the balance can only tip in one direction. Before jumping to questions of balance, however, it is essential to first of all thoroughly assess the security and privacy interests that are at stake.[5] That is the aim of this chapter. The chapter begins with an explanation of what dataveillance is and an evaluation of its effectiveness as a counterterrorism tool, focusing specifically on pattern-based queries. As we will see, not only are there reasons to doubt the efficacy of this predictive technique, it also risks significant opportunity and collateral security costs. The second half of the chapter examines different views of what privacy means in this context and of its importance, arguing that privacy is often understood too narrowly and given insufficient weight. Having examined the security and liberty interests independently, the chapter concludes by offering a more nuanced view of what it means to balance these interests in this context.

[*] College of Law, Swansea University.

[1] Lucia Zedner, 'Pre-Crime and Post-Criminology?' (2007) 11 *Theoretical Criminology* 261.

[2] Ronald Dworkin, 'The Threat to Patriotism' *New York Review of Books* (New York), 28 February 2002, 44; Ronald Dworkin, 'Terror & the Attack on Civil Liberties', *New York Review of Books* (New York), 6 November 2003, 37; Stuart Macdonald, 'Why We Should Abandon the Balance Metaphor: A New Approach to Counterterrorism Policy' (2008) 15 *ILSA Journal of International and Comparative Law* 95; Stuart Macdonald, 'The Unbalanced Imagery of Anti-Terrorism Policy' (2009) 18 *Cornell Journal of Law and Public Policy* 519; Christopher Michaelsen, 'Balancing Civil Liberties against National Security? A Critique of Counterterrorism Rhetoric' (2006) 29 *University of New South Wales Law Journal* 1; Jeremy Waldron, 'Security and Liberty: The Image of Balance' (2003) 11 *Journal of Political Philosophy* 191; Lucia Zedner, 'Securing Liberty in the Face of Terror: Reflections from Criminal Justice' (2005) 32 *Journal of Law and Society* 507.

[3] Waldron (n XX); Zedner, 'Securing Liberty in the Face of Terror' (n XX).

[4] Daniel J Solove, '"I've Got Nothing to Hide" and Other Misunderstandings of Privacy' (2007) 44 San Diego Law Review 745.

[5] Macdonald, 'Why We Should Abandon the Balance Metaphor' (n XX).

Security

Dataveillance and its use in counterterrorism

Record keeping is not new. In the sixth century BC Servius instituted a census in the Roman Empire.[6] The Domesday Book, which covered much of England and parts of Wales, was completed in the eleventh century on the orders of William the Conqueror.[7] Modern day censuses have been conducted in many countries – including the US, UK, Canada, France and Germany – since the nineteenth century.[8] But in recent years there has been a surge in the quantity of information that is collected and stored. Three developments in particular have contributed to this. First, increases in computer power. Over the past four decades Moore's Law – which predicts that computer chip capacities will double roughly every two years – has proven to be quite accurate.[9] Thanks to this exponential growth storage capacities are now measured in zettabytes and yottabytes, not bytes and megabytes.[10] Second, decreases in data storage costs. In 1984 it cost roughly two hundred US dollars to store a megabyte of data. By 1999 the cost had decreased to seventy-five cents.[11] Today it is possible to buy a one terabyte (one trillion bytes) hard drive for significantly less than one hundred US dollars. Third, a wider range of actors are now interested gathering personal data.[12] Surveillance is not just for totalitarian regimes. Democratic governments have also invested heavily in surveillance technologies, particularly in the years since 9/11. There are even private data brokers – like Acxiom and PrivateEye – which collect personal data and then sell it, with claims like 'At PrivateEye.com we have access to millions of public records all in one spot!'[13] In fact, Acxiom estimates that it holds on average approximately 1500 pieces of data on each adult in the US.[14] Since government and law enforcement are amongst these companies' most important clients they have been dubbed 'Big Brother's Little Helpers'.[15]

These developments have led one commentator to describe data as 'the perspiration of the Information Age'.[16] Today, a mind-boggling array of information is collected and stored, including: contact information (postal address, phone numbers, email addresses); emergency contact information; educational records; library records; financial records (including credit and debit card use); communications data (including phone numbers dialled, originating numbers and the time and duration of calls); online searches; websites visited; retailer records; tax records; immigration records; drivers' licence information; vehicle registration records; health information; criminal records; and, flight bookings. Moreover, this list will grow still longer over the coming years thanks to the development of smart objects such as vehicles and home appliances. This so-called 'Internet of

[6] Paul A Zoch, *Ancient Rome: An Introductory History* (University of Oklahoma Press 2000).
[7] Frederic William Maitland, *Domesday Book & Beyond: Three Essays in the Early History of England* (Cambridge University Press 1987).
[8] Hyman Alterman, *Counting People: The Census in History* (Harcourt, Brace & World 1969).
[9] Paul Rosenzweig, 'Privacy and Counter-Terrorism: The Pervasiveness of Data' (2010) 42 Case Western Journal of International Law 625, 627.
[10] A zettabyte (ZB) is 1000^7 bytes. A yottabyte (YB) is 1000^8 bytes. By contrast, a megabyte is 1000^2 bytes.
[11] Rosenzweig (n XX) 628.
[12] Neil M Richards, 'The Dangers of Surveillance' (2013) 126 Harvard Law Review 1934.
[13] <www.privateeye.com> accessed 9 May 2014.
[14] Rosenzweig (n XX) 631.
[15] Chris Jay Hoofnagle, 'Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement' (2004) 29 North Carolina Journal of International Law and Commercial Regulation 595.
[16] Daniel J Solove, 'Privacy and Power: Computer Databases and Metaphors for Information Privacy' (2001) 53 Stanford Law Review 1393, 1407.

Things' may enable new conveniences, but will also subject 'more and more previously unobservable activity to electronic measurement, observation, and control'.[17]

Information databases have been used in a number of contexts. An example from the field of public health is Google's Predict and Prevent initiative, which sought to match online search query patterns with the actual occurrence of influenza. The aim of such 'disease surveillance' is to proactively identify hotspots where diseases may emerge in order to enable an earlier response to potential outbreaks.[18] Meanwhile, marketers have been using information databases for targeted marketing campaigns since the 1970s.[19] Companies organise and sort their master lists of customers, then identify and profile the most profitable ones and use the profile to hunt for other similar customers. In 2012 an article in the *New York Times* told of a marketing campaign by the superstore Target which had targeted pregnant women.[20] A data analyst was quoted as saying, 'We knew that if we could identify them in their second trimester, there's a good chance we could capture them for years … As soon as we get them buying diapers from us, they're going to start buying everything else too'.[21] The analyst had been given the task of constructing a pregnancy prediction model. The premise underlying this model was that women's shopping habits change during pregnancy. The model assigned all female customers a pregnancy prediction score, based on customers' purchases of 25 specific products. The model turned out to be fairly accurate, not only in telling whether or not a woman was pregnant but also in identifying the stage of her pregnancy. Women identified as pregnant were sent coupons timed to the specific stage of their pregnancy. The article went on to tell how one angry father – whose daughter had received coupons in the post – walked into a Target store outside Minneapolis, demanded to see the manager and said to him, 'My daughter's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?' The manager apologized, and a few days later phoned the father to apologize again. This time the father was rather embarrassed: 'I had a talk with my daughter … It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology'.

The degree of predictive potential offered by data analytics in these other contexts raises the question whether it could also be used as a counterterrorism tool. There have certainly been some bold claims of its potential effectiveness. Some management consultants and IT specialists, for example, have asserted that if such techniques had been in use before 9/11 the attacks of that day may not have occurred.[22] Indeed, following 9/11 the Pentagon launched the controversial Total Information Awareness (TIA) programme. This project – whose logo was the all-seeing Eye of Providence – proposed combining databases held by state and federal governments with private data held by companies like Acxiom to create 'a new kind of extremely large, omni-media, virtually-centralized, and semantically-rich information repository that is not constrained by today's limited commercial

[17] Richards (n XX) 1940.
[18] Google.org, 'Predict and prevent: an initiative to help prevent local outbreaks of emerging disease from becoming pandemics' (14 October 2008) <http://www.google.org/Predict_Prevent_Brief.pdf> accessed 9 May 2014.
[19] Solove, 'Privacy and Power' (n XX).
[20] Charles Duhigg, 'How Companies Learn Your Secrets' *New York Times* (New York 16 February 2012) <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?ref=magazine> accessed 10 May 2014.
[21] ibid.
[22] Louise Amoore and Marieke de Goede, 'Governance, risk and dataveillance in the war on terror' (2005) 43 Crime, Law & Social Change 149, 160.

database products'.[23] Network analysis would then be used to search for suspicious behaviour.[24] The programme was fiercely criticized and, in 2003, the US Senate voted to cease its funding.[25] Nonetheless, the legacy of TIA lives on. For example, across the US there is now a network of Fusion Centers whose stated goal is to detect and prevent all crimes and all hazards.[26] As the following description illustrates, Fusion Centers are reminiscent of TIA:

> Data-mining tools analyze a broad array of personal data culled from public- and private-sector databases, the Internet, and public and private video cameras. Fusion centers access specially designed data-broker databases containing dossiers on hundreds of millions of individuals, including their Social Security numbers, property records, car rentals, credit reports, postal and shipping records, utility bills, gaming, insurance claims, social network activity, and drug- and food-store records. Some gather biometric data and utilize facial-recognition software. On-the-ground surveillance is collected, analyzed, and shared as well[27]

Similarly, the documents leaked by Edward Snowden in 2013 revealed the existence of: the Prism programme, which allows the NSA and FBI to collect materials including search history, emails, file transfers and live chats from nine leading US Internet companies including Google, Facebook and Apple;[28] the XKeyscore programme, which allows analysts to search through vast databases containing the emails, online chats and browsing histories of millions of individuals without prior authorisation;[29] and, a secret court order requiring Verizon to give the NSA information on all telephone calls in its systems, both with the US and between the US and other countries, on an 'ongoing, daily basis'.[30]

Roger A Clarke – who first coined the term 'dataveillance' in the 1980s – draws a distinction between personal dataveillance and mass dataveillance.[31] Personal dataveillance involves 'subjecting an identified individual to monitoring'.[32] In general, a specific reason will exist for the investigation or monitoring. By contrast, in mass dataveillance 'groups of people are monitored in order to generate suspicion about particular members of the population'.[33] In a co-authored article, the former Head of the Information Awareness Office John Poindexter – who resigned in 2003 after the TIA programme

[23] 'EPIC Analysis of Total Information Awareness Contractor Documents' (Electronic Privacy Information Center 2003) <http://epic.org/privacy/profiling/tia/doc_analysis.html> accessed 10 May 2014.
[24] Jeffrey Rosen, 'The Naked Crowd: Balancing Privacy and Security in an Age of Terror' (2004) 46 Arizona Law Review 607.
[25] Solove, '"I've Got Nothing to Hide"' (n XX).
[26] Michael German and Jay Stanley, *What's Wrong with Fusion Centers?* (American Civil Liberties Union 2007).
[27] Danielle Keats Citron and David Gray, 'Addressing the Harm of Total Surveillance: A Reply to Professor Neil Richards' (2013) 126 Harvard Law Review Forum 262, 264.
[28] Glenn Greenwald and Ewen MacAskill, 'NSA Prism program taps in to user data of Apple, Google and others' *The Guardian* (London 7 June 2013) <http://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data> accessed 10 May 2014.
[29] Glenn Greenwald, 'XKeyscore: NSA tool collects "nearly everything a user does on the internet"' *The Guardian* (London 31 July 2013) <http://www.theguardian.com/world/2013/jul/31/nsa-top-secret-program-online-data> accessed 10 May 2014.
[30] Glenn Greenwald, 'NSA collecting phone records of millions of Verizon customers daily' *The Guardian* (London 6 June 2013) <http://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order> accessed 10 May 2014.
[31] Roger A Clarke, 'Information Technology and Dataveillance' (1988) 31 Communications of the ACM 498; Roger Clarke, 'Profiling: A Hidden Challenge to the Regulation of Data Surveillance' (1993) 4 Journal of Law and Information Science 403.
[32] Clarke, 'Profiling' (n XX) 403.
[33] ibid 403.

was defunded[34] – has argued that both types of dataveillance are necessary.[35] He explains that two types of database searches/queries are possible. Subject-based queries start with known suspects and look for links to other suspects, people, places, things, or suspicious activities. On the other hand, pattern-based queries look for patterns of activity indicative of a terrorist plot. The hypothesis is:

> If terrorists plan to launch an attack, the plot must involve people (the terrorists, their financiers, and so forth). The transactions all these people conduct will manifest in databases owned by public, commercial and government sectors and will leave a signature – detectable clues – in the information space[36]

Pattern-based queries seek to imagine and understand the signatures that terrorist plots will create and then use advanced search methods to find instances of these in the information space before the plots materialise. So, for example, a traveller who buys fertilizer and a one-way ticket and takes flying lessons might be tagged for further investigation.[37]

One of the claimed benefits of pattern-based queries is that terrorists will find it hard to engage in counter-surveillance:

> Furthermore, because the patterns relied on – particularly those that are computer-detected – will be non-obvious, terrorists will find it difficult to engage in counter-surveillance tactics. They will not know the patterns to avoid. Unless cash-only transactions were engaged in, a transactional signature would be unavoidable, even if that signature were distributed amongst proxies[38]

The most important claimed benefit, however, is that pattern-based queries have the potential to identify individuals who have not yet aroused any suspicion:

> In the commercial context, these individuals are called 'potential customers.' In the terrorism context, they are often called 'clean skins' because there is no known derogatory information connected to their names or identities. In this latter context, the individuals are dangerous because nothing is known of their predilections. For precisely this reason, this form of data analysis is sometimes called 'knowledge discovery,' as the intention is to discover something previously unknown about an individual. There can be little doubt that data analysis of this sort can prove to be of great value[39]

Whilst this argument may sound appealing – particularly given the fears of an insider threat that are evident in several chapters in this book – it is a double-edged sword. Pattern-based queries envisage 'the State having broad access to many individuals' personal information, when there is no basis for even a suspicion of wrong-doing'.[40] This raises important questions about the privacy of individuals.

---

[34] Hannah C Bloch-Wehba, 'Global Governance in the Information Age: The Terrorist Finance Tracking Program' (2013) 45 *New York University Journal of International Law and Politics* 595.
[35] Robert Popp and John Poindexter, 'Countering Terrorism through Information and Privacy Protection Technologies' [2006] IEEE Security & Privacy 18.
[36] ibid 18-19.
[37] Rosen (n XX).
[38] Wayne N Renke, 'Who Controls the Past Now Controls the Future: Counter-Terrorism, Data Mining and Privacy' (2006) 43 Alberta Law Review 779, 788.
[39] Rosenzweig (n XX) 632-33.
[40] Renke (n XX) 796.

Before considering these questions, however, it is important to first evaluate the likely effectiveness of pattern-based queries as a counterterrorism tool.

The effectiveness of dataveillance as a counterterrorism tool

In its 2008 report *Protecting Individual Privacy in the Struggle Against Terrorists* the US National Research Council insisted that, when evaluating an information-based programme, the first task should be to assess its effectiveness:

> Too frequently the argument is heard that national security is too important and the terrorist threat too great to pause to ask hard questions of the systems to be deployed to protect the nation. In the committee's view, that is the wrong approach. It is precisely because national security is important and the threats to it are great that it is so important to ensure that the systems to be deployed to protect the nation are effective and are consistent with U.S. values[41]

This is undoubtedly correct. As one commentator has observed, 'data mining and behavioral surveillance programs that fail the effectiveness test protect neither the nation's privacy nor its security'.[42]

Unsurprisingly, security services are reluctant to disclose details of cases in which they claim that dataveillance has proven successful for fear that it will aid terrorist groups by revealing methods and capabilities. Nonetheless, in a speech in Berlin in June 2013 President Obama declared: 'We know of at least 50 threats that have been averted because of this [NSA mass surveillance] information'.[43] Similarly, in *ACLU v Clapper* Judge Pauley claimed that 'The effectiveness of bulk telephony metadata collection cannot be seriously disputed', citing three examples taken from Congressional testimony.[44] In sharp contrast, in *Klayman v Obama* Judge Leon stated that 'The Government does *not* cite a single instance in which analysis of the NSA's bulk metadata collection actually stopped an imminent attack, or otherwise aided the Government in achieving any objective that was time-sensitive in nature'.[45] He rejected the three examples cited by Judge Pauley, saying that none of them involved any apparent urgency. In fact, a detailed study of 225 cases involving individuals recruited by al Qaeda or similar groups and charged in the US with an act of terrorism since 9/11 concluded that surveillance of US phone metadata has 'had no discernible impact on preventing acts of terrorism and only the most marginal of impacts on preventing terrorist-related activity, such as fundraising for a terrorist group'.[46] It should also be noted that none of the three cases cited by Judge Pauley involved suspicion-less pattern-based queries. In all three instances the authorities were already investigating known suspects and only used metadata analysis to ensure that all of the suspect's contacts had been identified.

---

[41] US National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment* (National Academies Press 2008) 47.
[42] Susan Landau, 'The NRC Takes on Data Mining, Behavioral Surveillance, and Privacy' [2009] IEEE Security & Privacy 58, 58.
[43] Jackie Calmes, 'Obama Says Surveillance Helped in Case in Germany' *New York Times* (New York 19 June 2013) <http://www.nytimes.com/2013/06/20/world/europe/obama-in-germany.html?_r=0> accessed 11 May 2014.
[44] 959 F Supp 2d 724 (NY), 755.
[45] 957 F Supp 2d 1 (DDC), 40 (emphasis original).
[46] Peter Bergen and others, *Do NSA's Bulk Surveillance Programs Stop Terrorists?* (New America Foundation 2014), 1.

One of the principal difficulties with pattern-based queries is modelling. Terrorist conduct preceding an attack is likely to be designed to appear legitimate. So, unlike offences like credit card fraud, 'it will not be outlier conduct in the midst of legitimate conduct, but apparently legitimate conduct in the midst of legitimate conduct'.[47] Moreover, in the commercial context analysts have enormous datasets from which to construct models. By contrast, successful terrorist attacks are relatively rare and so the evidential basis for constructing patterns is small.[48] One response to this might be to target lower-level events which are commonly associated with terrorism and which occur more frequently, such as fund transfers and recruitment activities. However, since many of these activities are, in themselves, perfectly legal and may be carried out for purposes other than terrorism such an approach risks casting the net too wide.[49] A further difficulty is that constructing models based on past attacks is reactive, and so may miss innovative forms of attack.

Pointing to the attacks of 9/11 as an example, the American Civil Liberties Union has argued that what is needed is not more information, but better use of existing information: 'You don't find a needle in a haystack by bringing in more hay'.[50] In terms of pattern-based queries, more important than the volume of data per se is the knock-on effect on the number of false positives. False positives are individuals who are wrongly deemed worthy of suspicion. So, in our earlier example, the aspiring pilot who purchases fertilizer might merely be 'a retired businessperson who was a gardening aficionado'.[51] Since pattern-based queries work 'on the principle of draining the swamp to catch the snake'[52] an extremely high level of accuracy has to be achieved. Yet not only are there the modelling difficulties outlined in the previous paragraph, there are also the problems caused by incomplete, incorrect, incomprehensible and inconsistent data ('garbage in, garbage out'[53]) – which are exacerbated still further by the prevalence of identity theft and the difficulties that victims of this crime have trying to expunge prior convictions wrongly attached to them.[54] Indeed, there are numerous examples of individuals being wrongly identified by data mining.[55] When dealing with huge numbers of individuals, even extremely high accuracy rates result in large numbers of false positives:

> Out of an American-sized population of 250m, a 99.9 per cent level of accuracy in surveillance still means placing approximately 250,000 Americans at risk: even a 99.99 per cent level would still affect about 25,000 people, though the unacceptability of this politically might depend on which sectors they came from[56]

Added to this is the possibility of false negatives (individuals who are wrongly deemed to not be worthy of suspicion). Experts concede that 'the technical reality [is] that the number of false negatives

[47] Renke (n XX) 794.
[48] K A Taipale, 'Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data' (2003) 5 Columbia Science and Technology Law Review 1, 35.
[49] Renke (n XX) 794.
[50] 'Q&A on the Pentagon's "Total Information Awareness" Program' (American Civil Liberties Union 20 April 2003) <https://www.aclu.org/technology-and-liberty/qa-pentagons-total-information-awareness-program> accessed 11 May 2014.
[51] Rosen (n XX) 611.
[52] Michael Levi and David S Wall, 'Technologies, Security, and Privacy in the Post-9/11 European Information Society' (2004) 31 Journal of Law and Society 194, 207.
[53] Renke (n XX) 791.
[54] Stephen W Dummer, 'Falses Positives and Secure Flight Using Dataveillance When Viewed Through the Ever Increasing Likelihood of Identity Theft' (2006) 11 Journal of Technology Law & Policy 259.
[55] David Gray and Danielle Citron, 'The Right to Quantitative Privacy' (2013) 98 Minnesota Law Review 62.
[56] Levi and Wall (n XX) 207.

can never be zero'.[57] In the terrorism context, the likely number of false negatives is increased by the fact that many terrorists use privacy enhancing technologies (such as anonymisation and encryption) and privacy enhancing strategies (such as the use of cash and barter).[58]

So even if pattern-based queries do have security benefits – which, as we have seen, has been contested – they may also have security costs. Resources and time could be spent investigating (large numbers of) false positives. Meanwhile false negatives may escape attention. In addition, depending on the features of the profile used, the false positives returned by a pattern-based query could include a disproportionate number of members from particular minority or ethnic groups. This has the potential to generate resentment and ill-feeling, particularly given the suspicion-less nature of pattern-based queries. Empirical research in the UK has found that Muslim communities are more likely to resent counterterrorism measures if they do not require individualised suspicion and are perceived as targeting whole communities. So, for example, the broad and indiscriminate use of powers like stop and search – and, by extension, mass dataveillance – are more likely to generate resentment than the use of individualised measures such as Control Orders.[59] This is at odds with the importance counterterrorism strategies place on fostering community cohesion in order to make communities more resilient to radicalization.[60]

Liberty

As noted above, pattern-based queries involve the State having access to many individuals' personal information without any individualized suspicion. This raises important questions regarding both the scope of privacy interests in this context and the weight that should be attached to these interests.

What does protecting privacy mean in this context and is it important?

Some have suggested that dataveillance does not raise issues of privacy at all. Judge Posner, for example, has emphasised the limited role human beings play in data mining:

> [M]achine collection and processing of data cannot, as such, invade privacy. Because of their volume, the data are first sifted by computers, which search for names, addresses, phone numbers, etc., that may have intelligence value. This initial sifting, far from invading privacy (a computer is not a sentient being), keeps most private data from being read by any intelligence officer[61]

[57] US National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists* (n XX) 40.
[58] Levi and Wall (n XX).
[59] Joint Committee on the Draft Enhanced Terrorism Prevention and Investigation Measures Bill, *Oral Evidence Taken on Wednesday 24 October 2012* <http://www.parliament.uk/documents/joint-committees/Draft%20ETPIMS%20Bill/HC%20495%20iii%2024%20October%202012%20Corrected.pdf> accessed 12 May 2014, 4-5.
[60] See, eg, Home Office, *CONTEST: The United Kingdom's Strategy for Countering Terrorism* (Cm 8123, 2011).
[61] Richard A Posner, 'Our Domestic Intelligence Crisis' *Washington Post* (Washington DC, 21 December 2005) <http://www.washingtonpost.com/wp-dyn/content/article/2005/12/20/AR2005122001053.html> accessed 16 May 2014.

Poindexter and Popp, meanwhile, define privacy even more narrowly. In their opinion, 'personal privacy is only violated if the violated party suffers some tangible loss, such as unwarranted arrest or detention'.[62]

Others accept that issues of privacy do arise, but argue that little weight should be attached to these concerns. Solove describes this as the 'I've got nothing to hide' argument.[63] This perspective emphasises that mass dataveillance programmes are concerned only with those pieces of information which are likely to be useful in identifying behaviour that threatens national security. This means: first, that the vast majority of information that dataveillance programmes collect on law-abiding citizens will be innocuous; and, second, that the vast majority of law-abiding citizens' embarrassing or discreditable information is unlikely to be collected in the first place, since this is not the sort of data that dataveillance programmes are likely to collect. Moreover, even if some embarrassing information did happen to be collected it will only be exposed to a few unknown officials at worst.

These suggestions adopt an unduly narrow view of what privacy entails and overlook important concerns about the societal effects of mass dataveillance. Solove states that the nothing to hide argument misses the point, because it assumes that privacy is about nothing more than 'hiding bad things'.[64] This tendency is exacerbated, he says, by the common use of the Big Brother metaphor and references to Orwell's *Nineteen Eighty-Four* and Bentham's Panopticon.[65] Where dataveillance is concerned, 'The most insidious aspect of the surveillance of Big Brother is missing in the context of databases: human judgment about the activities being observed (or the fear of that judgment)'.[66] Solove accordingly advances a broader taxonomy of privacy.[67] This not only encompasses surveillance (information collection) and disclosure (information dissemination) – the problems which the nothing to hide argument focuses on – but also information processing: how information that has already been collected is handled. It is here that dataveillance raises particular concerns.

The first set of concerns focuses on the aggregation of individuals' data. Even if each discrete piece of information about an individual is not in itself something they would hide, 'Well-established techniques in the field of information technology such as data-mining make it possible for those so-called meaningless bits zooming in and out of the ether of global networks and public and private databases to be quickly and inexpensively reassembled'.[68] This 'surveillant assemblage'[69] could lead to far more being discovered about the individual than would be available by, for example, surveilling their movements.[70] It is quite conceivable, then, that 'by combining pieces of information we might not care to conceal, the government can glean information about us that we might really want to conceal'.[71] Moreover, this information processing may occur without the individuals' knowledge or

[62] Popp and Poindexter (n XX) 24.
[63] Solove, '"I've Got Nothing to Hide"' (n XX).
[64] Solove, '"I've Got Nothing to Hide"' (n XX) 764.
[65] Solove, 'Privacy and Power' (n XX).
[66] Solove, 'Privacy and Power' (n XX) 1417.
[67] Daniel J Solove, 'A Taxonomy of Privacy' (2006) 154 University of Pennsylvania Law Review 477.
[68] Ian Kerr and Jena McGill, 'Emanations, Snoop Dogs and Reasonable Expectations of Privacy' (2007) 52 Criminal Law Quarterly 392, 416.
[69] Kevin D Haggerty and Richard V Ericson, 'The Surveillant Assemblage' (2000) 51 British Journal of Sociology 605.
[70] Julia Alpert Gladstone, 'A Call from the Panopticon to the Judicial Chamber "Expect Privacy!"' (2006) 1 Journal of International Commercial Law and Technology 62.
[71] Solove, '"I've Got Nothing to Hide"' (n XX) 766.

involvement.[72] The individual may not have consented to their information being used in this way, may not know that their information is being used in this way and may not know how their data will be used in the future. Indeed, as the documents leaked by Snowden illustrate, individuals may not even know that the dataveillance programme exists. These issues of exclusion and secondary use of individuals' data raise questions of 'technological due process'.[73] The individual is broken down into a series of data flows and then reassembled. Not only could individuals find themselves red-flagged at airports[74] or prevented from crossing borders[75] on the basis of this 'decorporealized body'.[76] The exclusion of individuals from the process means that in practice it is very difficult to correct any errors in one's virtual 'data double'.[77]

This leads to a related concern: the disparity of power between individuals and the state. As the earlier example involving Target illustrated, data-driven marketing gives companies considerable power in relation to their customers. Marketers can sort consumers into categories and then allocate opportunities on the basis of this classification. But when this approach is applied in the counterterrorism context, so that the state attempts to proactively identify individuals who pose a risk by detecting behaviour that is perceived as suspicious, 'The power of sorting can bleed imperceptibly into the power of discrimination'.[78] This results in the construction of 'suspect populations'.[79] As Amoore and de Goede explain, such constructions rest

> upon the representations of two worlds of globalization: one populated by legitimate and civilized groups whose normalised patterns of financial, tourist and business behaviour are to be secured; and another populated by illegitimate and uncivilized persons whose suspicious patterns of behaviour are to be targeted and apprehended[80]

Moreover, using the example of the war on terrorist finance and its effect on migrant communities, they explain that these two worlds are mutually reinforcing.[81] To control the risk of terrorist financing post-9/11 tougher financial regulatory regimes were introduced and informal money transmitters were criminalized and suppressed. One effect of this was to deny migrants a relatively cheap and efficient method of sending money back to their families, forcing many to resort to cash transfers – notwithstanding the fact that cash itself was increasingly being regarded as suspicious. Amoore and de Goede thus conclude, 'From downtown banking halls to city airport terminals, the techniques of dataveillance will continually inscribe and reinscribe a manufactured border between the licit and illicit worlds'.[82]

---

[72] ibid.
[73] Danielle Keats Citron, 'Technological Due Process' (2008) 85 Washington University Law Review 1249.
[74] Stephen W Dummer, 'Secure Flight and Dataveillance, A New Type of Civil Liberties Erosion: Stripping Your Rights When You Don't Even Know It' (2006) 75 Mississippi Law Journal 583.
[75] Amoore and de Goede (n XX).
[76] Haggerty and Ericson (n XX) 611.
[77] Haggerty and Ericson (n XX) 611.
[78] Richards (n XX) 1957.
[79] Levi and Wall (n XX) 200.
[80] Amoore and de Goede (n XX) 167-68.
[81] See also the chapter by Sproat in this collection.
[82] Amoore and de Goede (n XX) 168.

Lastly, mass dataveillance raises concerns about normalization, in two respects. First, Richards has warned of the potential effect of the 'normalizing gaze of surveillance'[83] on individuals' intellectual privacy. Surveillance, he argues, 'inclines us to the mainstream and the boring':

> [W]hen we are watched while engaging in intellectual activities, broadly defined – thinking, reading, websurfing, or private communication – we are deterred from engaging in thoughts or deeds that others might find deviant. Surveillance thus menaces our society's foundational commitments to intellectual diversity and eccentric individuality[84]

Indeed, using data from Google Trends, a recent study of search terms from before and after the Snowden revelations has found substantial empirical evidence of a chilling effect – not only in respect of search terms that might be deemed suspicious but also ones that could be personally embarrassing.[85] So, as well as privacy, mass dataveillance also raises issues of freedom of speech and expression. Furthermore, as mass dataveillance is routinized people's expectations of privacy will be eroded. The effect is that perspectives like the ones outlined at the start of this section – which define privacy narrowly and/or attach little weight to privacy concerns – become internalized and normalized. So, over time, people 'will not recognize and no less expect, that privacy was once possible in that encroached area'.[86] This observation is especially relevant to the Fourth Amendment to the US Constitution, whose protection of privacy is dependent on individuals' expectations. It is to this that we now turn.

Legal protection of the right to privacy

The protection of privacy takes a number of different forms, ranging from domestic legislation[87] to regional agreements[88] and from sector specific regimes[89] to overarching ones.[90] The focus here is on the protection conferred by the Fourth Amendment to the US Constitution and Article 8 of the European Convention on Human Rights. These provisions not only inform the content of other legislative and regulatory regimes. Their entrenched status is significant in the context of counterterrorism, where the temptation to erode the protection of rights is particularly strong.

In the US, the principal source of protection for individuals' privacy rights vis-à-vis the government is the Fourth Amendment. This guarantees a right against unreasonable searches and seizures and states that warrants may only be issued with probable cause. Importantly, for the purposes of the Fourth Amendment a search occurs if the government violates a person's reasonable expectation of privacy.[91] The scope of this protection is diminished, however, by two doctrines. The first is the public observation doctrine, according to which law enforcement officers can freely make observations from

[83] Richards (n XX) 1950.
[84] Richards (n XX) 1948.
[85] Alex Marthews and Catherine Tucker, 'Government Surveillance and Internet Search Behavior' (SSRN, 24 March 2014) <http://dx.doi.org/10.2139/ssrn.2412564> accessed 16 May 2014.
[86] Gladstone (n XX) 65.
[87] For example, the UK's Regulation of Investigatory Powers Act 2000.
[88] For example, the Data Protection Directive (95/46/EC).
[89] Examples include the Communications Data Privacy Directive Directive 2002/58/EC (telecommunications) and the Agreement between the European Union and the United States of America on the Processing and Transfer of Passenger Name Record (PNR) Data by Air Carriers to the United States Department of Homeland Security (OJ 2007 L 204/18).
[90] For example, the UK's Data Protection Act 1998.
[91] *Katz v United States*, 389 US 347 (1967).

any place where they lawfully have a right to be.[92] So, for example, police officers may stand on the street and watch through open windows and may monitor movements on public roads.[93] They may also use devices like binoculars and beeper-type trackers to enhance their observational abilities.[94] The second is the third party doctrine, which states that information shared with a third party falls outside the scope of the Fourth Amendment. So, for example, a person has no expectation of privacy in data they share with their bank.[95] Importantly for present purposes, in reliance on this doctrine the Supreme Court in *Smith v Maryland* has also held that the person who makes a telephone call surrenders the right to privacy in the numbers dialled.[96]

As has been explained, the principal claimed benefit of pattern-based queries is that they aim to identify clean skins. The corollary of this, however, is that they involve the state having access to the personal information of vast numbers of individuals who have not aroused any suspicion. This raises the question whether pattern-based queries violate the Fourth Amendment's requirement of probable cause, which in turn hinges on whether mass dataveillance programmes violate individuals' reasonable expectations of privacy. Much of the information these programmes aggregate – such as financial transactions, communications data, online searches and website browsing history – is covered by the public observation and third party doctrines. The Fourth Amendment would not apply to any of these pieces of data taken on their own. Arguably, however, the process of aggregation should be regarded as constitutionally significant. Even if each individual piece of data was publicly observable and/or disclosed to a third party, the amalgamation of this information produces a 'transactional narrative'[97] that would not otherwise be available. This same issue arises in relation to the protection against unreasonable searches and seizures conferred by section 8 of the Canadian Charter of Rights and Freedoms. Applying the test developed by the Canadian Supreme Court, on their own none of the pieces of data involved in a dataveillance programme may relate to a person's 'biographical core'.[98] But this core could be reconstituted using this data via the process of aggregation.[99]

One method of protecting privacy against the possibility of data aggregation would be to adopt a quantitative approach to the Fourth Amendment.[100] Such an approach finds some support in the decision of the US Supreme Court in *United States v Jones*.[101] Antoine Jones was suspected of drug trafficking. Without a valid warrant, federal law enforcement agents installed a GPS device to his car and used it to monitor his travel on public roads around the clock for four weeks. The tracking data showed that he made regular visits to stash houses. He was subsequently convicted of drugs conspiracies offences and sentenced to life imprisonment. On appeal, the US Court of Appeals for the District of Columbia Circuit reversed, ruling that there had been a violation of the Fourth Amendment. Before the US Supreme Court the government contended that there had been no search for the purposes of the Fourth Amendment. The car's journeys were on public roads and were publicly observable, so the driver had no reasonable expectation of privacy. The Supreme Court, however, held unanimously that the Fourth Amendment applied. The majority opinion – written by

[92] *Florida v Riley*, 488 US 445 (1989).
[93] *United States v Knotts*, 460 US 276 (1983).
[94] *United States v Knotts*, 460 US 276 (1983).
[95] *United States v Miller*, 425 US 435 (1976).
[96] *Smith v Maryland*, 442 US 735 (1979).
[97] Renke (n XX) 808.
[98] *R v Plant*, [1993] 3 SCR 281.
[99] Renke (n XX).
[100] Gray and Citron, 'The Right to Quantitative Privacy' (n XX).
[101] *United States v Jones*, 132 SCt 945 (2012).

Justice Scalia and joined by Chief Justice Roberts with Justices Kennedy, Thomas and Sotomayor – held that the installation of the GPS device involved a search because it was accomplished by a trespass for the purpose of obtaining information. Importantly, though, Justice Alito's opinion – written for himself and Justices Ginsburg, Breyer and Kagan and with which Justice Sotomayor expressed some sympathy – rejected the trespass-based reasoning and instead adopted a quantitative approach. In the past, he said, resource constraints meant that long-term surveillance could only be justified in investigations of 'unusual importance'.[102] So, whilst short-term monitoring of a person's movements on public streets accords with people's reasonable expectations of privacy, long-term monitoring does not. It was not necessary to identify the precise point at which surveillance crosses this threshold, since four weeks was 'surely' long enough to impinge on expectations of privacy.[103]

Whether a quantitative approach to the Fourth Amendment will be applied in cases involving mass dataveillance remains to be seen. Early signs are mixed. In *Klayman v Obama* Judge Leon relied on *United States v Jones* to distinguish *Smith v Maryland*, ruling that it is 'significantly likely'[104] that government collection and storage of telephone metadata for five years violates reasonable expectations of privacy. On the other hand, in *ACLU v Clapper* Judge Pauley emphasised that in *United States v Jones* the majority decided the case on the basis of trespass. Since the majority did not adopt a quantitative approach and did not overrule *Smith v Maryland*, he concluded that 'The collection of breathtaking amounts of information unprotected by the Fourth Amendment does not transform that sweep into a Fourth Amendment search.'[105] Academic reaction to Justice Alito's opinion has also been mixed. Kerr argues that Alito's 'mosaic approach' raises a series of difficult questions regarding when the aggregation of data triggers the Fourth Amendment and that, at a time when technologies are developing rapidly, the uncertainty generated by these questions is undesirable.[106] For others, by contrast, Alito's opinion represents an opportunity to develop a radical new approach to the Fourth Amendment which is more protective of the 'right to quantitative privacy.'[107]

An alternative course, which has been charted by Cole, is comparative constitutionalism.[108] In its jurisprudence on the Article 8 right to respect for one's private and family life the European Court of Human Rights (ECtHR) has adopted a more nuanced approach to questions of privacy. In contrast to the public observation and third party doctrines, the Court has held that Article 8 'is not limited to the protection of an "inner circle" in which the individual may live his own personal life as he chooses and to exclude therefrom entirely the outside world'.[109] On the contrary, 'private life is a broad term not susceptible to exhaustive definition', which includes – but is not limited to – 'a right to identity and personal development, and the right to establish and develop relationships with other human beings and the outside world'.[110] A person's reasonable expectations of privacy may be a significant factor, but are not necessarily conclusive.[111] The Court has held that Article 8 encompasses the

[102] *United States v Jones*, 132 SCt 945, 963 (2012).
[103] *United States v Jones*, 132 SCt 945, 964 (2012).
[104] 957 F Supp 2d 1 (DDC), 37.
[105] 959 F Supp 2d 724 (NY), 752.
[106] Orin S Kerr, 'The Mosaic Theory of the Fourth Amendment' (2012) 111 *Michigan Law Review* 311.
[107] Gray and Citron, 'The Right to Quantitative Privacy' (n XX).
[108] David Cole, 'Preserving Privacy in a Digital Age: Lessons of Comparative Constitutionalism' in Fergal David, Nicola McGarrity and George Williams (eds), *Surveillance, Counter-Terrorism and Comparative Constitutionalism* (Routledge 2014).
[109] *Shimovolos v Russia* (2014) 58 EHRR 26, para 64.
[110] *Uzun v Germany* (2011) 53 EHRR 24, para 43.
[111] *Uzun v Germany* (2011) 53 EHRR 24, para 44.

systematic collection and storing of data by security services on particular individuals, even if the data was collected in a public place[112] or concerned exclusively the person's professional or public activities.[113] So, for example, Article 8 applied where a person's name had been registered in a surveillance database which collected information about his movements by train or air.[114] And, in a case with similar facts to *United States v Jones*, Article 8 applied where GPS tracking data had been collected for almost three months from a suspect's car.[115]

According to Article 8(2), an interference with a person's Article 8 right may be justified if it was: first, in accordance with the law;[116] and, second, necessary in a democratic society for one of several stipulated purposes, including the interests of national security. The second of these conditions requires that the 'interference corresponds to a pressing social need and, in particular, that it is proportionate to the legitimate aim pursued'.[117] The assessment of proportionality will include consideration of whether other less intrusive methods of investigation were available which would have proved effective. In a similar vein, the law governing wiretapping warrants in Canada requires an assessment of investigative necessity.[118] To obtain a warrant, investigators must show that no other reasonable alternative method of investigation is available. This is not as strict as a last resort test, but requires more than a claim that it is the most efficacious method available.[119]

Pointing to these other sources of law, Cole argues that:

> [A] legal system need not treat privacy as an on/off affair, but can – and in my view, should – recognize that private details of an individual's life can be gleaned by the gathering, recording, collation, and analysis of hundreds of pieces of information about the individual's purchases, travels, communications, contacts, and viewing and reading habits[120]

Recognition that privacy is implicated in such cases, coupled with an acceptance that there would be some flexibility in the procedural rules governing intrusions into privacy, would, Cole argues, allow a measure's necessity to be evaluated as part of the Fourth Amendment's reasonableness requirement. If other less intrusive means of investigation are available, the search may not be reasonable. He thus concludes that 'the Canadian and ECtHR approaches suggest that such assessments of the relative intrusiveness of different monitoring tactics may provide an important constraint on the use of new technologies'.[121] It is worth adding that, as well the degree of intrusiveness, evaluations of necessity should also include an assessment of a measure's utility as an investigative device. As explained above, there are significant concerns about the effectiveness of pattern-based queries. Before suspicion-less searches of (large numbers of) individuals' personal records can be deemed necessary,

[112] *Peck v United Kingdom* (2003) 36 EHRR 41.
[113] *Amann v Switzerland* (2000) 30 EHRR 843.
[114] *Shimovolos v Russia* (2014) 58 EHRR 26.
[115] *Uzun v Germany* (2011) 53 EHRR 24, discussed further in the chapter by Dickson in this collection.
[116] This means that 'the impugned measure must have some basis in domestic law and be compatible with the rule of law'. It must, therefore, 'be adequately accessible and foreseeable' and 'formulated with sufficient precision to enable the individual – if need be with appropriate advice – to regulate his conduct' (*MM v United Kingdom* App no 24029/07 (ECtHR, 13 November 2012).
[117] *Uzun v Germany* (2011) 53 EHRR 24, para 78.
[118] *Criminal Code*, RSC 1985, c C-46, s 186(1).
[119] *R v Araujo* [2000] 2 SCR 992.
[120] Cole (n XX) 110.
[121] Cole (n XX) 112.

questions regarding the likely number of false negatives and false positives – and the resultant opportunity cost in terms of resources – should also be addressed.

Conclusion

The balancing of competing interests may take different forms. Much discussion of dataveillance programmes assumes that a choice must be made between the demands of national security and individuals' privacy, so that one set of concerns must trump the other. But to balance competing interests can also mean to reconcile them without eroding the essence of either. In other words, it need not be a 'zero-sum game'.[122] This could potentially be the case in the context of mass dataveillance, as experts begin to develop technologies which have privacy protection embedded in their design but are still effective for counterterrorism.[123] But if it is necessary to balance in the sense of one-trumps-the-other, it is important to first of all carefully assess the different sets of interests that are at stake. As this chapter has shown, mass dataveillance raises important privacy-based concerns, which should not be overridden for the sake of pattern-based queries whose security benefits are speculative and which have potentially significant opportunity and collateral security costs.

[122] Zedner, 'Securing Liberty in the Face of Terror' (n XX) 511.
[123] Ann Cavoukian and Khaled El Emam, *Introducing Privacy-Protective Surveillance: Achieving Privacy and Effective Counter-Terrorism* (Information and Privacy Commissioner, Ontario, Canada 2013).