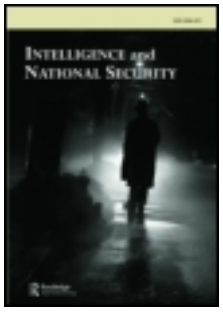

Intelligence and National Security
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/fint20

Introducing Social Media Intelligence (SOCMINT)

Sir David Omand, Jamie Bartlett & Carl Miller

Version of record first published: 28 Sep 2012.

To cite this article: Sir David Omand, Jamie Bartlett & Carl Miller (): Introducing Social Media Intelligence (SOCMINT), Intelligence and National Security, DOI:10.1080/02684527.2012.716965

To link to this article: http://dx.doi.org/10.1080/02684527.2012.716965

Intelligence and National Security
2012, 1–23, iFirst

ARTICLES

Introducing Social Media Intelligence (SOCMINT)

SIR DAVID OMAND,* JAMIE BARTLETT AND CARL MILLER

ABSTRACT We introduce the latest member of the intelligence family. Joining IMINT, HUMINT, SIGINT and others is 'SOCMINT' – social media intelligence. In an age of ubiquitous social media it is the responsibility of the security community to admit SOCMINT into the national intelligence framework, but only when two important tests are passed. First, that it rests on solid methodological bedrock of collection, evidence, verification, understanding and application. Second, that the moral hazard it entails can be legitimately managed. This article offers a framework for how this can be done.

Introduction

On Thursday 4 August 2011, Mark Duggan was shot and killed by a police officer in Tottenham. By the morning of the 6th, social media channels showed burgeoning hostility, peppered with explicit threats against the police. From the 7th, social media information indicated the possible spread of disorder to other parts of London, then England. Over the next few days, content indicating criminal intent or action ratcheted in huge numbers through both open-source social networking, such as Twitter, and closed system networks, such as the BlackBerry Messaging Service and closed groups such as chat forums. Similarly, huge numbers of messages appeared trying to provide information to the police, either about an outbreak of disorder or the identities of the people behind it.[1]

In the aftermath, the police acknowledged that they had been insufficiently equipped to deal with intelligence gathering via social media.

*Email: david.omand@kcl.ac.uk

[1]Her Majesty's Inspectorate of the Constabulary (HMIC), *The Rules of Engagement: A Review of the August 2011 Disorders* (London: Crown Copyright 2011) especially pp.36–9.

ISSN 0268-4527 Print/ISSN 1743-9019 Online/12/000001-23 © 2012 Taylor & Francis
http://dx.doi.org/10.1080/02684527.2012.716965

One intelligence professional said it was like 'searching the British Library for a page in a book without an index to refer to'.[2] Social media did not fit into their systems of receiving, corroborating, prioritizing and disseminating information, and therefore was not properly acted on. Her Majesty's Chief Inspector of Constabulary noted: 'With some notable individual exceptions, the power of this kind of media (both for sending out and receiving information) is not well understood and less well managed'.[3] He concluded that 'the police have much to learn about social media, and the quickly shifting modern communications of today'.[4]

Since then, Government has reacted. The Metropolitan Police has established a social media hub, in time for the London Olympics. A number of police forces in the UK and elsewhere are believed to be trialling various types of automated social media collection and analysis to collect information to help criminal investigations and gauge the 'temperature' of communities they are working with.[5] Police forces have used Flickr to crowd source identifications of suspects from photographs. In the UK, the Ministry of Defence's Cyber and Influence Science and Technology Centre has released a number of calls for research to develop capabilities including 'cyber situational awareness', 'influence through cyber' and 'social media monitoring and analysis: crowd sourcing'.[6] Underlying these developments is significant planned public investment in the capabilities that will allow the authorities to continue to access communications data and access under warrant where necessary the content of internet communications including social media. In the UK, new legislation has been proposed to ensure law enforcement agencies maintain the ability to tackle crime and terrorism as criminals use modern technology and new ways of communicating in the digital space to plan and commit crime.

This rapid growth of interest by law enforcement in intelligence derived from social media (which we term SOCMINT) prompts questions about the methodological and ethical framework within which it will be used. Public acceptability lies at the heart of any form of intelligence collection, and this can only be secured if SOCMINT is properly used and properly authorized. This article suggests a framework for how this can be achieved.

[2]Ibid., p.31.
[3]Ibid., p.30.
[4]Ibid.
[5]'Facebook Crimes Probed by Humberside Police', *Hull Daily Mail*, 24 August 2011, <www.thisishullandeastriding.co.uk/Facebook-crimes-probed-Humberside-Police/story-13191231-detail/story.html> (accessed 17 April 2012); see also Westminster City Council's 'Your Choice' programme: *Choose Life, Not Gangs: Problem Kids Told to Clean Up or Face the Consequence* (City of Westminster, 29 September 2011), <www.westminster.gov.uk/press-releases/2011-09/choose-life-not-gangs-problem-kids-told-to/> (accessed 17 April 2012).
[6]Ministry of Defence and Centre for Defence Enterprise, Cyber and Influence Science and Technology Centre, *CDE Call for Research Proposals*, 1 November 2011, <www.science.mod.uk/controls/getpdf.pdf?603> (accessed 17 April 2012).

The Age of Social Media

We live in the age of social media. Facebook, Twitter, Google+ and LinkedIn are all examples of the rapid transfer of people's lives – interactions, identities, arguments and views – onto a new kind of public and private sphere; a vast digital social commons. This transfer is happening on an unprecedented scale. On Facebook alone, 250 million photos are added per day,[7] as are 200 million tweets on Twitter.[8] There are four billion video views per day on YouTube.[9]

As people transfer more of their lives onto social media platforms, they become an increasingly significant public space, and therefore of interest to, and used by, public bodies. Understanding the content of social media presents an opportunity for public bodies better to understand, and respond to, the public they serve. Public health experts are learning to scan tweets and search requests to identify pandemics earlier than traditional methods.[10] US psychologists believe Facebook contains valuable indicators of mental health, and indeed the social media profiles of a number of participants in school shootings, such as the suspect in the Ohio School Shooting, TJ Lane, seem to show some indicative content.[11] The United Nations believes that tapping into social media can help tackle global unemployment and food insecurity.[12]

Social media spaces are also now significantly relevant to security and public safety. Facebook has been used to try to hire hitmen, groom the targets of paedophiles, violate restraining orders, steal identities and fatally cyberbully victims.[13] Al-Qaeda's Somali affiliate Al-Shabaab runs a twitter account, whilst pirates operating in the Gulf of Aden use blogs, Twitter and

[7]'The Value of Friendship', *Economist*, 4 February 2012, <http://www.economist.com/node/21546020> (accessed 17 April 2011).

[8]Twitterblog, *200 Million Tweets a Day*, 30 June 2011, <http://blog.twitter.com/2011/06/200-million-tweets-per-day.html> (accessed 17 April 2012).

[9]YouTube, *Statistics*, <www.youtube.com/t/press_statistics> (accessed 17 April 2012).

[10]A. Signorini, A.M. Segre and P.M. Polgreen, 'The Use of Twitter to Track Levels of Disease Activity and Public Concern in the US During the Influenza A H1N1 Pandemic', *PLoS ONE* 6/5 (2011) pp.1–10.

[11]J. Hoffman, 'Trying to Find a Cry of Desperation Amid the Facebook Drama', *New York Times*, 23 February 2012, <www.nytimes.com/2012/02/24/us/facebook-posts-can-offer-clues-of-depression.html?_r=3> (accessed 17 April 2012); 'T.J. Lane Facebook Photos: Suspect Faces Charges in Chardon High School Shooting (Slideshow)', *Huffington Post*, 28 February 2012, <www.huffingtonpost.com/2012/02/28/tj-lane-facebook-photos_n_1307836.html#s736080&title=TJ_Lane_Facebook> (accessed 17 April 2012).

[12]For instance the UN Global Pulse Programme, *UN Unveils Initial Findings on Uses of Real-time Data for Development Work* (UN News Centre, 8 December 2011), <www.un.org/apps/news/story.asp?NewsID=40667&Cr=global&Cr1=pulse> (accessed 17 April 2012).

[13]Criminal Justice Degrees Guide, *20 Infamous Crimes Committed and Solved on Facebook*, <http://mashable.com/2012/03/01/facebook-crimes/> (accessed 1 July 2012).

Facebook to plan and coordinate with each other.[14] The Daily Mail reported that 12,300 alleged offences were linked to Facebook in 2011.[15]

When society develops and adopts new methods of communication and organization – such as social media – public institutions, including the police and intelligence services, have a responsibility to react and adapt. The explosion of social media is the latest in a long line of disruptive technological innovations, and now requires a response from the authorities in turn.

The Opportunity of SOCMINT

Measuring and understanding the visage of millions of people digitally arguing, talking, joking, condemning and applauding is of wide and tremendous value to many fields, interests and industries. A family of 'big data' approaches already exists to make sense of social media. Known as social media analytics, these tools constitute a broad church, ranging from advertisers listening to social media 'buzz' to track attitudes surrounding their brands and companies monitoring their social media reputation, to mapping 'social graphs' of relationships between people, to drawing 'wisdom of the crowd' solutions to emergency situations, to conducting linguistic analysis of forum posts and network analysis of Twitter users. Fledgling academic efforts have used social media to inform investments into hedge funds.[16]

Looking at the current SOCMINT technologies now on the horizon – as well as the threats we now face – the following capabilities could for example contribute in the future to public security:

- *Crowd-sourced information.* This could help ensure a better flow of information between citizens and the government, especially in times of emergency.[17] With access to social media, passive bystanders can become active citizen journalists, providing and relaying information from the ground. The report by Her Majesty's Inspectorate of Constabulary into the riots notes, for example, a messaging service on West Midlands

[14]Diego Laje, '#Pirate? Tracking Modern Buccaneers Through Twitter', *CNN*, 15 March 2012, <http://edition.cnn.com/2012/03/15/business/somalia-piracy-twitter/index.html> (accessed 1 June 2012).

[15]Jack Doyle, 'A Facebook Crime Every 40 Minutes', *Daily Mail*, 4 June 2012, <http://www.dailymail.co.uk/news/article-2154624/A-Facebook-crime-40-minutes-12-300-cases-linked-site.html> (accessed 1 June 2012).

[16]T.O. Sprenger and I.M. Welpe, *Tweets and Trades: The Information Content of Stock Microblogs*, 1 November 2010, <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1702854> (accessed 17 April 2012).

[17]Twitter was used by pupils as an *ad hoc* emergency broadcasting system during the Ohio school shooting. See L. Dugan, 'Twitter Used as an Impromptu Broadcast System During Ohio School Shooting', *Media Bistro*, 28 February 2012, <www.mediabistro.com/alltwitter/twitter-used-as-impromptu-emergency-broadcast-system-during-ohio-school-shooting_b19030> (accessed 1 June 2012).

Police's website, which allowed citizens to post messages and questions, allowing the police to build up a picture of the situation on the ground in real-time, as well as allowing people to identify pictures of suspects uploaded to the site.[18] Tapping into the 'wisdom of the crowds' is already of great, demonstrated value. For example, the open-source platform Ushahidi has allowed large groups of people to provide collective testimony on everything from the earthquake in Haiti to blocked roads in Washington, DC.[19] These applications, impressive as they are, are only the beginning, and the stronger the techniques to make sense of information of this kind, scale and dynamism, the more effective the responses, from providing snow ploughs to drinking water, that can be made.

- *Research and understanding*. Research based on social media could contribute to our understanding of a number of phenomena. This could include the thresholds, indicators and permissive conditions of violence; pathways into radicalization; an analysis of how ideas form and change; and investigation of the socio-technical intersections between online and offline personae. Beneath the tactical and operational level, a background of more generic and distanced understanding is important for security work. For instance, the British counter-terrorism strategy aims to reduce the threat from terrorism so that people can go about their normal lives, freely and with confidence, and it is understood that the long-term way to do this is through tackling the underlying social, ideational and political causes of terrorism.

In addition, the rise in use of social media, together with the rapid development of analytics approaches, now provides a new opportunity for law enforcement to generate operational intelligence that could help identify criminal activity, indicate early warning of outbreaks of disorder, provide information and intelligence about groups and individuals, and help understand and respond to public concerns. Some of this access will come from 'open source' information derived from Twitter and other social media content authorized for public access. Some, however, will require legal authorization to override privacy settings and encryption of communications. We can group the advantages of such operational exploitation in terms of:

- *Near real-time situational awareness*. This is the ability to collect and cluster social media and output in a way that indicates and describes unfolding events. Analysis of Twitter has shown that, while the majority of Twitter traffic only occurred after an event had been reported by a

[18]HMIC, *The Rules of Engagement*, p. 31.
[19]J. Howe, 'The Rise of Crowdsourcing', *Wired*, June 2006, <www.wired.com/wired/archive/14.06/crowds.html> (accessed 17 April 2012).

mainstream news outlet, 'bursts' of tweets indicating a significant event often pre-empt conventional reporting.[20] Social media traffic analysis could allow for a more rapid identification of emerging events than traditional reporting mechanisms. With the application of geo-location techniques this could lead, for example, to a constantly evolving map showing spikes in possible violence-related tweets, facilitating a faster, more effective, and more agile emergency response.

- *Insight into groups*. This would include the ability to better understand activities and behaviour of certain groups already of interest to police or intelligence agencies. Given the appropriate legal authorization, the police could use SOCMINT to spot new, rapidly emerging 'hot topics' that spring up within group-specific conversations and how the group reacts to a specific, perhaps volatile, event. Through these and other techniques, SOCMINT might indicate the overall levels of anger within a group, and their key concerns and themes that animate intra-group discussions. At a higher level of specificity, information can also be identified and extracted regarding when a group is planning demonstrations or flashmobs, which could lead to violence or increasing community tensions; football fans planning 'meets', which could cause major economic disruption; groups planning counter-demonstrations, which could change the kind of policing required to maintain public order.
- *Identification of criminal intent or criminal elements in the course of an enquiry both for the prevention and prosecution of crime*. Similarly, law enforcement could use the warranted surveillance of social media use by individuals suspected of involvement in a crime or criminal conspiracy, the cross referencing of such individuals' accounts, the identification of accomplices, the uncovering of assumed identities, the identification of criminal networks that operate through social media sites, and the provision of social media content suspected of being evidence of a crime to the Crown Prosecution Service.

Such potential suggests that SOCMINT will merit a significant place in the national intelligence framework. However, whenever a new form of technology emerges, it takes some time before legitimate and rigorous systems of capture, analysis and interpretation are developed. There are a number of key challenges that need to be addressed before SOCMINT can be fully exploited in the interest of national and public security. It is to these challenges, and their suggested solutions, that we turn next.

The Challenges of SOCMINT: Necessity and Legitimacy

The full promise of SOCMINT as a law enforcement tool in addition to its use as an open source of information must be tempered against the reality

[20]'Reading the Riots: Investigating England's Summer of Disorder' [interactive], *Guardian*, <www.guardian.co.uk/uk/interactive/2011/aug/24/riots-twitter-traffic-interactive> (accessed 17 April 2012).

that the methods employed to protect society rest ultimately on some form of public acceptability and involvement. Britain's National Security Strategy recognizes that security and intelligence work in general is predicated not only on the public's consent and understanding, but also on the active partnership and participation of people and communities. Serious and recognized damage to security occurs when the state's efforts are not accepted or trusted.[21]

Public acceptability can be secured and maintained through two important public demonstrations. First, that the collection of intelligence is able to make an effective and necessary contribution toward safety and security; second, that this contribution is being proportionately and appropriately balanced against other desirable public goods – such as the right to private life. In sum, intelligence activity must effectively contribute to a public good but not detract from or threaten any others in ways that are not recognized and appropriately managed. These are the challenges of necessity and legitimacy.

Necessity

The first demonstration that the use of SOCMINT must make is that it works. If it did not have a reasonable prospect of contributing towards public safety there would be no moral, or indeed financial, argument for it to be collected or used. If SOCMINT is not efficacious, it risks harm to suspects who turn out to be innocent, risks of collateral damage to others to whom a duty of care is owed, and the risk of confounding otherwise sound intelligence efforts.

The 'success' of intelligence is not the information or even secrets that it collects, but the value it adds to decision-making. Indeed, the justification for creating SOCMINT (or any intelligence) capabilities and applying them, with all the recognized hazards this entails, is that it contributes to the public good of safety and security. It is therefore morally imperative that SOCMINT operations present a reasonable chance that they will yield actionable, useable intelligence that contribute to consequential decisions, such as deploying emergency services to the right place at the right time.

For information to be considered successful 'intelligence' it needs to meet certain thresholds of how it is gathered, evidenced, corroborated, verified, understood and applied. Different sources and kinds of information have developed signature ways of meeting this challenge. For example, open-source intelligence (OSINT) triangulates reliable sources; human intelligence (HUMINT) might consider the track record of the agent; imagery intelligence (IMINT) needs to pay attention to the technical characteristics of the collection platform; and signals intelligence (SIGINT) would need to understand the context of the language used. All source intelligence

[21]Cabinet Office, *A Strong Britain in an Age of Uncertainty: The National Security Strategy* London: HMSO 2010) p.5.

assessments try to get an overall picture on the basis of these different types and reliability of contribution.

SOCMINT has its own characteristics that can be seen by examining how SOCMINT fits into the traditional intelligence cycle, in particular the functional steps of collection, processing and analysis, and dissemination.

Data access. One of the difficulties of SOCMINT which it shares with much of modern SIGINT is not a paucity of data, often the key problem in the collection of Cold War secret intelligence, but a deluge. The term 'access' is preferred over 'collection' to indicate that in the internet we are dealing with a very different process from that of traditional intelligence gathering. During the week of the August 2011 riots, for example, millions of riot-related tweets were sent, consisting of news stories, rumours, reactions and indications of criminal intent.[22] Knowing what data ought to have been accessed and analyzed – sorting the wheat from the chaff – is the critical consideration. Selection and filtering tools (including such techniques as semantic search) are available and have, for example, been extensively used in signals intelligence and in email searches and legal discovery and will be needed for social media analysis by law enforcement of their targets.

Rather different problems arise for the analyst when trying to distil general meaning from large open data sets. One useful way to approach this challenge is to draw on how traditional quantitative disciplines deal with overwhelmingly large data sets. Most turn to the statistical technique of sampling, wherein a manageable amount of data is collected that represent the unmanageably large 'population' being researched. The reliability and validity of inferences or extrapolations made on this basis depend on the quality, especially representativeness, of the sample collected. Over the past century, statisticians have developed techniques that allow small data sets to be representative, and therefore permit more general conclusions to be drawn – particularly through the use of randomized sampling. Simply put, however, inferences and conclusions can only be reasonably drawn if one knows how the sample was constructed and the data collected, and what this means about the inferences that are then made.

The broad problem is that social sciences have not developed an approach to robustly sample social media data sets. Only a very limited amount of work has been done to develop different types of sampling for automated systems of data collection.[23] More attention has been paid to methodologies that produce a large sample (something that computational approaches are

[22]R. Proctor, F. Vis and A. Voss, 'Riot Rumours: How Misinformation Spread on Twitter During a Time of Crisis', *Guardian*, 7 December 2011, <www.guardian.co.uk/uk/interactive/2011/dec/07/london-riots-twitter> (accessed 17 April 2012).

[23]See for instance J. Leskovec, J. Kleinberg and C. Faloutsos, 'Graph Evolution: Densification and Shrinking Diameters', *ACM Transactions on Knowledge Discovery from Data* 1/1 (2007) <www.cs.cmu.edu/~jure/pubs/powergrowth-tkdd.pdf> (accessed 16 April 2012); J. Leskovec and C. Faloutsos, 'Sampling from Large Graphs' in T. Ellassi-Rad (chair), *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery*

good at delivering), rather than methodologies that produce a representative one (something social sciences are good at delivering). Moreover, the emerging, highly technical and computer science-driven developments in social media sampling practices, including 'forest-fire' (wherein links and nodes 'burn' outward from a random seed to create the sample), user-activity and location-attribute-based techniques, have had very little uptake within the social science community.

Very little research based on social media data sets acknowledges the sampling frame applied, and how this might limit or bias the results that are drawn. Typical data acquisition strategies within the small but growing field of academic work of remain 'samples of convenience' or 'incidental sampling', which means the most readily available or easily accessible – rather than the most representative – are collected.[24] For obvious reasons this type of sampling limits the strength of conclusions drawn. One prominent example is the recent *Reading the Riots* collaboration involving the Guardian, a number of British universities, freelance journalists and community researchers. The project gathered 2.6 million tweets about the August 2011 riots and drew a number of conclusions, including that there were very few examples of Twitter being used to express or encourage criminal activity. However, the data set of tweets was collected using around 150 'hashtag clouds', which means only tweets that included an identifying hashtag, such as #londonriots, were collected and analyzed. It is possible, however, that people who use a hashtag when tweeting are not representative of all tweets about the riots; for example, they might be less likely to talk about criminal activity because hashtags are usually employed by users to disseminate tweets to a wider audience. In statistics, this is known as 'missing at non-random data', which means certain data might be systemically absent as a result of the sampling method. This is considered a serious problem when drawing conclusions, because when people are absent from a data set for a reason other than chance, they share a related and likely important trait (or traits) that could have a substantial impact on the research findings.

Unlike in traditional social science research, technical considerations might also affect the quality of sample. For example, in the case of Twitter, the publicly available application programme interface is limited to 150 requests per hour, going up to 20,000 when it is 'whitelisted'. While this can capture an enormous data set by traditional social science standards, it can only capture a small amount, and not an automatically representative sample, of the total number of tweets. Researchers can gain access to larger proportions of the tweet-feed. The 'firehose' gives access to all tweets, the

and Data Mining, *2006* (Philadelphia: KDD 2006), <www.stat.cmu.edu/~fienberg/Stat36-835/Leskovec-sampling-kdd06.pdf> (accessed 16 April 2012).

[24]See, for instance, B. O'Connor, R. Balasubramanyan, B.R. Routledge and N.A. Smith, 'From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series', *Proceedings of the AAAI Conference on Weblogs and Social Media* (Washington, DC: AIII Press 2010). The authors collected their sample using just a few keyword searches.

'gardenhose' gives 10 per cent, and the 'spritzer' gives 1 per cent. Unfortunately, precisely how these different access levels affect the quality of the data, and what sorts of systemic bias they might hide, are not fully known – and very rarely stated.

For the interests of public policy, including security strategies, what are especially required are data acquisition plans that allow online phenomena to be related to offline behaviour. Demographic representativeness is key for this. The people who use social media such as Twitter and Facebook tend to be younger, richer, more educated and more urban than the population in general.[25] Additionally, when looking at demography, it is not only the general population that is important, but also the community that accounts for the information that is gathered. Online social content is subject to the enduring influence of the Pareto principle of the 'vital few' that 80 per cent of the user-generated content for any given site will tend to come from a highly productive 20 per cent of users.[26] A 2010 study of Twitter found this to be broadly true: 22.5 per cent of users accounted for around 90 per cent of all activity.[27]

Processing and analysis. As described in the previous section, drawing meaning from open social media data presents great challenges for the intelligence analyst. Many of the technologies that have been developed in the private sector by the advertising and public relations sectors have been moulded by the metrics traditions and to suit the needs of these industries. They aim to gain a general understanding of attitudes toward a product or whether a new advertising campaign is creating a 'buzz'. Security and intelligence efforts, however, demand modes of analysis that can deliver levels of confidence that these technologies cannot, yet, deliver.

Because social media data sets are so large, a number of broadly computational approaches have been developed to infer and extract 'meaning' automatically, without the routine presence of a human analyst. The most important approach is a variant of artificial intelligence – 'machine learning' – where algorithms are taught to recognize patterns and therefore meaning within pieces of information that human beings need therefore never see. Machine learning has a number of important applications, from identifying clusters and anomalies in large data sets to the extraction of semantic information from text. A particularly important application is 'sentiment analysis', where an algorithm looks for certain qualities and properties in a piece of text that it considers to correlate statistically with an emotion or 'sentiment'. Once human input has defined what sentiments are

[25]For information Twitter and Facebook demographics see, 'Infographic: Facebook vs. Twitter Demographics', *Digital Buzz Blog*, 21 December 2010, <www.digitalbuzzblog.com/infographic-facebook-vs-twitter-demographics-2010-2011/> (accessed 16 April 2012).

[26]See C. Shirky, *Here Comes Everybody: The Power of Organizing Without Organizations* (New York: Penguin 2008).

[27]'Twitter Statistics for 2010', *Sysomos*, December 2010, <www.sysomos.com/insidetwitter/twitter-stats-2010/> (accessed 16 April 2012).

being searched for, and what textual examples of these sentiments are, the algorithm is able, with varying degrees of specificity and accuracy, to classify enormous volumes of data automatically on the same basis. Sentiment analysis has been applied for a number of aims, from measuring Twitter users' feelings towards political parties to predicting the future of box office revenues.[28] In the future we may see sentiment analysis used by the police to gauge the mood of demonstrators and the possibility of criminal violence erupting.

The ability to extract automatic meaning from unstructured data such as tweets opens many research opportunities, and social researchers can now contemplate handling bodies of information, and compiling sample sets, on a previously unmanageable scale. However, a crucial consequence of the rise of machine learning approaches within social media analytics is that we are currently much better at counting examples of online human behaviour than critically explaining why they are and what it might mean.

To make this sort of sense of any form of communication, context is critical. A central tenet of all semiotics and linguistics is that language is textured: the intent, motivation, social signification, denotation and connotation of any utterance is mutable and dependent on the context of situation and culture. The accuracy of any interpretation depends on a very detailed understanding of the group or context that is being studied. For example, most groups of people use vernacular and group-specific language that a generic or standardized sentiment lexicon or thesaurus would often misinterpret.

However, because automatic data collection is required to process the sheer volume of data now available, many of the contextual cues – the thread of a conversation, information about the speaker, the tone of the utterance and the information about the speaker – are often lacking in analysis of social media data. Therefore utterances have to be abstracted out of the wider situational, contextual and cultural picture – what we would call their 'naturalistic setting'. The act of 'scraping' a social media platform – such as collecting tweets or Facebook posts – like filtered selection in signals intelligence usually does not by itself provide the utterance's position in a social network (such as whether they were talking to their friend) or a conversational network (such as whether the utterance was a heated rebuttal in an argument). Again, these are issues that have been faced by signals intelligence agencies in the internet age and where analytic experience and judgment have been shown to be key.

Context is also shaped by the norms and mores of the medium we are using. A number of studies are beginning to identify norms, rules and behaviours that dictate communication via social media that differ in

[28] J. Weng, Y. Yao, E. Leonardi and F. Lee, 'Event Detection in Twitter', *HP Laboratories*, 6 July 2011, <www.hpl.hp.com/techreports/2011/HPL-2011-98.html> (accessed 17 April 2012); S. Asur and B.A. Huberman, 'Predicting the Future With Social Media', *HP Laboratories*, 29 March 2010, <www.hpl.hp.com/research/scl/papers/socialmedia/socialmedia.pdf> (accessed 17 April 2012).

significant ways to how people might communicate offline. Some studies for example argue for an 'online disinhibition effect' – that the invisible and anonymous qualities of online interaction lead to disinhibited, more intensive, self-disclosing and aggressive uses of language.[29] Identification with groups or movements has also changed. Traditional forms of membership to a group or a movement are relatively intense, often involving subscription fees and membership lists. For many online groups, however, a single click of a mouse is sufficient to express some form of affiliation. This is a more ephemeral, looser and possibly less involved form of affiliation. Indeed, a recent study of 1300 Facebook fans of the English Defence League found that only three-quarters considered themselves 'members' of the group, and only one-quarter of those had ever actually been on a march.[30]

Taken together, these phenomena constitute the rapid emergence of distinct social media sub-cultures, which are developing new understandings, social mores and uses of language in clearly distinct ways.[31] Indeed, a new branch of sociology – digital sociology – is devoted to understanding these socio-cultural consequences of the internet and the new ways it is being used.

When context is not considered, there can be profound consequences and potential for misinterpretation. In 2010, Paul Chambers declared to his 650 Twitter followers his intention of 'blowing [Robin Hood] airport sky high!!'[32] Undoubtedly in jest, his initial conviction for the 'menacing use of a public communication system' under the Communications Act 2003 has attracted wide public criticism and was overturned on appeal. Jonathan Bennett, the district judge, noted the 'huge security concerns' within the context of the times in which we live, but perhaps not the Twitter-specific situational and cultural context of the utterance.[33] In a similar case, Leigh Van Bryan and Emily Bunting were denied entry to America after tweeting 'free this week for a quick gossip/prep before I go and destroy America? x'.[34]

Although there are no simple solutions to these difficulties, some steps forward are possible. First, big data computational tools must become more 'human-sized' – sympathetic to the human subject they wish to measure.

[29]J. Suler, 'The Online Disinhibition Effect', *Journal of Cyberpsychology and Behaviour* 7/3 (2004) pp.321–6. See also J. Suler, *The Psychology of Cyberspace: The Online Disinhibition Effect*, <http://users.rider.edu/~suler/psycyber/disinhibit.html> (accessed 17 April 2012).

[30]J. Bartlett and M. Littler, *Inside the EDL* (London: Demos 2011).

[31]'Twitterology High and Low', *The Economist*, 31 October 2011, <www.economist.com/blogs/johnson/2011/10/technology-and-language?fsrc=scn%2Ftw%2Fte%2Fbl%2Ftwitterologyhighandlow> (accessed 16 April 2012).

[32]'Robin Hood Airport Tweet Bomb Joke Man Wins Case', *BBC News*, 27 July 2012, <www.bbc.co.uk/news/uk-england-19009344> (accessed 16 April 2012).

[33]'Jack of Kent' (David Allen Green), *Paul Chambers: A Disgraceful and Illiberal Judgment*, 11 May 2010, <http://jackofkent.blogspot.com/2010/05/paul-chambers-disgraceful-and-illiberal.html> (accessed 16 April 2012).

[34]A. Parker, 'US Bars Friends over Twitter Joke', *Sun*, 30 January 2012, <www.thesun.co.uk/sol/homepage/news/4095372/Twitter-news-US-bars-friends-over-Twitter-joke.html> (accessed 16 April 2012).

Statistical analytical methods, such as sentiment analysis, must involve analysts and experts who understand the norms and behaviours of the groups involved. Second, any analysis of social media data sets should always be based on an understanding of the medium itself: the specific online culture, language and behaviour. Project Raynard is a good and relatively recent of an intelligence programme that stresses the importance of first establishing norms in online environments before looking for deviations from the norm.[35] Most important is that any organization using SOCMINT must recognize the analytical and interpretative limitations of the field, how they reflect on the kind of insight that can be drawn, and the kind of decisions that can be made on the basis of those limitations.

Dissemination. The effective use of intelligence from internet sources including social media data sets also depends on it getting to the right people quickly, securely and presented in a format that makes sense to strategic and operational decision-makers. Depending on the purpose of the SOCMINT, this may range from a footnoted, caveated and in-depth strategic analysis paper, to the operational use of single-screen, real-time visualizations of data available on portable devices.[36]

Several challenges will need to be addressed. First, SOCMINT dissemination must reflect the general difficulties in using SOCMINT: its complexity, scale, dynamism and – given the problems outlined above relating to both access and interpretation – any presentation of data needs to be presented with new procedures and caveats.

Second, SOCMINT dissemination must slot into existing intelligence channels – police, emergency service response, the Security Service, the Joint Terrorism Analysis Centre, Cabinet Office Assessments Staff and so on. However, this requires specific training for gold, silver and bronze commanders and additional training for frontline officers who could benefit from the daily use of such intelligence but who are not today regarded as direct intelligence customers.

Third, SOCMINT dissemination and retention must comply with the highest standards of information assurance. Existing controls must be applied to ensure the safekeeping of SOCMINT data accessed under warrant

[35]Intelligence Advanced Research Projects Agency, *Broad Agency Announcement: Reynard Program*, 16 June 2009 <http://www.iarpa.gov/solicitations_reynard.html> (accessed 12 June 2012).

[36]For instance, in deprived areas of Berlin, civil servants have increasingly used portable devices connected to database records when visiting care homes for the elderly and hospitals. These devices give constant, mobile access to databases, enabling public servants to understand the needs of individuals and families, track their previous contact and check for problems and underlying issues that may have been recorded by other agencies. See J. Millard, 'eGovernance and eParticipation: Lessons from Europe in Promoting Inclusion and Empowerment', paper presented to UN Division for Public Administration and Development Management (DPADM) workshop 'e-Participation and e-Government' (Budapest, Hungary, 27–28 July 2006), <unpan1.un.org/intradoc/groups/public/documents/UN/UNPAN023685.pdf> (accessed 23 January 2012).

and regulating their dissemination, including overseas. The unregulated dissemination of SOCMINT data would risk jeopardizing public confidence in this form of intelligence. More generally, whether lost, insecurely held, or subject to hostile access, as government increases the amount of personal information it holds, the potential for various data compromises, and the harm it might cause to public confidence, will inevitably grow.

Fourth, the effective application of data visualization techniques will be required to render complex and often interlinked intelligence in a way that is intuitively comprehensible, but conserves the networked nature of the information. More experience of using such SOCMINT data analysis techniques by law enforcement in particular is needed in order to draw up detailed rules and regulations for its safe management.

Validation and use. The way SOCMINT can add value relates to how the operators – such as frontline police officers – will actually use the information, and how they ought to interpret and act on it (such as deploying reserve forces in the build up to a march). In addition to the many methodological hurdles that stand in the way of the responsible interpretation of data, the social media being monitored is itself prone to contain misleading information of a polemical nature, which may involve the recirculation of selective half-truths, mistakes and outright distortions. Validation of SOCMINT intelligence is therefore an important function for the social media analyst.

One risk that must be accounted for when considering validating SOCMINT data is the risk of engineering the 'observation effect': the tendency of individuals to change their behaviour if they believe they are being observed. In 2009, the LSE's Policy Engagement Network warned of this 'effect' in a briefing paper responding to the then-government's Interception Modernisation Programme. The report feared that when the public became aware that communications data were being collected and collated, there would be a risk that 'it will generate a chilling effect on the individual's right to free expression, association and might dissuade people from participating in communications transactions'.[37] On the other hand, previous predictions such as the decline of communications intelligence due to the ready availability of hard encryption have not proved correct, and similarly it is unlikely that changes in public behaviour as a result of knowledge of social media monitoring will significantly limit the effectiveness of social media and online communications data as sources of intelligence.

Related to this issue is the problem of 'gaming' – the deliberate use of this media as a means of misleading or confusing an observer, in this case the law enforcement agencies. In the context of intelligence work, this problem is not new, although experience such as the Allied deception operations in the Second World War illustrates the care with which deception operations have to be planned if they are not to backfire on the originator. The nature of

[37]Briefing on the Interception Modernisation Programme, *Policy Engagement Network Paper 5* (2009), p.56.

SOCMINT may make attempts at deception more likely given the ubiquity of social media, its widespread use and the democratization of computer coding and technical know-how. In a recent example, a leaked cache of emails allegedly belonging to Bashar al-Assad indicated that an aide, Hadeel al-Assad, posted pro-regime commentary under an assumed Facebook identity that was mistaken as genuine and given international coverage by CNN.[38]

For these reasons, there must be a thorough (yet sufficiently rapid) process to ensure that an item of SOCMINT can, as far as possible, be validated before it reaches the final user. The validation of SOCMINT is ideally done further up the 'food chain' from the functions of access and processing of data when all sources of intelligence, including open source material, can be brought to bear.

As with other intelligence sources, this validation process must take the form of a reporting framework that rates the 'confidence' in any piece of freestanding piece of SOCMINT. By pointing out potential vulnerabilities and biases in the acquisition and analysis of the information, we may gauge the importance of the information collected and caveat the conclusions that may be drawn.

We must also be able to relate SOCMINT to other kinds of evidence to produce an overall picture – the equivalent of an 'all-source assessment'. The value of SOCMINT relative to other forms of intelligence must be evaluated and the ways in which various types of intelligence can be used in tandem needs to be investigated. The crucial points here are the exact application of SOCMINT in a given circumstance and its 'strength' in relation to other forms of intelligence. To complicate the issue, both will of course vary according to the situation ranging from identifying broad societal-level trends on the one hand and the context of a riot or crowd control on the other.

A number of strategies will be useful to create processes to validate SOCMINT. More methodologically mature forms of offline research can be conducted in parallel to SOCMINT projects to allow the results to be compared. For example, it would be especially useful to establish rules about how online phenomena maps onto offline behaviour. Retrospective analysis can also be used to quantify SOCMINT accuracies and diagnose instances where accuracy was confounded. In addition to the specific validation responsibilities placed on the agency that collected the intelligence, there needs to be a very general up-skilling of all the branches of government that might be involved in this work. It will be impossible to use this medium without analysts, police officers or judges who understand its norms and mores. Ultimately, the value of SOCMINT can only really be understood through using it. Understanding will slowly emerge as SOCMINT is trialled – we must expect varying degrees of success in different contexts.

[38]'Shopping Amid a Massacre: Leaked E-mails from Syria's Regime', *CNN International*, 16 March 2012, <http://edition.cnn.com/2012/03/15/world/meast/syria-al-assad-e-mails/index.html?iphoneemail> (accessed 16 April 2012).

In general, the origin of the main risks in using information from social media arises from a lack of interaction between the humanities and the statistical and computational disciplines. Those disciplines best equipped to understand and explain human behaviour – the social and behavioural sciences, political science, psephology, anthropology and social psychology – have not kept pace in relating this insight to the big-data approaches necessary to understand social media. Conversely, these very same big-data approaches that form the backbone of current SOCMINT capabilities have not used sociology to employ the measurements and statistics they use to the task of meaningfully interpreting human behaviour.

Taken together, the methodological steps forward point towards a fundamental evolution in the capabilities available to exploit social media. If SOCMINT is to be methodologically robust enough to base decisions on and change policy, it must rest on a new, applied, academic inter-discipline: social media science. This would embody a more meaningful and intensive fusion of the computational, technological and humanities approaches. Only through this fusion can data-led explanations of human behaviour also be humanistic explanations of human behaviour. This will require new relationships with industry and academia, and concerted, long-term investment to build technological, methodological and presentational capabilities.

Legitimacy

The second important condition of SOCMINT use is that it be legitimate. Broadly speaking, all security and intelligence work rests on a delicate balance between three classes of public goods: the maintenance of national security including public order and public safety; citizens' right to the rule of law, liberty and privacy; and the overall economic and social wellbeing of the nation and its citizens. To be legitimate, any use of SOCMINT similarly has to recognize where it risks harm to a public good, and balance this against any contribution it makes to another.

In most circumstances these three classes of public goods should be mutually reinforcing: security promotes inward investment and market confidence promoting economic wellbeing and social harmony that in turn supports the maintenance of security. There are times however when choices have to be made. Within a rights-based approach, the only justification for one public good to be hazarded is the provision of another. Yet social media is a potentially disruptive phenomenon that is already affecting and in some cases redefining how these three classes of public goods can be attained, for the following reasons:

- *Fungibility and heterogeneity.* SOCMINT cuts across several categories and can be in more than one category at a time. The broad scanning of tweets has similarities to mass surveillance such as the deployment of CCTV in crowded places. The close following of an individual's Facebook page during the course of an investigation has similarities to *de visu* surveillance as 'authorizable' by a senior police officer. Accessing

encrypted BlackBerry messaging by cracking the security PIN is an interception of communications under RIPA 2000 for which a warrant would be required.

- *Generality*. When used for intrusive investigation there may be no named suspect or telephone number to target and the output may be general rather than specific to an individual (such as noting an increase in social media communications in a specific area where demonstrations are taking place).
- *Scalability*. Many of the automated techniques can be scaled up from the collection of hundreds to millions of pieces of social media data easily and cheaply. The scale may be difficult to pin down in advance.
- *Flexibility*. The scope of many 'scraping' technologies (for instance, the keywords they scan for) can be changed easily. This means they can easily be redirected away from their original mission and function, which may be justified operationally by tactical changes on the ground.
- *Invisibility*. Like other forms of covert surveillance, the operation of SMA techniques will normally not be visible to the social media users themselves and will override what they may assume are their privacy settings.
- *Broader public concerns with digital surveillance*. SOCMINT must be understood within the context of public concerns about digital surveillance driven by the broad rise in information systems that have vast capacities to capture, stockpile, retrieve, analyze, distribute, visualize and disseminate information. Concerns arise in the proliferation of surveillance opportunities in this data-rich environment, and the consequences of collateral intrusion, the possibility of data compromise, and a general implication of suspicion of wrongdoing entailed by the widespread collection of information.[39]

As with other forms of intelligence, public concerns over privacy must be managed. Privacy itself is an elusive concept. Article 8 of the European Convention of Human Rights enshrines the right to respect for 'a person's private and family life, his home and correspondence'. Respecting privacy can mean that data are kept confidentially, gathered anonymously, used in a self-determined way (the principle of 'informed consent'), and that people are able to see them and correct errors, or, of course, that no data are gathered at all.

Many broad and fundamental changes in society are nonetheless transforming what privacy means to people and social media use in particular challenges clear-cut distinctions of what is private and what is not. McKinsey Global Institute has calculated that 30 billion pieces of content are shared on Facebook each month, many of them personal.[40] This sharing

[39]For a description of many of these trends, see Information Commissioner's Office, *Information Commissioner's Report to Parliament on the State of Surveillance* (November 2010).

[40]S. Sengupta, 'Zuckerberg's Unspoken Law: Sharing and More Sharing', *New York Times*, 23 September 2011, <http://bits.blogs.nytimes.com/2011/09/23/zuckerbergs-unspoken-law-sharing-and-more-sharing/> (accessed 17 April 2012).

of such a large amount of voluntarily uploaded personal data, and the number of people and institutions to whom these data are accessible, is unprecedented; depending on the user-selected privacy settings employed, personal information added to Facebook can be viewed by all of Facebook's 845 million other users. Far from being incidental, this move towards the widespread dissemination of personal information is fundamental to the ethos of social networking sites. Facebook's privacy settings inform users that the ability to share information 'allows us to provide Facebook as it exists today', while Twitter states more explicitly that 'most of the information you provide to us is information you are asking us to make public'.[41] Indeed as a result of these changing behaviours, Mark Zuckerberg, Facebook's CEO, declared that privacy is 'no longer a social norm'.[42] Most of us accept that both private and public bodies – from Tesco through its Clubcards to Amazon, Oyster and Google – learn and record a vast amount about us daily. In a Eurobarometer poll, a bare majority of UK respondents considered photos of themselves to be personal data, less than half considered 'who your friends are' to be personal data, 41 per cent thought that details of the websites they visit were personal data, and only 32 per cent thought their tastes and opinions were personal data, yet in contrast, large majorities regard financial data as personal.[43] However, although research suggests that users recognize disclosing personal information is an increasingly important part of modern life, the majority have concerns about what this means.[44] In a 2008 European Commission Poll, around 80 per cent of people agreed that 'people's awareness about personal data protection in the UK is low'.[45]

Attitudes towards privacy – especially broad, generic and in-principle attitudes – are notoriously hard to measure objectively. Broad behavioural norms, such as the amount of information we now share, suggest the concept is in a state of flux, where its boundaries of definition are being fundamentally redrawn. The debate will continue to rage about where these redrawn boundaries on the possession, sharing and use of personal information, now lie – indeed what privacy is.[46]

---

[41]Twitter, *Privacy Policy*, <http://twitter.com/privacy/previous/version_2> (accessed 17 April 2012); Facebook, *Data Use Policy*, <www.facebook.com/about/privacy/your-info> (accessed 17 April 2012).

[42]B. Johnson, 'Privacy No Longer a Social Norm, Says Facebook Founder', *Guardian*, 11 January 2011, <www.guardian.co.uk/technology/2010/jan/11/facebook-privacy> (accessed 17 April 2012).

[43]*Attitudes on Data Protection and Electronic Identity in the European Union*: *Special Eurobarometer 359* (Brussels: European Commission 2010).

[44]See D. Boyd and E. Hargittai, 'Facebook Privacy Settings: Who Cares?', *First Monday* 15/8 (2010).

[45]European Commission, *Data Protection in the European Union: Citizens' Perceptions* (2008), <http://ec.europa.eu/public_opinion/flash/fl_225_en.pdf> (accessed 17 April 2012).

[46]For recent deliberative research into people's conceptions of privacy, see P. Bradwell, *Private Lives: A People's Inquiry into Personal Information* (London: Demos 2010), <www.demos.co.uk/files/Private_Lives_-_web.pdf> (accessed 17 April 2012).

The crucial implication for the use by government of different types of SOCMINT is that the framework for recognizing and managing incursions into privacy is struggling to keep pace with changing social attitudes and habit. There are many ways the state can collect and use information about people, while different systems exist for carrying this out. Each system identifies and limits potential harm from accessing private information. When the state conducts clinical research, for example, consent is requested and anonymity often guaranteed; when the state draws on the research of others we apply fair usage and ascribe credit to the individuals concerned. When the state obtains biometric information such as a DNA sample from a suspect, consent is not required but restrictions are applied to the retention of material. When the state carries out surveillance, the activity is usually covert and individual consent is irrelevant. Instead, to deal with concerns of privacy and intrusiveness, society's consent is needed, expressed through enabling legislation. Whilst enabling legislation differs between states, they are predicated on an association between the level of harm on the one hand, and a series of steps to mitigate harm (including ways of establishing authorization, accountability, and necessity) on the other. With the concept of privacy now so mutable, the calculation of this particular kind of moral hazard is difficult.

Economic and social wellbeing. The internet as a free and open space – of course within reasonable limits – provides an immense economic and social benefit. Government activity is intended to protect prosperity, not undermine it. As the British Foreign Secretary William Hague commented in 2011 'nothing would be more fatal or self-defeating than the heavy hand of state control on the internet, which only thrives because of the talent of individuals and of industry within an open market for ideas and innovation'.[47] On the other hand, the risk must be recognized that the unregulated large-scale collection and analysis of social media data by business and government alike (even if open source) risks undermining confidence in, and therefore the value of, this space. The idea that the economic and social benefit of the internet is premised on its openness and freedom of government control is not new. From the early 1990s, a powerful argument and vision has existed about what the internet is for and what it should be like: an opportunity to evolve past the nation-state system into post-territorial, self-governing communities who operate under their own floating social contracts of consent-through-use. John Perry Barlow's famous Declaration of Cyberspace Independence declared to the 'weary giants of flesh and steel' that cyberspace was developing its own sovereignty and 'you are not welcome among us'.[48]

[47]M. Hick, 'Hague: Governments Must Not Censor Internet', *Huffington Post*, 1 November 2011, <www.huffingtonpost.co.uk/2011/11/01/william-hague-government-internet-censorship_n_1069298.html> (accessed 17 April 2012).

[48]J.P. Barlow, *A Declaration of the Independence of Cyberspace*, 8 February 1996, <https://projects.eff.org/~barlow/Declaration-Final.html> (accessed 17 April 2012).

We believe that it is important to distinguish, as we have tried to do in this article, between open-source, non-intrusive SOCMINT and closed-source, intrusive SOCMINT. Any legitimizing and enabling framework for the collection and use of SOCMINT must therefore begin by distinguishing between what is a form of intrusive surveillance and what is not. This recognizes that there are times when we can legitimately seek control of what information we give away and how it is used, but there are also times when individual control must be over-ridden. The circumstances where this can happen are based on collective decisions and assent about the state's authority.

The key concept relevant to making this distinction is whether the user controls the use of their data through consent. For SOCMINT to be non-intrusive and open-source, it should not be able to identify individuals, be used as a means of criminal investigation, or puncture the privacy wishes of the user.

Any Government use of open SOCMINT should be put on the same footing as private companies and academia, with conditions relating to anonymity and data protection. For open SOCMINT, then, harm is conceived not as intrusion into someone's private space, nor the wider issues of trust and implied suspicion (since neither of these would happen within open SOCMINT), but by the loss of control over the information through use beyond what can be reasonably expected. Reasonable expectation can be protected through a characteristic openness of how or when this kind of SOCMINT is conducted. All such collection, retention, and sharing policies are publicized and justified and the reason why the information is collected is publicized.

Intrusive SOCMINT is more straightforward since most states already have a legislative framework for regulating intrusive intelligence gathering, for example for the purposes of national security and the prevention and detection of crime (in the case of the UK such authority is provided in the Regulation of Investigative Powers Act 2000). The application of such legislation can be codified for intrusive SOCMINT to cover the powers of access required and the procedures to be followed at a legal and operational level. What is important is that there is public acceptance of the arrangements based on recognition that at the heart of an enduring, effective settlement for state surveillance rest sound ethical principles. We suggest an adaptation of the set of principles earlier suggested for the intelligence community by Sir David Omand in his book, *Securing the State*.[49]

Principle 1: There must be sufficient, sustainable cause. This first and overarching principle forces the big picture to be taken into account: the overall purposes that could justify the acquisition by a public body of capabilities to gather, understand and use social media data. There is a danger that a series of SOCMINT measures – each in themselves justifiable – together creep towards an undesirable end point: a publicly unacceptable

[49]D. Omand, *Securing the State* (London: Hurst & Co 2010).

level of overall surveillance; the possession of a dangerous capability; and the overall damage to a medium that is of obvious intrinsic value beyond security. Just because it can be done does not mean that it should be done.

Application of the principle of requiring sufficient, sustainable cause is therefore necessary to ensure that SOCMINT remains within the boundaries required to deliver social, economic, security and law enforcement benefits to the public, resisting bureaucratic empire building, finding ways to employ spare capacity or simply the *banalization* of the technology available from commercial suppliers.

Principle 2: There must be integrity of motive. This principle refers to the need for integrity throughout the whole 'intelligence' system, from the statement of justification for access, accessing the information itself, through to the objective analysis, assessment and honest presentation of the results. The justification for seeking SOCMINT in individual cases must in particular be clear and not mask other motives on the part of the investigating officers. Intelligence is by its nature usually incomplete and fragmentary, and can be wrong or subject to deception. In presenting intelligence to end-users the limitations and caveats must be made clear. Nor must the decision to use (or not to use) SOCMINT, or the conclusions drawn from it, be influenced by local or national political considerations or media pressure.

Principle 3: The methods used must be proportionate and necessary. There is a well-established principle in law enforcement that the extent of harm likely to arise from any specific action being taken should be proportionate to the harm that it is being sought to prevent. In assessing proportionality, the extent of intrusion has to be assessed. That would mean a lower threshold for covertly monitoring material that the user has not restricted than in cases where they had only a limited list of friends who had access, or where they used a system such as BlackBerry that required a PIN.

Principle 4: There must be right authority, validated by external oversight. There is a general principle that there must be an audit trail for the authorization of actions that may carry moral hazard with an unambiguously accountable authority within a clear chain of command. Having an adequate paper trail (or its electronic equivalent) for key decisions is essential for confidence of staff and politicians, and for the operation of investigations and redress in any cases of suspected abuse of powers. We believe this principle should apply to SOCMINT as well as to any other intelligence operation. This is an important way in which proportionality and accountability is realized in practice.

Principle 5: Recourse to secret intelligence must be a last resort if more open sources can be used. Because of the moral hazards of all intrusive secret intelligence gathering methods, those authorizing such operations should ask whether the information could reasonably be expected to be obtained

through other means, ranging from fully open sources to information freely volunteered from within the community. Applying this general principle to SOCMINT, less intrusive forms should be preferred to more intrusive covert forms. SOCMINT should be based wherever possible on the clearest possible mandate of informed consent. The most preferred route is to access explicitly 'consenting' information from the online community, for example crowd-sourced information that has been explicitly volunteered on a particular Facebook wall or hashtag. Recourse to covert intelligence gathering, including via exploitation of social media, should be confined to cases where the information is necessary for the lawful purpose of the operation and cannot reasonably be expected to be gained by other means.

Overall, these principles maintain a series of crucial associations: as the type of surveillance becomes increasingly intrusive so three vital and increasingly narrow conditions are imposed onto it: the agencies that can conduct it, who must authorize it, and the reasons why the surveillance can be legitimately conducted. This is vital to balance the possible good of the collection and use of SOCMINT with the possible harm. Measuring intrusion, however, is far from straightforward. People often share what would usually be considered private things about their lives in ways that are unprecedentedly public, sometimes with unexpected or unrealized consequences. The scale of intrusion entailed by SOCMINT access varies greatly. To gather and analyze a suspect's open tweets looks similar to the surveillance of someone in public, whilst to gather and analyze someone's Facebook messages seems closer to reading their private correspondence. We do not yet have a conceptual framework of what should in future constitute privacy in social media use and the consequent sorts of harms associated with breaching that privacy.

Concluding Remarks

The opportunities that the explosion of social media use offers are remarkable. SOCMINT must become a full member of the intelligence and law enforcement family. At the heart of this process are the twin demonstrations of necessity and legitimacy.

To meet the challenge of necessity, a new, applied academic discipline – social media science – must be developed. This requires new relationships with industry and academia, and concerted, long-term investment to build technological, methodological and presentational capabilities. Those disciplines best equipped to understand and explain human behaviour – the social and behavioural sciences, political science, psephology, anthropology and social psychology – must be made to interweave with the big data approaches necessary to understand social media. Only through this fusion can data-led explanations of human behaviour also be humanistic explanations of human behaviour.

But technology and capability is only half the picture. To meet the challenge of legitimacy, the public must broadly understand and accept why, when and with what restrictions SOCMINT is undertaken. Any

Government that wishes to conduct SOCMINT must adopt an explicitly articulated approach grounded in respect for human rights and the associated principles of accountability, proportionality and necessity.

Notes on Contributors

Sir David Omand GCB is a visiting professor in the War Studies Department at King's College London. He was appointed in 2002 the first UK Security and Intelligence Coordinator, responsible to the Prime Minister for the professional health of the intelligence community, national counter-terrorism strategy and 'homeland security'. He served for seven years on the Joint Intelligence Committee. He was Permanent Secretary of the Home Office from 1997 to 2000, and before that Director of GCHQ. Previously, in the Ministry of Defence he served as Deputy Under Secretary of State for Policy, Principal Private Secretary to the Defence Secretary, and served for three years in NATO Brussels as the UK Defence Counsellor. He was educated at the Glasgow Academy and Corpus Christi College, Cambridge where he is an honorary fellow. His book, *Securing the State*, was published by C. Hurst in hardback (2010) and paperback (2012).

Carl Miller is an associate at Demos and a researcher at the International Centre for Security Analysis, King's College London.

Jamie Bartlett is head of the Violence and Extremism Programme and Director of the Centre for Social Media Analysis at the think-tank Demos.