

The signature of security

Big data, anticipation, surveillance

Claudia Aradau

‘We are not crystal ball gazers. We are Intelligence Agencies’, noted the former GCHQ director Iain Lobban in a public inquiry on privacy and security by the Intelligence and Security Committee of the UK Parliament (ISC) in the wake of the Snowden revelations about mass surveillance.¹ Several minutes later, Lobban went on to argue that the intelligence agencies ‘have to do detective work’, using the metaphor of finding the needle in a haystack:

If you think of the internet as an enormous hay field, what we are trying to do is to collect hay from those parts of the field that we can get access to and which might be lucrative in terms of containing the needles or the fragments of the needles that we might be interested in, that might help our mission.²

The ‘needle in a haystack’ has become a topos of all intelligence discourse after the Snowden revelations and has recently supplemented the language of ‘connecting the dots’, which security professionals routinely deployed post-9/11. Infamously used by Keith Alexander, former NSA director, to justify the NSA programmes, it has been invoked by the members of the UK’s Intelligence and Security Committee to make sense of GCHQ practices; it is used by MPs in House of Commons debates, by the UN High Commissioner on Human Rights in reports on privacy, and by journalists to render what is at stake in the practices of intelligence agencies concerning big data. It has also become one of the most used analogies in the public discourse of big data and its secret capabilities. While it draws on the language of computer engineering and, in particular, predictive analytics, and in that sense it is not an invention of these public debates, it structures the thinking about mass surveillance by placing it within both an indeterminate history of knowledge and a contemporary ‘big data revolution’.³ For intelligence professionals, the needle in a haystack captures a vision of globality and global threat – according to David Omand, ‘it is a global network and anyone’s information is liable to pop up anywhere in the world’⁴ – and epistemic assumptions of visibility and invisibility, of uncovering secrets and accessing that which is hidden and concealed.

The analogy between the practices of the NSA and GCHQ and ‘finding the needle in a haystack’ has structured the problematization of mass surveillance in terms of how big the haystack should be, how long it should be kept for, whether collecting the haystack constitutes privacy intrusion or mass surveillance, and how much oversight bigger haystacks require. All these questions focus on regulation and oversight and take for granted the self-authorization of knowledge – that big data can lead to knowledge discovery, that collecting a haystack of data can reveal the unknown needles. The production of knowledge by intelligence agencies is justified through the detective work that finding the ‘needle in a haystack’ requires and the spectre of the unknown terrorist who can be discovered through carefully combing the hay, or using a magnet to draw out the needles.⁵

How do the intelligence agencies know? What kind of security epistemics do they rely on? I argue here that the disavowal of 'crystal ball gazing' is as important as the image of finding the clue through the data deluge in order to locate potential dangerous events or individuals in the future. Intelligence work is no stranger to the anticipation of the future – rather, it justifies itself precisely through the capacity to peer into the future in order to prevent or pre-empt future events from materializing. Big data has intensified the promise of anticipating the future and led to 'exacerbate[ing] the severance of surveillance from history and memory', while 'the assiduous quest for pattern-discovery will justify unprecedented access to data'.⁶ 'Knowledge discovery' through big-data mining, and prediction through the recording of datafied traces of social life, have become the doxa of intelligence and security professionals. They claim that access to the digital traces that we leave online through commercial transactions or social interactions can hold a reading of the future. They repeat the mantra of data scientists and private corporations that the 'digital bread crumbs' of the online world 'give a view of life in all its complexity' and 'will revolutionize the study of human behaviour'.⁷

Unlike statistical technologies of governing populations, big data scientists promise that through big data 'we can escape the straightjacket of group identities, and replace them with more granular predictions for each individual'.⁸ To resist their unreasonable promise of predicting crises, preventing diseases, pre-empting terrorist attacks and overall reshaping society and politics, I recast it as divination rather than detection. Big-data epistemics has more in common with the 'pseudo-rationality' of astrology than the method of clues. As such, it renders our vocabularies of epistemic critique inoperative.⁹

Detective work: clues and errors

'The needle in a haystack' relies on an epistemology of the clue, the anomaly that can be uncovered, the concealed secret that can be brought to the surface through the minute and careful work of the intelligence agencies. Carlo Ginzburg has located this model of conjectural knowledge with the judge, the art historian, the doctor and the detective.¹⁰ Although Ginzburg seems to have found inspiration for the insignificant clue in the history of art, it is effectively the figure of Sherlock Holmes who stands for the command of conjectural reasoning. The parallels that Ginzburg has traced between the methods of the art historian, the private detective and the psychoanalyst belong to what he calls a medical semiotics or symptomology in which 'infinitesimal traces permit the comprehension of a deeper, otherwise unattainable reality' – symptoms, clues and pictorial marks are these 'traces' that need to be unearthed and interpreted.¹¹ Ginzburg argues that this conjectural method became used in the human sciences starting from the 1870s–1880s, but its history goes much further back to divinatory practices and hunters' methods of reading traces.

What counts in the production of conjectural knowledge for Ginzburg is not only the relation between surface and depth, between visible sign and deeper reality, but that between the part and the whole. Despite the long history of conjectural knowledge in the practices of ancient hunters, the clue, the insignificant detail, the minutiae of everyday life are symptomatic of the instability of perception in modernity and the emergence of attention as a site of regulation but also aesthetic experimentation. Holmes's attention to detail, his search for the clue, is not simply a re-enchantment of scientific reason, but a re-enchantment of humdrum repetitive industrial labour. Every insignificant detail can gain meaning within a larger whole. How is the clue to be located? An anomaly, as Boltanski notes, exists only to the extent to which it disrupts 'a coherent set of predictable expectations'.¹² For the detective, an anomaly emerges against the background of totality, of social reality and



its regularities. The detective makes ‘the most of his faculty for paying attention to details and his ability to relate these to general laws’.¹³

The logic of detection is also a political logic inasmuch as it stabilizes social reality against uncertain emergencies and reassures the reader about the meaningfulness of minute details and social order. The detective, the policeman and the intelligence agent are all ‘practitioners of a logic of inquiry’.¹⁴ Read through Ginzburg’s conjectural knowledge, this logic of inquiry has been extended not just to the historian and the judge, but equally to the anthropologist and the sociologist. In her analysis of the practices of the Romanian Securitate (secret police), the anthropologist Katherine Verdery finds astonishing similarities between the epistemic practices of the Securitate and her own methods as an ethnographer. For her, the Securitate was making ‘close examinations of everyday behaviour and interpret[ing] what they found’.¹⁵ Yet it is not the attention to insignificant detail that characterizes the Securitate’s logic of inquiry, but the production of banal, everyday or anodyne events as anomalies and clues of dangerous behaviour. This process of producing clues within a narrative of totality is simultaneously (social) scientific and ideological.

On the one hand, the scientific assumptions about the governing of social life meant that the Securitate was particularly concerned about the possibility of error in the production of knowledge. In 1980, a selection from Gilles Perrault’s *L’Erreur* (*The Error*) was translated by the Securitate. The translation remained classified and the text was only meant for the eyes of secret police agents given its ‘radiography’ of the methods used by the French intelligence services, but also the lacunae and ‘inadmissible errors’ that these made, the problems of the abundance of detail and information, and the thin line between plausibility and implausibility.¹⁶ The references to Perrault’s other work, such as *Dossier 51* and *Les Parachutistes*, suggest that these were known to Securitate agents as well. The scientific character of intelligence work was problematized time and again in the documents that the Securitate produced.

In 1989, to mark its forty-year anniversary, the Securitate published a secret document expounding the ‘scientific character’ of intelligence work. As set out in this lengthy document, the information relevant for intelligence purposes needs to meet a series of requirements: ‘veridicity, precision, value, opportunity, viability, frequency,

continuity'.¹⁷ Although it is not quite clear what each of these features are held to mean, they are subsumed to the overarching goals of 'veridicity' and 'valorization'. Veridicity relies on the scientific method, while valorization is defined through the prevention of enemy hostile acts. Here, errors are unthinkable, as party ideology cannot harbour any error. The epistemic contradiction between veridicity and valorization, between the requirement of scientific 'evidence' and party ideology, found a political resolution through the institutional division and hierarchization of labour. Higher officers 'synthesized' the disorderly multitude of information and made sure veridicity and valorization coincided:

the surveillance file was articulated at the intersection of two conflicting practices: that of the informer, denouncer, literary reviewer, and mail censor, who shared an inclination for collecting and recording information; and of the investigator, who collated and synthesized their contributions in a final characterization that reduced their cacophony.¹⁸

The analytics of the clue, of the insignificant detail, did not develop in opposition to dominant readings of reality. Effectively, nothing stands out from the details of everyday life, but a gap needs to be created in order for the ordinary to be produced as anomalous. If the Securitate could produce reality through recasting a multitude of anodyne details into a scenario of foreign threat, infiltration and prevention of hostile acts, the multitude of details also meant that 'veridicity' succumbed to both the material proliferation of different details, meaningless everyday information and an endless variety of interpretations of actions by different informers and officers. The production of files was a veritable industry with a labour force of informers who could give the Securitate access to the everyday banal details that could become the clues. The process was a laborious and time-consuming one, and files numbered thousands of pages of recording, information and analysis.¹⁹ According to Verdery, the archives of the Securitate cover approximately 24 kilometres; the Stasi archives extend over 100 kilometres.²⁰

As files circulate between different intelligence agents, they are read and reworked by different categories of officers; they multiply and expand to such volume that they give rise to a series of short circuits and paradoxes in the production of knowledge. In Verdery's case, one of the Securitate officers arrives at the conclusion that she was *not* an American spy; others that she was. The division of labour, the (non-) alignment of science and ideology, and the material conditions of knowledge production reinforced fallibility and error, whose spectre



haunted both the Securitate and the party. Handwritten comments in the margins of typed-up reports repeat the question 'Is this true?' and confirm the endless suspicion that intelligence officers, informers and citizens were caught in.

Holmes rarely finds himself in error: the admission 'I was in error' is overtaken by his quest to rectify errors committed by others.²¹ Appearances can be easily traversed through 'scientific' interpretation, thus reinforcing the idea that 'people remain totally predictable, or that, at least among those deserving of social power, the desire that could undermine logic and predictability would be self-policing'.²² The analytics of the

clue is bound with codes of class and social order, which make possible the reading of material traces, bodily details, footprints, fingerprints and facial expressions. When these codes are shattered and a process of radical transformation is subordinated to the 'monolithic intentionality ... to steer "all" social processes under the guidance of the latest pronouncements of the party', appearances are no longer stable cyphers of social order. Clues and errors become indistinguishable.

Big data: signatures and similitudes

With big-data surveillance, clues and errors enter a different economy of knowledge. In June 2013, in the wake of the Snowden revelations, the American Civil Liberties Union (ACLU) filed a case against the NSA bulk data collection on constitutionality grounds. The ACLU argued that the NSA programmes infringed the right to privacy and the rights of free speech and association. In a judgment dismissing the case on grounds of legal standing, Judge Pauley III from the South District Court of New York summarizes the epistemic logic of NSA big-data surveillance:

The Government learned from its mistake and adapted to confront a new enemy: a terror network capable of orchestrating attacks across the world. It launched a number of counter-measures, including a bulk telephony metadata collection program – a wide net that could find and isolate gossamer contacts among suspected terrorists in an ocean of seemingly disconnected data.²³

The metaphorical language of 'nets', 'gossamer contacts' and an 'ocean of data' relies on the same epistemic assumptions as finding 'needles in haystacks' and takes the necessity of collecting and creating an ever-larger 'haystack' of data for granted. These statements become meaningful and gain credibility through the promise of algorithms to unveil the 'unknown terrorists' through the anomalous clues and features that cannot be easily clustered and do not fall under a normal pattern. Courts and publics have also accepted the justification that new big-data methods can reveal the new, the anomalous and the atypical.

A very different method of detection is deployed here, one that does not – and cannot – connect the part to the whole. In a world where the 'next terrorist attack' is unexpected and unpredictable, where the future event is always already different from past events, there are no predictable expectations in reality against which the anomaly can stand out. Rather, the intelligence agencies argue that the hay field – all the data – is needed first in order to derive both expectations about normality and the anomalous elements. Big data is the new whole. The normal and the anomalous, the haystack and the clue, are supposed to emerge from big data. The epistemic logic is not that of locating and assembling insignificant details that stand out against the background of social order, but reading the data in order to produce the normal and the anomalous. For security professionals, the hunt for terrorists cannot be based on past knowledge as terrorists change their methods all the time.

How are the normal and the anomalous produced with big data? Rather than statistical regularities, a different economy of knowledge seems to be at stake in big-data mining. Historical regularities – or Sherlock Holmes's voluminous indexes of information and past cases – do not hold the clue to the unprecedented and the unexpected. The logic of big-data mining is that of resemblance and correspondence. It does away with surface and depth, appearance and reality. It privileges analogy, correspondence and similitude.

Through analogy and correspondence, 'clues' are placed within a different mode of epistemic intelligibility. Up to the sixteenth century, argued Foucault, the logic of resemblance 'made possible knowledge of things visible and invisible, and controlled the art of representing them'.²⁴ The logic of resemblance relied on four methods of

knowledge production: convenience (*convenientia*), emulation (*aemulatio*), analogy and sympathy. Convenience produced knowledge through a 'graduated scale of proximity' – things and bodies that were adjacent were also similar. Take, for instance, the associations and patterns that algorithms incessantly churn. Clustering algorithms measure distances between a series of data points and decide on clusters starting from the smallest distance and then enlarging this radius. Proximity to a suspect in a network of relations produces suspicion, and the technique of 'link analysis' is the method of gauging proximity. Emulation does not presuppose spatial proximity, but can eschew distance as it relies on imitation. Rather than proximity, it is the imitation of practices – for instance, using encrypted communications – that becomes the reason for suspicion. Analogy and sympathy are two other ways in which the doubling of the world reveals itself. Analogy is perhaps the most plastic, as it does not require visible similitude or correspondence, but relies on 'the more subtle resemblances of relations', and thus 'it can extend, from a single given point, to an endless number of relationships'.²⁵ Nothing is outside the purview of analogy. Any action, any characteristic, can become the sign of terrorist activity – from buying a one-way plane ticket to reading online material. Finally, sympathy captures links that do not require proximity, but work across large distances. Sympathy can draw together even the most distant things: it has the dangerous power of *assimilating*, of rendering things identical to one another, of mingling them, of causing their individuality to disappear'.²⁶ Discourses of terrorist radicalization via the Internet capture the fear of the sympathy that traverses and mobilizes in unknown ways, that assimilates between proximity or emulation.

The logic of correspondences means that knowledge production relies on an infinite chain of resemblances. Its practices of 'drawing things together' are different from the logic of discrimination that Foucault associated with the classical episteme. Drawing things together seems indeed to legitimate security practices as evading discrimination and differentiation – 'using big data we hope to identify specific individuals rather than groups; this liberates us from profiling's shortcoming of making every predicted suspect a case of guilt by association'.²⁷ Any detail can become a signature of a potential terrorist – in the absence of an ideology of foreign infiltration, socialist lifestyle or public dissidence. In order to be made knowable, resemblances need a doctrine of signatures, of marks that stand in for these resemblances. In Ian Hacking's explanation, 'Signatures are ultimately derived from the sentences in the stars, but a bountiful God has made them legible on earth. Everything is written'.²⁸ As developed by Paracelsus, the doctrine of signatures blurs the distinction between words and things, as one can be substituted by the other.

Although Foucault sees the replacement of the correspondence and resemblance by taxonomical knowledge focused on discrimination and ordering, these modes of epistemic intelligibility did not simply disappear even as they lost their 'scientific' dominance and credibility. Astrology, whose practices of divination were constituted through the logic of correspondence and signatures, was not simply forgotten. Despite public disavowals and legal restrictions, astrologers continued to dispense political advice, popular almanacs of yearly prognostications continued to be published, and the logic of correspondence between stars and society continued to be used throughout the scientific revolution and beyond. Today, horoscopes and astrology columns are a permanent and unquestioned fixture of everyday life, which treat astrology as 'something established and recognized, an uncontroversial element of our culture'.²⁹ Adorno's analysis of the *Los Angeles Times* astrology column resonates with Foucault's epistemes when it casts astrology as a 'vener of rationality ... fused with blind acceptance of undemonstrable contentions and the spurious exaltation of the factual'.³⁰

Big data reasoning combines a veneer of rationality – algorithmic logic and probabilistic calculations – with the irrationality of telling the future from data

'signatures'. Everything has a 'data signature' and everything can be derived from data in a never-ending loop of adding variables and correspondences. Big data is rendered as an inescapable system not only from which there is no place to hide, but where it is impossible to think the error of knowledge. Error does not undermine the production of knowledge, but is integrated in the production of knowledge:

To be sure, erroneous figures and corrupted bits have always crept into datasets. Yet, the point has always been to treat them as problems and try to get rid of them, in part because we could. What we never wanted to do was consider them unavoidable and learn to live with them. This is one of the fundamental shifts of going to big data from small.³¹



Not only do intelligence agencies eschew the problem of error in their practices, data scientists have a new motto: 'big data is messy data'.³² Big data comes from multiple sources and in heterogeneous formats – from commercial transactional data to mobile GPS signal data, and from digital pictures to social media data. Big data is produced by

citizens within daily interactions and collected by private and governmental organizations with the purpose of extracting economic and symbolic capital. Ownership of the instruments for data generation (such as a smartphone or Twitter account) is separated from access to and ownership of the instruments of collection, storage and processing of data. Security institutions and private corporations own and secretly guard their instruments and methods.

In the digital age, the dispersion of data does not create short circuits and paradoxes. Errors in messy data are accepted, and algorithms are designed to be error-tolerant. On the one hand, big data is assigned a 'data veracity' index which simply internalizes error. On the other, error is integrated within algorithmic design, so that error is no longer imaginable in the analysis of the results. The provenance of data, the sources of its production, measuring devices, methods and their credibility diminish in importance as bigger quantities of data are imagined as self-correcting.

For the intelligence agencies, error is also doubly erased through the conjunction of secrecy and divination. Intelligence agencies have always claimed to know, but their knowledge had to remain secret. The report by the ISC in the UK carefully redacted all details about the collection of data by GCHQ, and even examples of the usefulness of big-data mining:

As a first step in the processing under this method, ***. *** the system applies a set of 'selection rules'. As of November 2014, there were *** selection rules. ***. Examples of these initial selection rules are:

include ***;
include ***, and
discard communications ***.³³

The reader never knows the mechanisms through which the results are produced – but needs to rely and trust those who know – while not being able to participate in the production of knowledge. Big data promises security and the capacity to foretell the future; it acquires an occult quality of offering solutions to all social and political problems. The promise of revealing signatures and similitudes through big data – for

instance between a known terrorist and unknown suspect terrorists (raising the problem of how to distinguish between terrorist links and all sorts of everyday, social connections) – big data becomes an ‘abstract authority’ of knowledge. Data and things are linked through infinite resemblances that need to be traced, clustered, patterned.

‘There is nothing irrational about astrology’, concluded Adorno, ‘except its decisive contention that these two spheres of rational knowledge are interconnected, whereas not the slightest evidence of such an interconnection can be offered.’³⁴ The irrationality of big-data security is not in the data, its volume or messiness, but in how a hieroglyph of terrorist behaviour is produced from the data, without any possibility of error.

Notes

1. Ian Lobban, Uncorrected transcript of public evidence before the UK’s Intelligence and Security Committee (ISC), 7 November 2013, <http://isc.independent.gov.uk/public-evidence>.
2. Ibid.
3. See, for example, Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, John Murray, London, 2013.
4. David Omand, uncorrected transcript of public evidence before the UK’s Intelligence and Security Committee (ISC), 23 October 2014, <http://isc.independent.gov.uk/public-evidence>.
5. Both analogies are used in the UK’s Intelligence and Security Committee on ‘Privacy and Security: A Modern and Transparent Legal Framework’, <http://isc.independent.gov.uk/committee-reports/special-reports>.
6. David Lyon, ‘Surveillance, Snowden, and Big Data: Capacities, Consequences, Critique’, *Big Data & Society*, vol. 1, no. 2, 2014.
7. Alex Pentland, *Social Physics: How Good Ideas Spread – The Lessons from a New Science*, Penguin, New York, 2014.
8. Erez Aiden and Jean-Baptiste Michel, *Uncharted: Big Data as a Lens on Human Culture*, Riverhead Books, New York, 2013.
9. Theodor Adorno, *The Stars Down to Earth*, Routledge, London, 2002.
10. Carlo Ginzburg, ‘Morelli, Freud and Sherlock Holmes: Clues and Scientific Method’, *History Workshop Journal* 9, Spring 1980. See also Claudia Aradau and Rens van Munster, *Politics of Catastrophe: Genealogies of the Unknown*, Routledge, Abingdon, 2011.
11. Ginzburg, ‘Morelli, Freud and Sherlock Holmes’.
12. Luc Boltanski, *Mysteries and Conspiracies: Detective Stories, Spy Novels and the Making of Modern Societies*, Polity, Cambridge, 2014, p. 9.
13. Ibid., p. 51.
14. Katherine Verdery, *Secrets and Truth: Ethnography in the Archive of Romania’s Secret Police*, Central European University Press, Budapest, 2014, p. 74.
15. Ibid., p. 7.
16. Ministerul de Interne, *Eroarea*, ed. Departamentul Securitatii Statului, Serviciul editorial si cinematografic, Bucharest, 1980.
17. Ministerul de Interne, ‘Caracterul stiintific al activitatii de securitate desfasurate pentru cunoasterea, prevenirea si contracararea oricaror actiuni ostile, a faptelor si fenomenelor care pot genera sau favoriza comiterea de infractiuni impotriva securitatii statului’, Departamentul Securitatii Statului, Bucharest, 1989.
18. Cristina Vatulescu, *Police Aesthetics: Literature, Film, and the Secret Police in Soviet Times*, Stanford University Press, Stanford CA, 2010, p. 53. Andreas Glaeser arrives at a similar conclusion in the case of the Stasi through ‘continuous displacement of interpretive initiative from bottom to top’; Andreas Glaeser, ‘Power/Knowledge Failure: Epistemic Practices and Ideologies of the Secret Police in Former East Germany’, *Social Analysis*, 2003, p. 22.
19. The file of the writer Dorin Tudoran came to 9,862 pages, solely for the period between 1980 and 1985. Dorin Tudoran, *Eu, fiul lor. Dosar de securitate* [I, Their Son. Secret Police File], Polirom, Iași, 2010. Verdery’s own file came to the more ‘modest’ 2,980 pages.
20. Verdery, *Secrets and Truth*.
21. Arthur Conan Doyle, *The Adventures of Sherlock Holmes*, eBookMall, 2002, www.ebookmall.com.
22. Rosemary Jann, ‘Sherlock Holmes Codes the Social Body’, *ELH*, vol. 57, no. 3, 1990, p. 705.
23. *ACLU v. Clapper*, WL 6819708, Southern District Court of New York, 2013.
24. Michel Foucault, *The Order of Things. An Archeology of the Human Sciences*, Routledge, London, 2005 (1966), p. 19.
25. Ibid. p. 21.
26. Ibid., p. 26.
27. Mayer-Schönberger and Cukier, *Big Data*.
28. Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*, Cambridge University Press, Cambridge, 1975, p. 42.
29. Adorno, *The Stars Down to Earth*, p. 56.
30. Ibid., p. 159.
31. Mayer-Schönberger and Cukier, *Big Data*, ch. 3 ‘Messy’.
32. Adorno, *The Stars Down to Earth*, p. 29.
33. Intelligence and Security Committee, ‘Privacy and Security: A Modern and Transparent Legal Framework’, House of Commons, <http://isc.independent.gov.uk/committee-reports/special-reports>.
34. Adorno, *The Stars Down to Earth*, p. 159.