Big Data and Strategic Intelligence

KEVJN LIM

ABSTRACT This article examines the intersection of Big Data and strategic intelligence from a theoretical-conceptual viewpoint. Adopting Popperian refutation as a starting point, it approaches methodological issues surrounding the incorporation of Big Data into the intelligence cycle, and argues that Big Data analytics is best used to discern long-term developments, generate intelligence hypotheses, and adduce refuting facts. The article then briefly examines the use of Big Data via social media, an increasingly fertile platform for intelligence analysis. Finally, the article argues that despite its potential in filling our epistemic gaps, Big Data should continue to complement traditional subject-matter expertise, supported by game theory, as part of a tripartite analytical framework for strategic intelligence consisting of 'subtext', 'context' and 'metatext'. In this respect, Big Data may well become the midwife for more open modes of intelligence management and, ultimately, a more open society.

Introduction

This article examines the intersection of Big Data and strategic intelligence, and how Big Data may be exploited to reinforce the collection and analysis (or 'research') phases of the intelligence cycle. 'Strategic intelligence' concerns and anticipates events of far-reaching political, diplomatic, social, economic and military significance that often revolve around the questions of war, peace and stability. Where this has failed, strategic surprises have occurred. Foremost in mind are Pearl Harbor, the 1939 Molotov-Ribbentrop pact, the 1973 Yom Kippur War and Sadat's subsequent outreach to Israel, Iran's 1979 Revolution, the collapse of the Soviet Union, the 1998 Indian nuclear test, September 11 and, most recently, the Arab uprisings. Sherman Kent, writing at the start of the Cold War, called it 'high-level foreign positive intelligence … the constructive knowledge with which we can work towards peace and freedom throughout the world, and the knowledge necessary to the defense of our country and its ideals'.[1] In practical terms, this imposes an understanding of the grand strategy, force posture and foreign and national

security policies of the adversary – in other words his intentions and capabilities[2] – and how these elements measure up against and react to 'our own'. Such an understanding also encompasses elements as disparate as culture, psychology, identity and personality,[3] necessitating linkages between intelligence and other branches of government (particularly the foreign service), academia, the media, the corporate sector and, occasionally, other governments.

The present emphasis on the strategic level of analysis should in no way be taken to imply the irrelevance of operational, tactical and technical developments. Force buildup by Hezbollah and insurgents of the self-styled 'Islamic State', or the effective range of Iran's Shahab-5 missiles, for instance, can well give rise to profound strategic implications for the governments concerned. Likewise, cyber attacks in recent years have spawned an entirely different threat arena with increasingly far-reaching strategic repercussions. Consequently, such instances could likewise fall within the ambit of intelligence termed strategic.

Strategic intelligence as a professional discipline and force multiplier has depended foremost on human spies (Humint), and in modern times on an additional array of collection assets spanning signals (Sigint), imagery (Imint) and measurement-signatures (Masint) to open sources (Osint). If the hard core of intelligence analysis still revolves around qualitative subject-matter content analyzed by human specialists, some of these assets have over the years come to leverage on the increasingly massive collection and machine analysis of quantifiable, if not necessarily quantitative, data. On the one hand, 'Big Data' apply across such diverse asset classes in the intelligence toolbox as Sigint (and its two major subclasses electronic and communications intelligence), and so-called 'third-generation platforms' such as social media, smartphones and cloud computing,[4] and in this way behaves no differently from any other intelligence source. On the other hand, Big Data are distinguished by an unprecedented *order of magnitude* when applied to collection and analysis (known as 'Big Data analytics', hereinafter used more or less interchangeably with 'Big Data' unless specified), and it is this special feature with which this article concerns itself.

The open source literature on Big Data is extensive in relation to commercial applications, but scant with regards to national security and,

---

[2]Michael I. Handel rightly pointed out that the adversary's (military) capabilities lend themselves more easily to analysis compared to (political) intentions. For an interesting, if separate, discussion on the dynamic generated between bilateral perceptions of intentions and capabilities, see his 'Intelligence and the Problem of Strategic Surprise', in Richard K. Betts and Thomas G. Mahnken (eds.) *Paradoxes of Strategic Intelligence: Essays in Honor of Michael I. Handel* (Oxon: Routledge 2003) p.13.

[3]Ephraim Kam, 'HaMizrach HaTichon keEtgar Modi'ini [The Middle East as an Intelligence Challenge]', *Strategic Assessment* 16/4 (2014) p.91.

[4]First-generation platforms correspond to the early mainframe, while second-generation platforms comprise client-server technologies and conventional (i.e. structured) relational databases.

worse still, with respect to strategic intelligence.[5] The value that Big Data analytics brings to intelligence work, a realm whose primary stock-in-trade is accurate information, is immense. This article attempts to sketch out a conceptual basis for the incorporation of Big Data within strategic intelligence, and how it might enhance the latter.

The article is organized as follows. The main body begins with a brief description of Big Data *qua* phenomenon. Adopting Karl Popper's methodology as a starting point, it proceeds to some methodological issues surrounding the incorporation of Big Data into the intelligence cycle. The main arguments in this respect are that Big Data analytics is best used to (1) discern long-term developments; (2) generate intelligence hypotheses; and (3) adduce refuting facts. The following section examines one specific use of Big Data analytics in contemporary intelligence analysis, namely social media, especially in democracies and regimes which are highly sensitive to public opinion. Before the conclusion, the article argues that, despite its massive potential in filling our epistemic gaps, Big Data should continue to complement traditional subject-matter expertise, supported by game theory, as part of a tripartite analytical framework for strategic intelligence. The article neither attempts to present the full range of potential uses for Big Data in intelligence work (variants of C4I or integrated Command, Control, Communications, Computer and Intelligence applications for example are already part and parcel of today's kinetic battlespace, but nonetheless fall within the tactical and operational, rather than the strategic realm of activity), nor does it broach privacy and related legal, ethical and moral aspects. Furthermore, it retains a largely theoretical-conceptual approach to the subject and, as such, refrains from analyzing technical applications of Big Data in intelligence, a task better reserved for career data scientists.

The Big Data Phenomenon

The term Big Data refers to massively voluminous, highly varied (i.e. structured and especially unstructured[6]) and dynamic real-time datasets that do not lend themselves to traditional relational data analysis processes. Instead, because of the orders of magnitude involved, the datasets are

[5]See for instance Neil Couch and Bill Robins, 'Big Data for Defence and Security', Occasional Paper, *Royal United Services Institute* (September 2013) <https://www.rusi.org/downloads/assets/RUSI_BIGDATA_Report_2013.pdf>; and David Omand, Jamie Bartlett and Carl Miller, 'Introducing Social Media Intelligence (Socmint)', *Intelligence and National Security* 27/6 (2012) pp.801–23.

[6]'Unstructured' refers to data that are more complex to quantify such as photographs, video images, emails, text messages and so forth. Contemporary advances in automated analysis techniques, Bayesian-based machine learning and data mining allow for the datafication of such sources. According to one report, unstructured data make up over 90 per cent of the digital universe, see John Gantz and David Reinsel, 'Extracting Value from Chaos', IDC iView & EMC Corp. (June 2011) <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>.

captured, ingested and interrogated across a number of servers and the results (of successive iterations) are re-aggregated afterwards in a procedure known as massive parallel processing (MPP) – the functional basis for Big Data analytics.

The key vector for the rise of Big Data is the digitization of information. In 2000, only a quarter of the world's stored information was digital. In 2013, this figure rose to over 98 per cent of the approximately 1200 exabytes (1 exabyte equaling 1 billion gigabytes) of information stored worldwide in all forms.[7] But digitization is only a necessary, not sufficient, condition for Big Data applications. What is still required is *datafication*, that is the conversion of all structured, semi-structured and non-structured information packets into quantifiable units permitting the extraction of new forms of value.

These new forms of value depend on the analyst's ability to interrogate these datasets, via algorithms, to derive insights informing decision-making.[8] Whereas small, representative and hence presumably precise samples once lay at the heart of statistical analysis, Big Data speak to a different methodology and approach altogether; one in which sheer sample size, variety and messiness, backed up by unprecedented storage capabilities, compensate for what they lack in measurement precision.[9] The massive sample size creates something of a normalizing effect and enables higher confidence levels in the inferring of trends, anomalies and patterns which might normally escape notice with small datasets, with implications for future predictions. Ultimately, Big Data analytics shift the focus of inquiry from *causation* to *correlations*: that is the mere knowledge that something is happening, rather than why it is happening, suffices for the formulation of an adequate response.

Big Data applications in commerce, medicine, business intelligence, internal security, financial markets, machine translation, social media and so forth have been extensively documented. Google query inputs, for instance, have been shown to correlate with real-world and even future events such as sales spikes, stock market turnarounds, or even flu epidemics,[10] and Facebook, Twitter and Amazon are able to say a great deal about the likes, dislikes and networks within societies. A type of analysis known as normalized time series has been used to examine time lags in the stock market to detect highly profitable causal patterns.[11] The dangers of 'predictive policing' and how it prejudices free will have also been treated. In all these

---

[7]Kenneth Neil Cukier and Viktor Mayer-Schoenberger, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston: Houghton Mifflin Harcourt 2013) p.9.

[8]Chris Yiu, 'The Big Data Opportunity: Making Government Faster, Smarter and More Personal', *Policy Exchange* (2012) p.10 <http://www.policyexchange.org.uk/images/publications/the%20big%20data%20opportunity.pdf>.

[9]Cukier and Mayer-Schoenberger, *Big Data*, pp.13–14.

[10]Hyunyoung Choi and Hal Varian, 'Predicting the Present with Google Trends', *Economic Record* 88/Supplement s1 (2012) pp.2–9.

[11]Vincent Granville, 'The Curse of Big Data', *Analytic Bridge*, 5 January 2013 <http://www.analyticbridge.com/profiles/blogs/the-curse-of-big-data>.

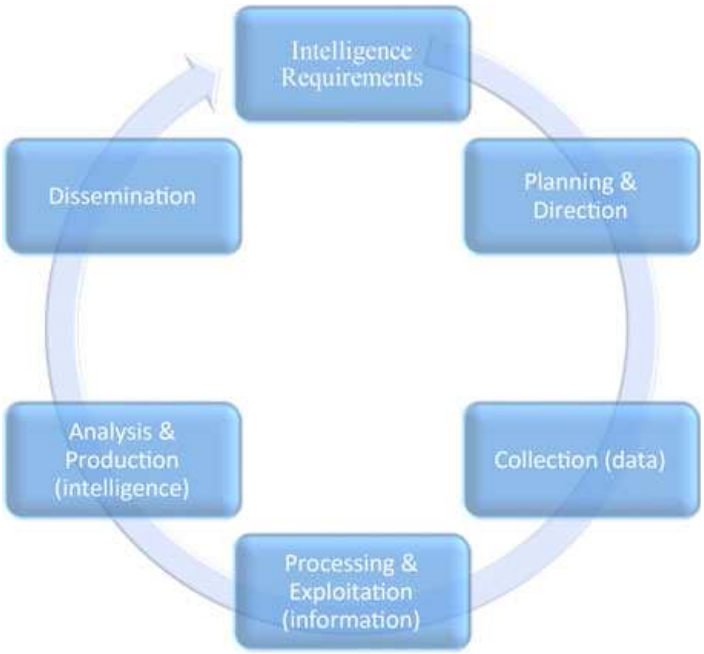

Figure 1. The Intelligence Cycle.[14]
Note: the terms 'data', 'information' and 'intelligence' in parentheses denote: raw unstructured data; data that has been given context and hence informational value; and information that has been analyzed for future implications and shaped into intelligence *for decision support*, respectively.

cases, Big Data can infer the 'probabilistic associations' of future events by relying on a variant of inductive (Bayesian) reasoning and repeated historical patterns (historical corpora).[12] Clearly, this same capability carries significant implications when wedded to strategic intelligence.

Methodological Issues

Intelligence ultimately aims at *reducing uncertainty* in order to support and inform political decision-making,[13] and much of the traditional intelligence method typically involves extrapolating from past events and historical patterns, a process known as induction. Figure 1 captures the standard intelligence cycle, which has its roots in Sherman Kent's pioneering work in the context of the US intelligence community.

---

[12]Kira Radinsky and Eric Horvitz, 'Mining the Web to Predict Future Events', paper given at the 6th ACM International Conference on Web Search and Data Mining/WSDM '13, Rome, Italy, 6–8 February 2012.
[13]'Hagdarat Tafkid haModi'in', 6 August 1958, AMAN, Alef Tzadde 400/765/2004; this notion, by then Head of Israel's Military Intelligence (AMAN) Yehoshafat Harkavy, is cited in David Simantov and Shay Hershkovitz, *Aman Yotze la'Or: HaAsor haRishon le-Agaf ha-Modi'in beTzahal* (Israel: Ma'arachot 2013) p.109.
[14]Adapted from Sean Fahey, 'Big Data and Analytics for National Security', Johns Hopkins University Applied Physics Laboratory, PDF slide presentation (2012) <http://www.stanford.edu/group/mmds/slides2012/s-fahey.pdf>; Appendix C, Loch K. Johnson (ed.), *Handbook of Intelligence Studies* (Oxon: Routledge 2007) p.366.

Against an inductive intelligence methodology,[15] Yitzhak Ben Yisrael has made a case for adopting the critical method defended by Karl Popper (d. 1994), and even much earlier by David Hume.[16] Popper held that a scientific theory can never be verified by experience or observation, *contra* the Baconians and the Newtonians of the earlier modern period. It can only be falsified, and by as little as a single, substantive counterproof (one black swan against a thousand white ones). This leads to a logical asymmetry between corroboration/verification and refutation/falsification, since such theories or propositions 'are never derivable from singular statements, but can be contradicted by singular statements'.[17] Until such a time, the theory merely remains the most robust in existence because it has *stood up to the test of elimination*. The trouble, Popper wrote, is that:

> if we are uncritical we shall always find what we want: we shall look for, and find, confirmations, and we shall look away from, and not see, whatever might be dangerous to our pet theories. In this way it is only too easy to obtain what appears to be overwhelming evidence in favor of a theory which, if approached critically, would have been refuted.[18]

At best, one can only know what is false to be 'true', and critique, he reminds us, is the basis of scientific progress. Ben Yisrael applied Popperian refutation to the intelligence process, where collected data are systematically marshalled to eliminate hypotheses. Figure 2 is an adaption of Ben Yisrael's schema, and implicit in this is a process in which analysis continuously unfolds in parallel with collection, rather than proceed from it along the standard linear pathway.[19] He further amended this by proposing that, rather than automatically refuting hypotheses on the basis of new information, the latter's own assumptions (or hypotheses) should themselves also first be subject to the same rigorous testing standards.[20]

---

[15]The reference is to analytical approaches that deploy some aspect of history to predict the future – historical analogy, situational logic and, in some ways, even the application of theoretical models.

[16]Hume's 'problem of induction', described in his *An Enquiry Concerning Human Understanding*, stemmed from the contradiction between the priority of empirical experience on the one hand, and the inadmissibility of experience-based inductive inferences on the other. See also David Hume, 'Sceptical Doubts Concerning the Operations of Understanding, Part 1, Section IV' in D.C. Yalden-Thomson (ed.) *Hume: Theory of Knowledge* (Austin, TX: University of Texas Press 1953) pp.24–5, 33–7.

[17]Karl Popper, *The Logic of Scientific Discovery* (NY: Harper Torchbooks 1968) pp.40–1.

[18]Karl Popper, *The Poverty of Historicism* (London: Routledge 1957) p.134; Karl Popper, *The Open Society and its Enemies, Vol. II: The High Tide of Prophecy – Hegel, Marx, and the Aftermath* (NJ: Princeton University Press 1971) pp.12–13.

[19]Arthur S. Hulnick, 'What's Wrong with the Intelligence Cycle', *Intelligence and National Security* 21/6 (2006) pp.961–2.

[20]See Yitzhak Ben Yisrael, 'Philosophy and Methodology of Intelligence: The Logic of Estimate Process', *Intelligence and National Security* 4/4 (1989) p.710; and his *Dialogim al mada u-modi'in* (Tel Aviv: Ma'arachot–IDF Ministry of Defense 1989) pp.147–8.

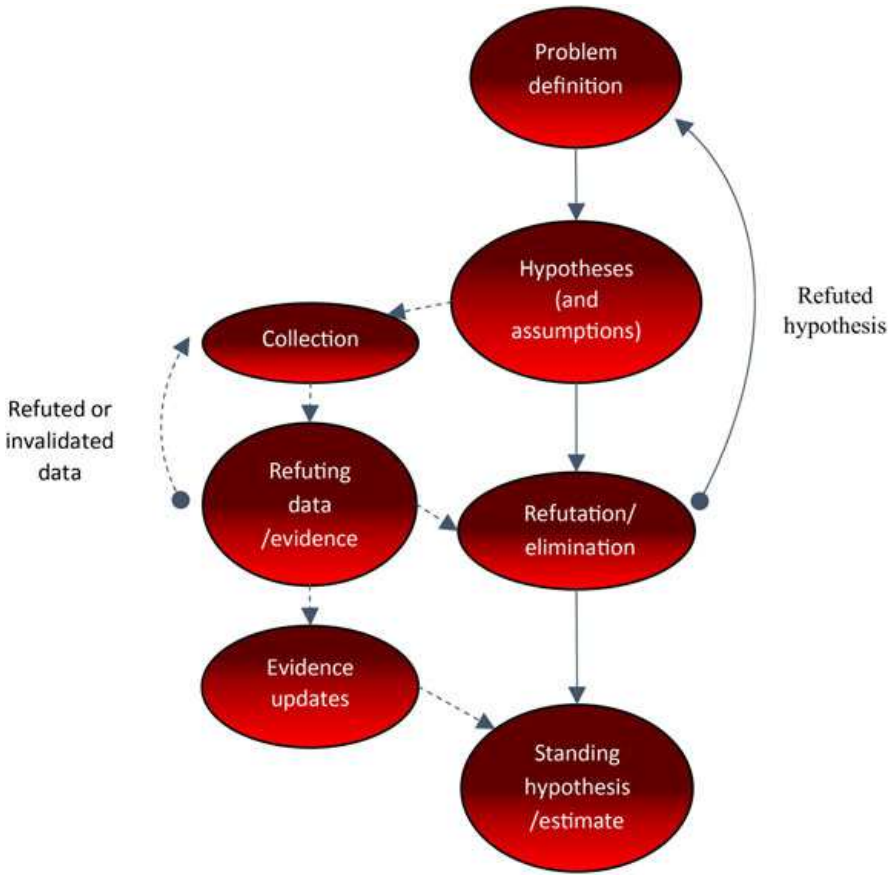

Figure 2. The Amended Refutation Method.[21]

Popper's refutation method has its echoes in the analysis of competing hypotheses, an approach CIA veteran Richards J. Heuer helped flesh out in greater detail in a 1999 book. Despite minor technical and methodological differences such as relatively lower inconsistency scores as a criterion in disproving hypotheses (as opposed to *absolute* inconsistency), and the concept of 'diagnosticity' (consistency of evidence with the various hypotheses) which implicitly recognizes degrees in corroborative evidence, Heuer's framework nonetheless adds valuable analytical nuance.[22]

The inadequacies of the inductive method for deriving the intelligence 'truth' have yielded costly lessons. The lead-up to the Yom Kippur war, which began on 6 October 1973, represents one example, during which the routine predictability of Egypt's 'Tahrir' border exercises effectively reinforced the cognitive preconception or mindset (Heb. *kontzeptzia*) held by Israeli military intelligence that war was not imminent so long as Egypt could not counteract Israel's superior air power. Yet, signs pointing to the contrary were

[21]Adapted from Ben Yisrael, *Dialogim*, p.149.

[22]Richards J. Heuer, Jr., 'Analysis of Competing Hypotheses', ch.8, in his *Psychology of Intelligence Analysis* (Washington, DC: CIA Center for the Study of Intelligence 1999) pp.95–110 <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/>.

already in evidence, including the large-scale evacuation of Soviet officials and their families from Egypt and Syria from 3–4 October onwards, and the deployment of significant bridging equipment on the Suez Canal.[23] In this connection, a subsequent study by Abraham Ben Zvi showed that such military catastrophes have also occurred whenever decision-makers stubbornly clung on to *possible* strategic assumptions even after these have been contradicted by *actualized* tactical indicators.[24]

For the ultimate proof of the poverty of induction, one only has to think of events or processes that have *never* happened before in a particular way in a given context, and for which history as such is no guide. A 1983 report on early warning failures prepared for the US Director of Central Intelligence pointed out that the Sino-Soviet split, the OPEC-induced spike in oil prices from December 1973, the slow-motion overthrow of Haile Selassie's Ethiopia, and the 1979 Khomeini Revolution, to cite only four examples, 'each involved *historical discontinuity*' that challenged the status quo bias (emphasis added).[25] 'The basic problem in each [case of estimate failure]', it continues:

> was to recognize qualitative change and to deal with situations in which trend continuity and precedent were of marginal, if not counterproductive, value. Analysts of the period clearly lacked a doctrine or a model for coping with improbable outcomes.[26]

Given its focus on historical patterns, Big Data-based predictive modelling would appear inherently inductive, and hence inconsistent with the arc of this article's argument. But there is a crucial difference as mentioned earlier. The value of Big Data analytics does not revolve around causal explanations as such, even if its inferences may be based on causal relations within datasets. Rather, it revolves around *correlations* and the identification of phenomena that may not be directly observable or evident, even if the accompanying profusion of data 'noise' and statistically meaningless correlations ironically may also reinforce uncertainty as much as a lack of data does. In scientific terms, Big Data

---

[23]According to this preconception, Egypt would not wage war against Israel if it didn't possess an adequate response to Israel's aerial superiority (especially long range bombers and Scud missiles), and Syria wouldn't attack Israel except in tandem with Egypt; in the event, Egypt surprised by resorting to the use of surface-to-air missiles; See Ephraim Kahana, 'Early Warning versus Concept: The Case of the Yom Kippur War 1973', *Intelligence and National Security* 17/2 (2002) pp.83–7; Michael I. Handel, *Perception, Deception, and Surprise: The Case of the Yom Kippur War* (Jerusalem: Leonard Davis Institute for International Relations 1976).

[24]Abraham Ben Zvi, 'Hindsight and Foresight: A Conceptual Framework for the Analysis of Surprise Attacks', *World Politics* 28/3 (1976) pp.394–5.

[25]Willis C. Armstrong et al., 'The Hazards of Single-Outcome Forecasting', CIA Senior Review Panel declassified report (16 December 1983) p.4 <http://www.foia.cia.gov/sites/default/files/document_conversions/5829/CIA-RDP86B00269R001100100010-7.pdf>

[26]Ibid. loc. cit.; for a deeper discussion of the (un)predictability of Iran's revolution for instance, see Nikki R. Keddie, *Iran and the Muslim World: Resistance and Revolution* (NY: NYU Press 1995) pp.13–33.

epistemology may be said to be *existential* or *nominal*, that is it concerns itself with the mere fact that something is happening and how, as opposed to *essential*, which inquires into the ontological nature of the occurrence.[27]

When integrated into the Popperian method, Big Data analytics serves three apposite and crucial tasks (see Figure 3). The first involves the inductive collection of data with the aim of *discerning general trends and anomalies*. In a sense, this is a variant of applied grounded theory given the emphasis on data 'speaking for themselves' as it were, and facilitates the defining of intelligence problems. Indeed, the discernment of general trends and longer-term developments in itself often constitutes a specific type of intelligence estimate.[28] Intimately tied in with this aspect is, in some ways, the second and perhaps more important function related to the formulation of intelligence hypotheses, a stage requiring as little inhibition from cognitive bias, and as much imagination and (informed) speculation as possible. That the tragedy of Pearl Harbor occurred, Thomas Schelling noted, exemplified a 'great national failure to anticipate' and a 'poverty of expectations'.[29] In another context, he also perceptively identified the 'tendency in our planning to confuse the unfamiliar with the improbable'.[30]

The third, following on from the generation of hypotheses, is that Big Data allow the intelligence analyst to cut through the overwhelming morass of *supporting* facts in order to adduce those with *refutative* value – the search for that one black swan also being naturally far more defined than for the thousandth white one – and this possibly in real-time, an invaluable advantage in intelligence work. This specific task is ideally complemented by the analytical rigor of 'devil's advocates', whose singular task is to challenge baseline or 'conventional wisdom' intelligence estimates with logical, if often far less probable alternatives that require explicit refutal. Woodrow Kuhns has raised the obvious question concerning the situation in which Popperian refutation still fails to eliminate two or more hypotheses.[31] Under such cases of ambiguity, greater weight ought to be shifted towards probability and impact assessments as arbiters (see below). The product, even if imperfect, would be the strengthening of hypothetical approximations in the absence of the true intelligence 'picture'.[32]

In addition, Big Data can increase, by several orders of magnitude, the time spent on analysis and sense-making in relation to collection, provided analysts first possess the tools to make effective sense of the data. A report by

---

[27]Popper, *The Poverty of Historicism*, pp.26–34.

[28]Shlomo Gazit, 'Intelligence Estimates and the Decision-Maker', in Loch Johnson and James Wirtz (eds.) *Strategic Intelligence: Windows into a Secret World, An Anthology* (Los Angeles, CA: Roxbury 2004) pp.137–8.

[29]Thomas Schelling, 'Foreword' in Roberta Wohlstetter (ed.) *Pearl Harbor: Warning and Decision* (CA: Stanford University Press 1962) pp.viii.

[30]Ibid., p.vii.

[31]Woodrow J. Kuhns, 'Intelligence Failures: Forecasting and the Lessons of Epistemology', in Betts and Mahnken (eds.) *Paradoxes of Strategic Intelligence*, pp.80–100.

[32]Ben Yisrael, 'Philosophy and Methodology of Intelligence', pp.693–4.

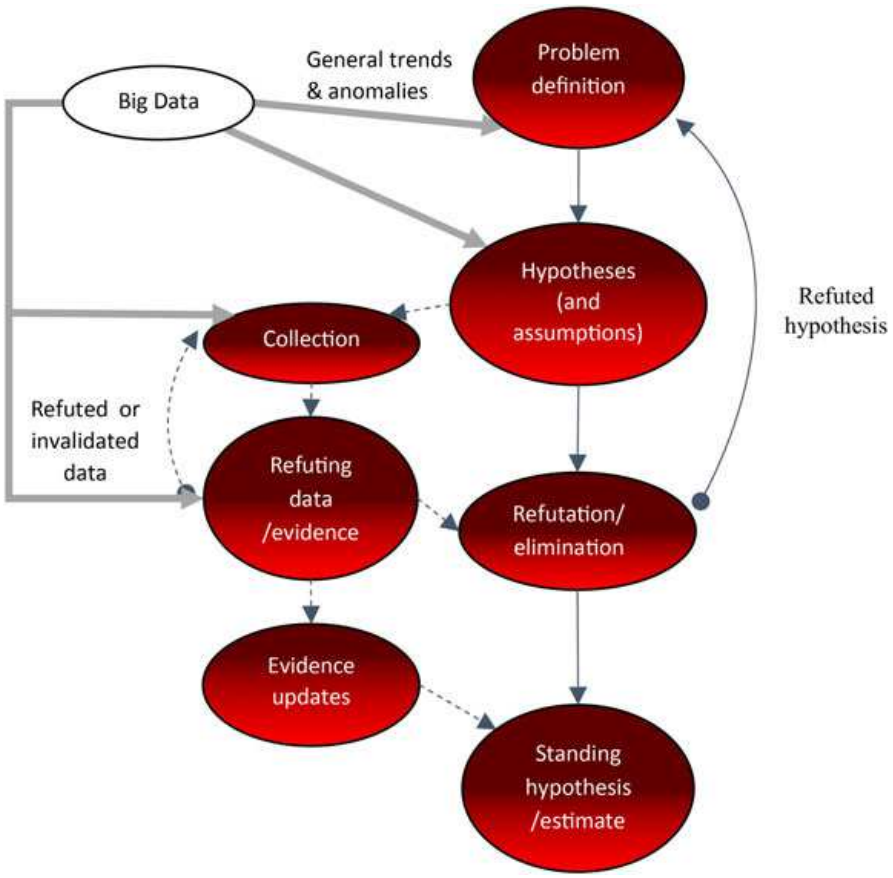

Figure 3. Big Data, Incorporated into the Intelligence Method.

the Royal United Services Institute noted that the 'intelligence world already collects more raw data than it can analyse, with perhaps as much as 95 percent of imagery never being viewed by analysts'.[33] According to the same report, senior Ministry of Defense officials believe that the UK 'has reached an inflection point in data deluge. We are now in danger of data asphyxiation and decision paralysis'.[34] Such is also certainly the case with the US' National Security Agency, where Sigint technology far outpaces the organization's human analytical capabilities.[35] Elsewhere, in order to track the movement of seagoing vessels worldwide, the US navy alone collects 200 terabytes of data approximately *every 48 hours*.[36] That more than a small proportion of all that collected data is processed is equally questionable. Finally, used correctly, Big Data analytics, with the aid of effective visualization and

[33]Couch and Robins, 'Big Data for Defence and Security', p.26.
[34]Ibid., p.9.
[35]Matthew M. Aid, 'The Time of Troubles: The US National Security Agency in the Twenty-First Century', in Johnson and Wirtz (eds.) *Strategic Intelligence*, p.81.
[36]Isaac R. Porche III, 'Why "Big Data" can't find the Missing Malaysian Plane', *US News*, 1 May 2014 <http://www.usnews.com/opinion/blogs/world-report/2014/05/01/why-big-data-cant-find-malaysia-airlines-flight-370>

presentation tools, can shorten the time required for key intelligence to reach decision-makers.

Two further issues with sensitive budgetary implications impose themselves in this context: just how much intelligence is enough (reasonable certitude); and what emphasis to place on which issues (impact). In the intelligence realm, words of estimative probability remain the best end-user tools in service. However, a 'highly probable' estimate cannot be *falsified* since the 'highly unlikely' alternative, should it occur, still absolves the analyst from error. From a purely scientific methodical perspective then, probability as a criterion is meaningless.[37] Yet, if minimizing uncertainty is the best result one can hope to achieve in the real world, then estimative probability remains inevitable. For this reason, the other criterion for determining intelligence and national security priorities, i.e. impact, must also be accorded importance. A critical intelligence task, and one superbly suited to Big Data analytics, is surveillance for warning against unlikely but nonetheless high-impact events or 'black swans', as they are now referred to in the popular literature.[38] One study focusing on civilian applications also cites this specific value of predictive models in identifying 'situations where there is a significantly *lower likelihood* of an event than expected by experts based on the large set of observations and feeds' (emphasis in the original).[39]

Big Data and Social Media

The former CIA contractor Edward Snowden's revelations of the NSA's sweeping internet and telephone data collection program known as PRISM have provoked tremendous controversy, yet they provide an indicator of the scale at which such possibilities operate.[40] An even more prominent and commonplace way in which Big Data are being integrated into national security assumes the form of social media intelligence (Socmint). Twitter, Facebook, YouTube, Instagram, LinkedIn and sundry social media applications have melded into a 'vast digital social commons' capable of facilitating complex analyses of sentiments, semantics, clusters and networks, for instance, in the effort to map, among other things, global Jihadist activity.[41] In the US intelligence community, analysts utilize and often combine a palette of different analytics or metrics software such as Visible:

---

[37]Popper recognizes and criticizes the roots of induction, and by extension the inevitability of an infinite regression in 'probable inferences', see Popper, *The Logic of Scientific Discovery*, p.29.

[38]Michael Herman, 'Intelligence and National Action', in Johnson and Wirtz (eds.) *Strategic Intelligence*, p.226.

[39]Radinsky and Horvitz, 'Mining the Web to Predict Future Events', p.2.

[40]George Friedman, 'Keeping the NSA in Perspective', *Stratfor*, 22 April 2014 <http://www.stratfor.com/weekly/keeping-nsa-perspective>

[41]Omand et al., 'Introducing Social Media Intelligence (Socmint)', p.803; Patrick Radden Keefe, 'Can Network Theory thwart Terrorists?', *The New York Times*, 12 March 2006 <http://www.nytimes.com/2006/03/12/magazine/312wwln_essay.html?_r=0>

Socializing the Enterprise, Geofeedia (which includes facial recognition), the CIA's Open Source Indicators, and the DoD's Information Volume and Velocity program. Commercially available open-source web intelligence programs include Recorded Future and Palantir. As powerful as these platforms may be, and even assuming that data points are all geospatially and time-tagged, they still require that the analyst know *specifically what to look for*. According to a ranking Chicago area security official, the 2008 Mumbai attacks, whose perpetrators were subsequently identified as associates of the Pakistan-based Lashkar-e Taiba, caught US intelligence by surprise despite their heavy pre-attack traffic on social media.[42]

Socmint may be used in a number of ways to inform strategic intelligence via Big Data analytics. At the country level, a wide-ranging analysis of mass sentiments, say social protests following in the wake of the Arab uprisings, can provide proxy indicators of how governments might (be obliged to) react to such 'upstream' threats,[43] and this is relevant for both democracies and authoritarian regimes where sensitivity to public opinion is high and domestic stability is precarious. Cluster and network analyses have been applied in the effort to identify and intercept Jihadist activity and logistics, which often exhibit highly compartmentalized, internal organizational structures with elusive vertical and horizontal node points. Such analyses can also be applied to interactions among state actors or between state and non- or sub-state actors to discern patterns relevant to both intentions and capabilities. Semantic analysis offers an additional tool to decoding and determining intentions. Key stakeholders in the Middle East exhibit increasing savvy and proficiency with social media such as Twitter, and even organizations which used to thrive in secrecy are rarely to be found these days without some kind of interactive online presence. Some such as the self-styled 'Islamic State', which in mid-2014 seized the northwestern third of Iraq from its bases in eastern Syria, have demonstrated an unprecedented level of online persistence and sophistication. Conversely, Socmint has its drawbacks, the most obvious being that social media users may exhibit socio-demographic particularities that do not represent the entire population, making them rather limited 'subsets of subsets of subsets' insofar as content is concerned (this applies even within these groups, for instance in the use of specific Twitter hashtags).[44]

Another similar approach, either through social media or dedicated portals, is crowdsourcing or even 'cloudsourcing' based on a lay crowd-only platform or 'a global network of subject-matter experts'.[45] The US

[42]Personal communication with Joel Vargas, Assistant Director of InterPort Police Global Force, and President and Director of Operations of Contingent Security Services, Ltd, 27 May 2014.

[43]Couch and Robins, 'Big Data for Defence and Security', p.10.

[44]Mark Graham, 'Big Data and the End of Theory?', *The Guardian*, 9 March 2012 <http://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>

[45]The description is Wikistrat's, which touts itself as 'the world's first Massively Multiplayer Online Consultancy', see <http://www.wikistrat.com/about/>

government's Intelligence Advanced Research Projects Activity (IARPA) has been funding a number of such forecast projects including SciCast, which goes beyond simple crowdsourced 'intelligence' by introducing response weighting in relation to previous predictions, inter-variable influences, and the option of modifying forecasts in light of fresh information.[46] In line with this everysource approach, another increasingly fertile avenue for Big Data mining is the so-called 'Internet of Things'.

Whether in the public domain or otherwise, Big Data mining usually depends on networks such as the internet. Yet capabilities currently exist that not only penetrate deepweb peer-to-peer networks known as 'dark nets' which circumvent central servers,[47] but also network-isolated computers, infiltrating and exfiltrating data via devices such as radio frequency transmitters (the 'Quantum' program for instance, a variant of which has now acquired notoriety as Stuxnet).[48] It should by now be clear that Big Data analytics and social media do not necessarily stop where cyber intelligence begins, even though the latter falls beyond the scope of this article.

Big Data within a Broader Analytical Framework

Big Data analytics is an unmistakable force multiplier in the grand intelligence campaign to minimize uncertainty. At the same time, algorithms cannot replace traditional subject-matter expertise or causality-driven theoretical models for that matter, but must complement and, arguably, remain subservient to it.[49] If the volatility of the human subject creates cognitive challenges for the area studies expert, the incorporation of Big Data demands, more than ever, a greater margin of maneuver for human intuition and 'a standard of judgment', in nineteenth century Clausewitzian terms, 'which [can be gained] only from knowledge of men and affairs and from common sense'.[50]

In the context of the US intelligence community, a number of factors currently blight the implementation of Big Data analytics. Robert Steele, a

[46]Patrick Tucker, 'This is How America's Spies could Find the Next National Security Threat', *Defense One*, 20 February 2014 <http://www.defenseone.com/technology/2014/02/long-overdue-return-crowdsourced-intelligence/79094/>

[47]Personal communication with Joel Vargas, 27 May 2014; perhaps more accurately, Peter Biddle et al. define a dark net 'not [as] a separate physical network but an application and protocol layer riding on existing networks', see 'The Darknet and the Future of Content Distribution', ACM Workshop on Digital Rights Management, Microsoft Corporation, Washington, DC, 18 November 2002.

[48]David E. Sanger and Thom Shanker, 'NSA Devises Radio Pathway into Computers', *The New York Times*, 14 January 2014 <http://www.nytimes.com/2014/01/15/us/nsa-effort-pries-open-computers-not-connected-to-internet.html>

[49]Chris Anderson, 'The End of Theory: The Data Deluge makes the Scientific Method Obsolete', *Wired*, 23 June 2008 <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>

[50]Carl von Clausewitz, *On War*, eds. and trans. Michael Howard and Peter Paret (NJ: Princeton University Press 1976) p.117.

leading proponent of Osint and formerly with the CIA and Marine Corps Intelligence, if also admittedly something of an industry iconoclast, identified the following to the author:

> indiscriminate collection in the digital arena with generally no collection in the analog and human arena;
> the absence of analytic model(s) to focus collection on what matters, and to identify gaps where gap-based iterative collection is needed;
> the absence of geospatial attributes for all Big Data (in the realm of national security);
> the absence of a holistic analytic framework that allows all information in all languages – and all human minds engaged with that information – to operate at exoscale.[51]

The NSA, for example, which processes no more than five per cent of the Big Data it collects, excels in 'precision interception for specific purposes … [but] is largely worthless when it comes to global situational awareness, anomaly detection, real-time warning, and pattern analysis across all mission areas'. The system, he continues, 'is designed to throw money at technology for collection, and is not held accountable for failing to process what it collects'.[52] This account appears to gel with others regarding the massive 'firehose' of data, especially from Imint (satellite, U-2 and UAV imagery) and Sigint (especially foreign language) sources.[53] Collection capabilities far outpace analytical capacity. Michael Handel, a pioneer in intelligence studies, argued that strategic surprises, or intelligence failures, are often linked to inefficiencies in analysis and acceptance (of intelligence by policymakers), rather than to collection.[54] The Big Data phenomenon has clearly given rise to an unprecedented glut in collection. Nevertheless, as this author contended earlier, an appropriate response exists in a matching analytics capability, again, provided the analyst knows what to focus on.

The intersection of Big Data analytics and traditional subject-matter expertise must also to an extent incorporate game theory (rational choice), the conceptual framework that best accommodates *animate* players who *react* on the basis of their own intelligence 'picture' and what they think the adversary will do, and whose goal is to maximize self-interest (or 'payoffs') amid uncertainty.[55]

---

[51]Personal communication with Robert D. Steele, formerly with the Central Intelligence Agency and Marine Corps Intelligence, 3–4 June 2014.
[52]Ibid.
[53]Johnson and Wirtz, *Strategic Intelligence*, pp.44–5.
[54]Handel, 'Intelligence and the Problem of Strategic Surprise', pp.8, 12–13; Richard K. Betts, 'Analysis, War, and Decision: Why Intelligence Failures are Inevitable', *World Politics* 31 (1978) pp.66–7.
[55]This is paraphrasing, again, Clausewitz's 'In war, the will is directed at an animate object that *reacts*', see *On War*, p.149.

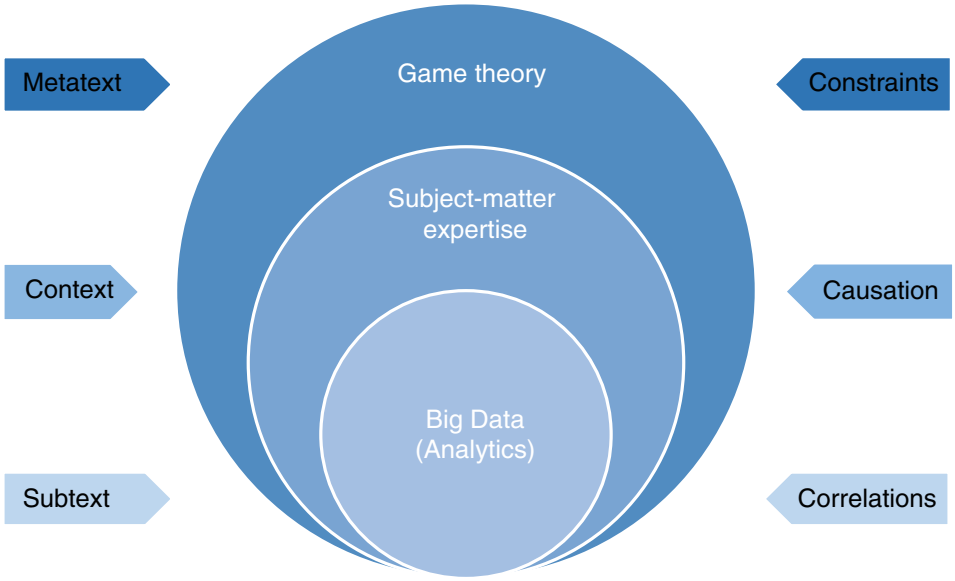

Figure 4. A Tripartite Analytical Framework for Strategic Intelligence.

The tripartite relationship among Big Data analytics, traditional subject matter expertise, and game theory may be schematized in the following manner (see Figure 4). Big Data comprise the raw *text*, or more accurately, *subtext* that conceals correlations and that sits unstirred in the ether, so to speak, awaiting exploitation. Subject matter expertise lends the necessary *context* by imbuing correlative (Big Data) analysis with historical, diplomatic, political, economic, social, cultural and linguistic meaning and hence a causal narrative, and for this reason retains its normative primacy. But to *subtext* and *context* must be added *metatext*, namely a higher, overarching level of analysis comprising strategic calculus. Game theory provides the overall, largely unspoken framework in which all possible strategic responses are conceptualized and the corresponding equilibria, known as best responses, played out by both or more sides. This is a 'framework' in the delimitative sense because strategy is usually driven by ambient circumstances and is hence obliged to operate within these constraints.

Improved capabilities at the *subtext* and *context* levels can render decisive the asymmetry of information – or more properly speaking, intelligence – in strategic assessments, which feed directly into *metatext*. However, game theory's utility in real world situations, given the impossibility of perfect information and the often far less calculated decisional processes of policymakers, means that its task should only be to guide and help refine, rather than drive, intelligence analysis.[56] By integrating data scientists and game theorists, the intelligence endeavor traditionally driven by the subject-matter expert is significantly

[56]Stephen M. Walt, 'Rigor or Rigor Mortis?: Rational Choice and Security Studies', *International Security* 23/4 (1999) pp.17–20 and FN 35; Graham Allison and Philip Zelikow, *Essence of Decision: Explaining the Cuban Missile Crisis*, 2nd ed. (NY: Longman 1999) pp.45–6.

expanded and enriched, without being unnecessarily bogged down in academic deliberations (not for lack of importance but rather for lack of time).

Conclusion

This article began with a description of the Big Data phenomenon, which, characterized by massive volume, variety and velocity, and combined with appropriate analytics capabilities, creates the conditions for a vastly novel epistemic mode concerned with simple *correlations* rather than deep *causation*. The article then examined the fit between Big Data analytics and the intelligence cycle, specifically the collection and analysis components, as it relates to strategic events of far-reaching implications. Presupposing the operationalization of Karl Popper's method, it argues that Big Data eminently suit the existing intelligence methodology in at least three ways: discerning general long-term trends and anomalies; generating hypotheses; and adducing data to refute these same hypotheses. Big Data likewise vastly increase the time spent on analysis and sense-making, whereas at the moment the bulk of the intelligence effort and its resources go towards collection, much of which at any rate is squandered for lack of matching processing and analytical capacity. Real-time parallel processing furthermore collapses the interval required for key intelligence to turn into impactful decisions. The article then provided an example of one area, social media, in which Big Data analytics can complement strategic intelligence. Before concluding, the article proposed conceptually situating Big Data as *subtext* within a tripartite analytical framework that incorporates traditional subject-matter expertise as *context*, and game theory as the overarching strategic *metatext*.

The implications of the Big Data-strategic intelligence intersection reach still deeper and further. From a sociological perspective, the emergence of the Big Data phenomenon is a direct correlate of the information supersociety and the crowdcentric century.[57] In the intelligence context, Big Data analytics goes hand-in-hand with, and is in some ways contingent upon, the rising importance of open source intelligence (Osint) given that the latter constitutes as much as 95 per cent of all useful intelligence;[58] indeed, the primary Big Data tasks sketched out in the chapter on methodological issues are particularly well suited to Osint environments. By the same stroke, the relative importance of secrets to the overall intelligence endeavor has decreased in proportion to the propagation and normalization of information technologies. From the organizational bureaucratic viewpoint, Big Data

[57]Robert D. Kaplan, *The Revenge of Geography: What the Map tells us about Coming Conflicts and the Battle against Fate* (NY: Random House 2012) p.122.

[58]In the era of internet and increased digital communications, the Osint proportion is thought to approach nearly 95 per cent, see Johnson and Wirtz, *Strategic Intelligence*, p.44; the implications of a shifting emphasis towards Osint have been treated extensively by Robert D. Steele, and even taken in the direction of civil and governance reform. See, for instance, his 'Open Source Intelligence (OSINT)', draft (7 April 2006) p.14 <http://www.slideshare.net/RDSWEB/chapter-for-strategic-intelligence-on-osint-single-spaced?qid=8627e280-07d1-4c7d-bd67-5c57b0a361bf&v=default&b=&from_search=4>

prompt and necessitate a shift away from the secretive, highly compartmentalized and rigidly hierarchical mold typifying the realm of intelligence and national security, and towards relatively more open modes of intelligence management. These latter may approximate translateral and highly networked structures not entirely unlike social media,[59] expanding the space for dissent and minority opinions to be expressed, even while a plurality of intelligence agencies may still be retained to improve overall effectiveness and 'competitiveness'. Importantly, this encourages greater levels of disclosure and by implication, *criticism*. For only in the presence of critical feedback can intelligence analysis, like its scientific counterpart, transform into an epistemological edifice capable of adjustment, self-reflexivity and, it follows, progress in the broad sense of the term.

Ultimately, the proper trajectory of all that has been discussed hitherto is likely to lead a step closer towards an open society,[60] one which increasingly engages with decision-making processes behind matters of national security import, and the debates these give rise to. Dare one look further, even the notion of national security under such conditions may conceivably shift from one based on statecentric insularity to one that strives towards a consensual, participatory, decentralized and, eventually, more sustainable form of global security.

Notes on Contributor

Kevjn Lim is an independent research scholar focusing on foreign and security policy in the wider Middle East. He is also an analyst with the UK-based Open Briefing: The Civil Society Intelligence Agency, and consults on developments in the Middle East, where he has been based for nearly a decade. Research for this article benefited from discussions with Yitzhak Ben Yisrael at Tel Aviv University's Security Studies Department.

[59]As David Siman-Tov and Ofer G. point out, such innovations as networked logs and fusion platforms for the sharing of updates in wartime already exist and challenge traditional closed-loop intelligence structures, see their 'Intelligence 2.0: A New Approach to the Production of Intelligence', *Military and Strategic Affairs* 5/3 (2013) pp.35–6. The same authors have forwarded a similar idea, that of a 'shared, networked intelligence space and dynamic, evolving intelligence communities of knowledge', p.41.

[60]If criticism is the hallmark of scientific progress, then it is only in open societies or systems permitting free thought that this is possible. This point is treated at great length, and with great conceptual rigor, in at least two of Popper's works, *The Poverty of Historicism*, and both volumes of *The Open Society and its Enemies*.