Plots, murders, and money: oversight bodies evaluating the effectiveness of surveillance technology

Michelle Cayford, Wolter Pieters & Constant Hijzen

To cite this article: Michelle Cayford, Wolter Pieters & Constant Hijzen (2018) Plots, murders, and money: oversight bodies evaluating the effectiveness of surveillance technology, Intelligence and National Security, 33:7, 999-1021, DOI: 10.1080/02684527.2018.1487159

To link to this article: https://doi.org/10.1080/02684527.2018.1487159

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 22 Jun 2018.

Submit your article to this journal

Article views: 373

View Crossmark data
ARTICLE

Plots, murders, and money: oversight bodies evaluating the effectiveness of surveillance technology

Michelle Cayford, Wolter Pieters and Constant Hijzen

ABSTRACT
Intelligence agencies routinely use surveillance technology to perform surveillance on digital data. This practice raises many questions that feed a societal debate, including whether the surveillance technology is effective in achieving the given security goal, whether it is cost-efficient, and whether it is proportionate. Oversight bodies are important actors in this debate, overseeing budgets, legal and privacy matters, and the performance of intelligence agencies. This paper examines how oversight bodies evaluate the questions above, using documents produced by American and British oversight mechanisms.

Introduction

In September 2002, the U.S. Department of Defense received a Hotline complaint accusing the National Security Agency (NSA) of 'fraud, waste, and abuse' related to the development of the TRAILBLAZER surveillance system, a data collection and processing program. The complaint alleged that the NSA had wasted money on TRAILBLAZER and had chosen TRAILBLAZER over the better (more effective) THINTHREAD program. As a result of this complaint, one of the oversight bodies for the NSA performed an audit on these two systems, concluding in favor of the complaint.1

Recent years have seen an explosion of digital data. This has been followed by intelligence agencies trying to keep up with the flood of information in their endeavor to protect their countries against potential security threats. Rather than drown in the data, they strive to use it to more effectively identify and inform on security issues. As the flood of data rises, so too does the agencies' surveillance of that data.

Subsequently, the increase in surveillance provokes an increased concern of risk of abuse and privacy invasion – more surveillance powers to collect more data increases the likelihood that innocent persons' data may be swept up. To protect against this risk, governments often introduce increased oversight. A law introduced in the Netherlands in 2017 is a case in point. It gives more surveillance powers to intelligence services, allowing them to collect all the data traffic in a certain area in search of a terrorist; to hack the mobile phones of potential acquaintances of suspects; and to share collected, unanalyzed data with foreign intelligence services. To counterbalance this increase in powers, the law also introduces more oversight: the use of any of these new powers requires the prior permission of a special, new committee composed of two judges and a technical expert, in addition to the existing oversight of the Minister of the Interior and the Review Committee (CTIVD).2

Both examples above demonstrate the key role oversight bodies play in intelligence agencies' use of surveillance technology. If this is the role (increasingly) given to oversight bodies, then how oversight bodies perform their evaluations becomes increasingly important. How, then, do these oversight
mechanisms evaluate if a technology is effective in achieving its security goal? And how do they consider cost and proportionality in this evaluation? Do they consider all three in their oversight? Which, if any, of the three take priority? This comparative study investigates these questions, focusing on U.S. and U.K. oversight bodies, particularly those overseeing the NSA, CIA, and GCHQ. It is aimed at these two countries due to the Snowden leaks, which specifically focused on the surveillance of American and British intelligence agencies.

This study complements the authors’ previous paper, which analyzed what intelligence officials of these agencies state regarding effectiveness. It also compares the results of the previous study with what oversight bodies report on effectiveness, cost, and proportionality. Whether these two groups focus on the same things or different ones (e.g., counting money spent, plots thwarted, and murders averted) may impact audit results, like the one on TRAILBLAZER. Understanding how these groups consider the issue of effectiveness is an important first step for any subsequent dialog on the use of surveillance technology and its governing laws.

This study is not an evaluation of whether or not surveillance technologies are effective nor is it a judgment on how oversight bodies evaluate effectiveness. Rather, it is an examination, through analysis of public documents, of how oversight bodies deal with the question of effectiveness as part of their oversight function.

The next section of this paper presents related work in the fields of surveillance and oversight. The study’s methods and approach are then presented, followed by an overview of the oversight mechanisms of the American and British intelligence communities. The data are then analyzed and the findings presented, followed by a discussion of these findings.

Related work

Two bodies of literature were reviewed for this paper: evaluations of the effectiveness of surveillance technology within the security domain and existing studies on intelligence oversight.

Within the broad body of security and surveillance literature, resides a set that deals with the strict effectiveness of surveillance technology. Strict effectiveness assesses and measures whether or not a given security program accomplishes its security goal. A significant body of works exists, which strives to evaluate the effectiveness of counterterrorism measures. In addition, two government reports by the U.S. Privacy and Civil Liberties Oversight Board (PCLOB) address the question of strict effectiveness of security measures, revealing measures of effectiveness used by intelligence officials and drawing their own conclusions about the effectiveness of NSA surveillance programs. In the U.K., David Anderson, an Independent Reviewer of Terrorism Legislation, reviewed the utility of the bulk powers used by the intelligence services. These powers include intercepting and acquiring telecommunications, equipment interference, and using personal datasets, all in bulk (‘bulk’ refers to large quantities of data, including those not associated with current targets). Anderson concluded that these powers are effective in achieving operational aims.

Mueller and Stewart evaluate the strict effectiveness of surveillance technology through cost-benefit analysis. Methods and frameworks related to strict effectiveness include Ekblom’s work, which establishes a framework for crime prevention and security in the community, and Sproles’, which develops a method of establishing measures of effectiveness that can be applied to any field.

A small body of work deals with the strict effectiveness of specific kinds of surveillance technology, including assessing the effectiveness of U.S. Air Force drones, of border security, of wiretapping programs, and of advanced imaging technology full body scanners. Lastly, there is an entire body of literature on CCTV cameras and their effect on crime. Certain recent studies have identified conditions in which CCTV operates most effectively, namely in small, defined spaces such as car parks, and against property crimes rather than violent crimes.

Our previous paper identified several measures that intelligence practitioners use to measure effectiveness: thwarted attacks, lives saved, criminal organizations destroyed, output, context, support, and informed policy-maker. The concept of effectiveness was found to be strongly linked to cost, with
the goal of evaluating surveillance programs being efficient spending rather than effectiveness itself. Intelligence officials were found to rely on the law to determine proportionality and to consider collection by computer versus selection by human beings an important distinction. In the current paper, we compare these findings with those of oversight bodies.

Literature on intelligence oversight addresses the ‘basic problem’ of ‘how to provide for democratic control of a governmental function and institutions which are essential to the survival and flourishing of the state but which must operate to a certain extent in justifiable secret’. This question has raised considerable academic attention within the intelligence studies since 1975, the American ‘year of intelligence’, when major scandals led to reforms in the oversight system. A ‘pattern of exposure, report, and strengthening of oversight’ followed suit in many other countries throughout the 1990s.

Scholars have focused on the historical development of oversight, as well as performing comparative research. Comparing oversight systems and practices in Argentina, Canada, Norway, Poland, South Africa, South Korea, the United Kingdom, and the United States, Born et al. concluded that factors such as independence from the executive, proper investigative powers, access to documents, the possibility to keep secrets, and sufficient support staff make oversight ‘strong’.

Much of this research focuses on legal, formal, and institutional factors that influence the institutionalization of oversight and control bodies. It describes different systems of oversight, ranging from parliamentary committees which exercise oversight on security and defense policies in general, to specialized committees, and non-parliamentary bodies. The research compares different aspects of these systems, such as their composition, selection of members, resources, mandates, the criteria they use, and temporal dimension (ex post or ex ante).

Recently, several authors have started to research the cultural norms and social values that may influence intelligence and security practices, thus exploring the ‘soft side’ of oversight and control. Loch K. Johnson has shown how factors such as member motivation and cooperation by the executive influence the success of oversight and control in the field of intelligence and security. In the case of Dutch oversight, it has been explored how informality and the lack of political weight characterize oversight practices.

Much literature exists which examines the success of oversight bodies – how well do they perform their functions, and are they formally equipped and actually using their powers to successfully oversee intelligence communities? Aside from calls for modernization and ‘digitization’, however, there is no literature on how oversight bodies assess whether intelligence agencies’ use of surveillance technology is effective.

This current study works to fill this gap.

**Framework and methods**

Our conceptual framework relies on distinguishing different elements of effectiveness on the one hand, and the roles of technology, programs, and institutions on the other. This reflects the fact that the question of effectiveness is not determined in a vacuum. Other factors, such as cost and proportionality, are inevitably brought into consideration when determining whether or not to use a particular surveillance technology. We refer to this as overall effectiveness. Ultimately, a decision on overall effectiveness includes considerations of strict effectiveness (whether or not the technology achieves the security goal), as well as of expense and proportionality.

This study defines effective as ‘an impact that is desirable and can be observed as contributing toward the sought-after security goal’. Note that this differs from performance, which refers to the technology’s ability to function correctly. For example, performance tells us whether a technology accurately captures targeted emails, while effective considers whether capturing those emails contributes toward dismantling the criminal organization.

Intelligence oversight is rarely properly defined – instead it is considered a catch-all term for all kinds of practices and institutions and used alongside terms such as ‘accountability’, ‘review’, and ‘control’.
However, most authors understand it along the lines of Hans Born’s definition\textsuperscript{27}: ‘a means of ensuring public accountability for the decisions and actions of security and intelligence services’\textsuperscript{28}.

In the documents analyzed for this study, the term ‘surveillance technology’ as such is not used. Rather, oversight bodies refer to ‘surveillance programs’ or ‘collection programs’. A program could entail only one surveillance technology, but more often it refers to several technologies that together perform a certain collection action, such as collecting internet data, filtering it, and selecting and storing the pertinent data. Consequently, this study often refers to surveillance programs rather than technology. This is with the understanding, however, that the programs focused on here are composed of technologies, and that considerations of effectiveness and the like can equally be applied to both. The types of surveillance programs being discussed in this paper are primarily those dealing with communications data (i.e., surveillance systems monitoring and collecting data on internet and phone activity). Figure 1 shows how the different elements of this framework fit together.

This study analyzes public documents and statements issued from 2006 to 2016 by the oversight bodies of the NSA and CIA in the U.S. and of the GCHQ in the U.K. These oversight mechanisms include the following: the British Intelligence and Security Committee (ISC), Investigatory Powers Tribunal (IPT), Interceptions Communications Commissioner, and Intelligence Services Commissioner; the American NSA and CIA General Counsels and Inspectors General, Director of National Intelligence (DNI), Attorney General, Department of Defense Office of Inspector General, NSA and CIA offices of privacy and civil liberties, Privacy and Civil Liberties Oversight Board (PCLOB), Foreign Intelligence Surveillance Court (FISC), and House of Representatives and Senate Select Committees on Intelligence. The timespan of 2006 to 2016 was chosen to provide a good amount of time (7 years) prior to the Snowden documents to potentially compare differences in evaluation pre- and post-Snowden. It is also the same timespan analyzed in the authors’ paper on intelligence practitioners, allowing for possible comparison between the two studies.

The documents reviewed include all the documents available on the websites of the House and Senate Committees on Intelligence, excluding documents on their rules of procedure (47 documents reviewed). As concerns the FISC, all available – i.e., declassified – orders and opinions were analyzed (30). For the remaining U.S. oversight bodies, any available statements oral or written by these authorities
were examined (27). One source – the audit on TRAILBLAZER – dating from 2004 was analyzed due to its extreme relevance. For the U.K., 13 ISC sources were reviewed, 10 reports by the Intelligence Services Commissioner, and 17 documents produced by the Interception of Communications Commissioner. All relevant judgments of the IPT were analyzed (7) – i.e., those in which the GCHQ was the defendant – as well as the two existing IPT reports.

Analysis was performed by identifying all mentions of effectiveness, cost, and proportionality. These statements were then evaluated, identifying measures of effectiveness and reoccurring themes in all three categories.

**Overview of intelligence oversight bodies**

Before presenting our findings, we briefly introduce the American and British oversight mechanisms and the bodies charged with overseeing the NSA, CIA, and GCHQ. The U.S. intelligence oversight mechanism is expansive and fairly complicated. Here we focus on the oversight bodies studied in this paper.

As the U.S. government has three branches – executive, legislative, and judicial – so does oversight. U.S. oversight can be considered as layers of an inverted pyramid. That is, the first layer of oversight is within the respective agency itself, and from there each successive layer fans out into increasingly broader layers. (see Figure 2) There are several layers of oversight within the executive branch, including boards with particular oversight mandates, such as the PCLOB. The next layer of oversight is judicial – the Foreign Intelligence Surveillance Court (FISC) – but it is also partial, as its jurisdiction is limited to certain forms of investigative actions, such as electronic surveillance for foreign intelligence purposes.

The final oversight layer is legislative and consists of the Permanent Select Committees on Intelligence in the House of Representatives and Senate.

The British intelligence oversight mechanism consists of four pillars, which operate somewhat in successive order, from ministers to judicial commissioners, Parliament, and the IPT. (see Figure 3)

British oversight begins with ministerial oversight. For an intelligence agency to perform any given surveillance activity, a warrant or authorization is required. This authorization is given by certain designated ministers. The requested power is granted only if it is (1) legal, (2) necessary, and (3) proportionate. The second pillar is composed of Commissioners who review the agencies retrospectively, auditing their compliance with the law. The Intelligence Services Commissioner reviews all intrusive actions except interception; interception oversight falls to the Interception of Communications Commissioner.

Under the new Investigatory Powers Act, the Investigatory Powers Commissioner (IPC) will take over the responsibility of both the Intelligence Services and Interception of Communications Commissioners. Warrants will require the approval of both the IPC and a judicial commissioner.
Security Committee (ISC) of Parliament forms the third pillar. The ISC oversees the administration, policy, spending, and activities of the intelligence bodies. Lastly, the IPT exists for individuals to make complaint against the intelligence agencies if they believe they have been the victim of unlawful action or human rights infringement. The Tribunal then investigates and rules whether or not the complaint is justified, issuing orders for the agency in question, if necessary.

What oversight bodies are reporting related to …

Effectiveness

As a whole, it was found that oversight bodies minimally treat the question of the effectiveness of surveillance programs. The PCLOB was the only oversight body found to explicitly and of its own initiative address the effectiveness of specific surveillance systems, recognizing this as a necessary step to address proportionality. The Board’s reviewing of these programs was initiated by requests from Congress and the President, which was a result of public outcry following the Snowden leaks. In the context of its oversight mandate related to privacy protection, the Board was tasked with reviewing two NSA surveillance programs – the bulk collection of domestic phone metadata under Section 215 of the Patriot Act and the collection of foreign electronic communications under Section 702 of the Foreign Intelligence Surveillance Act (FISA).

The one other instance found of an oversight body evaluating the effectiveness of a surveillance system was initiated by a Defense Hotline complaint. The hotline is for reporting fraud, waste, or abuse anonymously. The complaint alleged that the NSA surveillance program TRAILBLAZER was more costly and yet inferior to the THINTHREAD system, but the NSA chose it over THINTHREAD regardless. The Inspector General of the Department of Defense consequently performed an audit on these two systems.
related to cost and effectiveness. Overall the audit’s findings seem to agree with the complaint that TRAILBLAZER was not the best system. TRAILBLAZER was created specifically to effectively exploit the global network. It was a question of which of the two systems was most effective at achieving this goal. The report notes that a separate study ‘observed that the TRAILBLAZER was poorly executed’.

It is significant that this is only the second example we have of an oversight body evaluating the effectiveness of a surveillance technology, and that this evaluation was prompted by a complaint – i.e., it was not a systematic review. Likewise, the origin of the PCLOB’s reports was public outcry. These effectiveness evaluations were reactive and in response to someone crying foul.

In the two reports above, and in other instances where oversight bodies’ reports do point to effectiveness certain trends were identified. These trends are indicated with italics. Oversight bodies place value on surveillance systems providing information that results in the identification of criminals and terrorists and their plots, knowledge about the functioning of their organizations, and the prevention of criminal acts occurring. The U.K. Surveillance Commissioner testifies in his reports of the importance of these collection programs: ‘I have been impressed … by how interception has contributed to a number of striking successes. It has played a key role in numerous operations including … the prevention of murders, tackling large-scale drug importations, … gathering intelligence … on terrorist and various extremist organisations, … serious violent crime and terrorism.’ In his 2010 report, the Commissioner describes an investigation that successfully utilized interception technology, which led to members of the drug organization being identified, a better understanding of the organization’s operations and interactions with other criminal organizations, the prevention of a murder, the seizure of drugs, and the arrests and convictions of principle members. The ISC judged bulk interception to be effective because it has exposed plots.

Likewise, in the U.S., the Attorney General’s Office argues that Section 702 is effective based on it yielding information about the identities and plans of terrorists, and the support and functioning of their organizations. It equally makes an argument for the effectiveness of the metadata collection program when it states that the FBI has opened 27 international terrorism investigations from May 2006 through the end of 2008 based, at least in part, on tips gained from this program.

The PCLOB found the Section 215 metadata collection program to be ineffective based on, in its seven years of existence, there not being a single instance in which it significantly contributed to a counterterrorism investigation, to identifying an unknown terrorist plot, or to disrupting a terrorist attack, and there being only one instance in which a semi-unknown terrorist suspect was identified. Conversely, the Board reported that Section 702 collection led to identifying terrorists or plots in approximately 30 cases, and that it contributed to existing investigations in about 20 cases, ultimately judging it to be ‘valuable and effective.’ Additional measures of effectiveness identified are knowledge gained about the target, as well as the location and movements of suspects.

Oversight bodies also consider the number of reports generated to be a measure of effectiveness. That is, a surveillance program can be measured to be effective or not based on the number of reports generated containing information the system has gathered. Programs that are ‘effective’ are implied to be those that result in reports, testimonies, and briefings of Congress. The PCLOB states that over one fourth of the NSA’s reports on international terrorism include information based in whole or in part on Section 702 collection. The British ISC cites the GCHQ’s increased number of reports and the quality of analysis as an indication of the agency’s effectiveness. This is a judgment in relation to the agency as a whole and not to surveillance technology, but it indicates what the oversight body considers to be a measure of effectiveness.

The PCLOB’s report on Section 215 established seven ‘categories of success’ by which to measure the value of a counterterrorism program. In addition to categories covered in the preceding paragraphs these include measures of enabling negative reporting, adding or confirming details, and triaging. ‘Negative reporting’ refers to establishing that a known terrorist does not have a U.S. nexus. Triaging refers to prioritizing leads based on urgency in a time-sensitive scenario. In the Section 702 report, the Board places value on the flexibility of the program, which allows the government to continue monitoring suspects when they change modes of communication, and to the execution speed of the Section 702 process, which is faster than the traditional warrant process and therefore saves resources.
to a certain kind of data collected, the ISC reports that this data helps the agencies quickly determine
who is a potential target and who to filter out.44 Here again value is given to speed and resources.

Oversight bodies rarely evaluate the effectiveness of surveillance programs themselves. In the doc-
uments analyzed, there were no indications that any of the U.K. oversight bodies explicitly perform
evaluations of effectiveness. And even in the U.S., the two cases cited above – the PCLOB reports and
the Department of Defense Inspector General’s audit – appear to be the exceptions to how oversight
bodies handle the question of the effectiveness of surveillance technology. The norm is not to perform
evaluations of effectiveness themselves, but to depend on the intelligence agencies to do so.45

For example, the U.S. Congressional Committees do not themselves determine effectiveness, but
press the intelligence community to do so. In one Senate Committee report, effectiveness is very spe-
cifically mentioned as something intelligence agencies should assess in a detailed way, to include
measures of effectiveness:

26. Measures of effectiveness

The Committee continued to press the Intelligence Community … to establish quantitative measures of effective-
ness to provide insight into how effectively a program is performing … The Committee is pleased that the IC is
developing more meaningful measures of effectiveness for its programs.46

It is worth mentioning here, a report produced by David Anderson evaluating the operational use
of bulk powers used by British intelligence services (GCHQ, MI5, MI6). At the request of Parliament,
Anderson reviewed these bulk powers (bulk interception, bulk acquisition, bulk equipment interference,
and bulk personal datasets) to assess whether they were useful for the operations in which they were
used, and whether or not other techniques could have been used in their place. Anderson found these
powers to be effective and necessary, using as his measure of effectiveness whether using the power in
question ‘has made a significant contribution’ to the process of identifying potential threats or sources
of intelligence, understanding more fully the intelligence picture, or taking action.47 More specific
activities identified were discovering targets, gaining knowledge about targets, detecting anomalies,
analyzing networks, and triaging and prioritizing. Although not listed specifically as a measure, in his
evaluation Anderson clearly also places value on speed – the speed at which a given power provides
the information over an alternative method. These are, obviously, many of the same measures identified
above. This report is significant in that it focuses specifically on evaluating the effectiveness of surveil-
lance programs within intelligence agencies, and it is public – a rarity – however, it is not produced by
an oversight body and thus does not strictly fall within the bounds of our research.

In the court documents analyzed for this study (i.e. those publically available), the question of strict
effectiveness is assumed by the U.S. FISC and the U.K. IPT. The documents consider the government’s
national security needs, but not whether the program in question is effective in contributing toward
meeting those needs. The program is assumed to be effective and to therefore contribute toward
national security. It is unknown whether non-public documents by the FISC might address effectiveness.
In the case of the IPT, it has not been called upon to address effectiveness.48 It is hypothetically possible
that it could, although this seems unlikely since its mandate revolves around investigating complaints
of unlawful action and human rights infringement.

To summarize, oversight bodies rarely treat the question of the effectiveness of surveillance tech-
nology. Rather, they expect and press the intelligence agencies to do so. When effective or successful
programs are spoken of, oversight bodies place value on systems that thwart plots, provide knowl-
dge on criminal organizations, and result in reports generated. Validating information gathered by
other means, negative reporting, prioritizing leads, and speed are additional measures of effectiveness
important to oversight bodies.

Cost

Oversight related to cost is a frequent subject in the reports of the parliamentary arms of oversight –
the British Intelligence and Security Committee and the American Senate and House Committees.
Their mandate includes overseeing the spending of the intelligence agencies. The ISC’s annual reports review the spending of each intelligence agency. In some instances exact amounts are classified, but the reports document the relative increases or decreases in relation to previous years. One report criticizes an agency for failing to effectively manage its expenditures for the fourth consecutive year. Another states that the Committee’s most significant concern is related to a collaborative savings program, which requires the intelligence agencies to achieve 220 million pounds of savings. It reports that ‘considerable improvements’ are needed if the agencies are to meet this goal by the deadline. Other issues include ‘an SIS payment of several million pounds relating to an operation with a foreign intelligence service which was not adequately documented; spending in excess of Treasury limits on advertising and marketing’.

One of the Senate Committee’s main set of reports is on the Intelligence Authorization Act for each fiscal year. While the Committee itself does not authorize spending, it reports on the authorizations and recommends whether the bill pass. Each report contains a classified section detailing the authorizations. One example of titles within the report further illustrate the cost focus: Budget and Personnel Authorizations; Increase in employee compensation and benefits; and Major System Cost Reports.

These parliamentary oversight bodies link the value of surveillance programs to their cost. The ISC reports on value for money and the efficiency of the agencies’ spending. In one report, it requested the National Audit Office to assess specific projects for value to money. In another, it chides an agency for putting efficiency and ‘value-for-money gains at risk’. The Senate Committee calls for vulnerability assessments of major systems in order to determine ‘whether funding for a particular major system should be modified or discontinued’. Another report requests cost and feasibility studies related to the adoption of certain business systems. These are assessments of the value of a program in relation to its cost.

This leads to the matter of effectiveness being discussed in the context of cost by oversight bodies. While ‘effectiveness’ is mentioned, the focus is really more a question of cost than of strict effectiveness. The ISC reports that the government was developing ‘a framework for monitoring efficiency and effectiveness across the Agencies; and then goes on to discuss resources being used in an effective and efficient manner. Thus, it is actually cost-effectiveness that is being examined. The Committee assesses how the British agencies have performed, stating, ‘[I]t is essential that this level of funding can be justified’. Likewise, the Senate Committee, due to budget cuts, calls for data ‘on the effectiveness of all of the intelligence disciplines … relative to their costs to the taxpayer … Therefore, the Committee directs the ODNI to complete a detailed analysis comparing the effectiveness and costs of the Geospatial, Human, Measurement and Signatures, Open Source, and Signals Intelligence disciplines. The study must include detailed analysis of the costs and effectiveness of subcomponents and major programs.’ Although effectiveness analysis is called for it is relative to and in the context of cost.

The previously mentioned Department of Defense audit of the TRAILBLAZER and THINTHREAD systems is the only example we have of an oversight body performing a cost evaluation of specific surveillance technology. The audit’s findings seem to agree with the complaint that TRAILBLAZER was a costly system. The report quotes another study which states ‘that the TRAILBLAZER was poorly executed and had an overly expensive [classified]’. The audit includes a whole classified section devoted to a cost analysis of THINTHREAD.

Proportionality

Ensuring that intelligence agencies stay within the bounds of the law is a central function of oversight. There are several aspects of conducting legal surveillance, such as following the correct procedure and covering only the permitted persons and communications. Surveillance can be conducted legally according to conditions such as these and yet be disproportionate. Proportionality refers to the impact on privacy versus the benefits for security. Proportionality can be clearly built into the legal statute, or can be more vaguely referenced as something to be sought after, but not specifically required for
legality. When proportionality does appear, it is a sub-category of legality – as such, it is difficult to treat proportionality without also mentioning the broader subject of legality.

We observed that some statements by oversight bodies have more of a legal focus, while others concentrate on the sub-category of privacy and proportionality. Some U.S. oversight bodies seem to address either legality generally or proportionality in particular, while all the U.K. oversight bodies appear to address both broad legality and the proportionality sub-category equally.

**Legality**

Examples abound of various oversight bodies determining the lawfulness of intelligence agencies’ actions. The ISC investigated allegations that GCHQ acted illegally in regards to accessing information gained through PRISM and found that contrary to the allegations, GCHQ acted legally.62 The Intelligence Services Commissioner yearly reports on the lawfulness of the issuing of warrants by the intelligence agencies. In all of the IPT’s judgments, the Tribunal considers the legality of the actions of the intelligence agency concerned.

The NSA Inspector General’s report on the President’s Surveillance Program states that the NSA General Counsel, the DOJ Office of Legal Counsel, and the NSA Inspector General all arrived at the conclusion that the President’s authorization for collection of communications with one end in the U.S. was legal.63 The FISC found the metadata collection program to be lawful 35 times.64 Even U.S. Congressional Committees’ jurisdiction includes aspects of legal oversight, although they do not render legal opinions (e.g., reviewing FISC orders authorizing targeted collection of communications entering or leaving the U.S. if there was probable cause of one of the parties being a terrorist).65

**Compliance** is a common subset theme of legality. Certain oversight bodies produce yearly compliance reports. For example, the Attorney General and DNI jointly produce a semiannual assessment of the NSA’s compliance with procedures and guidelines related to Section 702. These reports detail and number the errors, and determine whether or not intentional violations have been made. Based on one declassified report, we can deduce that these reports also detail the more significant incidents of non-compliance, i.e., those involving U.S. persons.66

The U.K. Interceptions of Communications and Intelligence Services Commissioners report on compliance that relates to their respective jurisdictions. Both produce an annual report documenting the number of errors made by the intelligence agencies. They contextualize the errors by categorizing them and highlighting the severity of the error and the degree of privacy intrusion; examples of errors are also detailed. One such example is that of a GCHQ internal monitoring system of staff communications capturing more information than authorized. The Commissioner concluded this was a technical error. GCHQ deleted the relevant data and reconfigured the system to ensure compliance.67

U.K. and U.S. reports are similar in that they both discuss the types of non-compliance and the number of incidents separated by agency, give examples of the errors, and describe what action was taken to correct the error and to prevent it from happening again. The difference is that the U.K. reports are originally intended for the public, while the U.S. reports are classified.

The term ‘compliance’ is used in discussing errors made that subsequently mean that the agency’s actions were not according to the law. Interestingly, it seems that ‘legal’ is used most often to refer to a surveillance program as a whole or the carrying out of surveillance duties as a whole. ‘Compliance’ is used primarily to refer to the mistakes made within these legal programs.

The oversight documents that report on compliance also address integrity and whether or not the errors were intentional. In all the reports reviewed, the oversight mechanisms stressed the integrity of the intelligence personnel and their desire to act within the law. The language was found to be slightly stronger when the report’s audience was the public at large.

The Interception of Communications Commissioner reported that he found no evidence of a desire to act unlawfully within the intelligence agencies, but rather a clear desire to ensure that their actions are within the law.68 The Intelligence Services Commissioner takes care to stress that in instances of non-compliance, ‘None of the cases involved bad faith or any deliberate departure from established practices.’69 The Attorney General and DNI state that NSA agency personnel demonstrate ‘a focused and
concerted effort’ to comply with requirements, and report that they found no intentional violations in the instances of non-compliance.\textsuperscript{70}

\textbf{Privacy and proportionality}

Proportionality falls within this broader theme of legality. It is closely associated with the notion of privacy protection, the (widely held) belief being that if surveillance is proportional, the privacy of innocent citizens will better be protected. In the U.S., it is most often the PCLOB and FISC that address proportionality and privacy. In both the U.S and the U.K., oversight bodies seek for privacy protections to be built into surveillance systems. The ISC called for privacy protections to ‘form the backbone’ of new legislation being drafted for investigatory powers, and not be handled as a mere ‘add-on’.\textsuperscript{71} The Interceptions of Communications Commissioner considers it his role to ensure that systems are in place to protect the privacy of British citizens.\textsuperscript{72}

The FISC imposes certain measures on NSA surveillance systems to protect privacy, such as instituting regulations regarding accessing and storage of metadata, as well as requiring random spot checks and authorizations for certain activities.\textsuperscript{73} The Court recognizes that the data collected by the NSA under the telephone metadata program will be broad, but qualifies that ‘the use of that information for analysis shall be strictly tailored to identifying terrorist communications’ and must be carried out according to prescribed procedures.\textsuperscript{74} The NSA is only permitted to search this metadata when it has reasonable suspicion that a telephone number is associated with a terrorist suspect.

While privacy controls are in place for various surveillance systems, our research revealed that whether these controls are adequate, and whether in given instances, the privacy invasion is proportionate to the security concern are matters of \textit{human judgment}. This judgment is passed by judges, commissioners, oversight committees and board members.

The ISC judges that the privacy concerns of examining data-sets containing large volumes of data of people of non-interest outweigh the practical considerations (significant increases in the number of warrants issued and therefore also time and cost) of the intelligence agencies; the intrusion merits requiring a specific warrant.\textsuperscript{75} The Intelligence Services Commissioner testifies that the question he focuses on in his oversight is that of proportionality, assessing whether the agencies have correctly balanced the security necessity against the privacy invasion.\textsuperscript{76}

In a report to the FISC, the Attorney General makes a proportionality judgment supporting the NSA’s balancing of security needs, cost, and protection of privacy: destroying credit card information contained in call records requires personnel, time and resources, which ‘are not justified given the operational need for certain information’ and the measures taken to ensure the records are secure.\textsuperscript{77} A FISC judge finds that a two-year rather than five-year retention of upstream acquisitions ‘strikes a more reasonable balance’ between security needs and protecting privacy.\textsuperscript{78}

The PCLOB found the Section 215 program to be disproportionate – one instance of identifying a not entirely unknown terrorist suspect hardly justifies the broad collection of phone metadata. This instance, in particular, is a good example of human judgment at play, because two of the Board’s five members wrote dissenting opinions disagreeing with the conclusion that Section 215 was disproportionate.\textsuperscript{79} One found that the limited amount of information collected by the program, along with the existing and PCLOB-recommended privacy protections renders the privacy intrusion small, while the potential benefit of the program remains significant.\textsuperscript{80}

All the above examples indicated that while laws and measures may be in place to protect privacy and ensure proportionality, the actual determination of whether or not proportionality is achieved must ultimately fall to individual human judgment. And, naturally, different individuals will often arrive at different conclusions.

\textbf{Proportionate ↔ legal}

One finding of this study is that there is a significant interplay between legality and proportionality. In the U.K. case, what is legal is determined, in part, by what is proportional. While proportionality is not the only aspect that determines legality, without proportionality being achieved neither can legality
be achieved. U.K. law stipulates that any surveillance performed with surveillance technology must be shown to be firstly necessary, and secondly proportionate to what it seeks to achieve. Thus, the case for proportionality is built directly into the law. The IPT states that ‘indiscriminate trawling for information … would be unlawful.’\(^{81}\) ‘Indiscriminate trawling’ is considered to be disproportionate which therefore makes it unlawful.

Across the Atlantic, the role of proportionality is more vague. The FISC states that to assess reasonableness (proportionality), a court must consider ‘the nature of the government intrusion and how the government intrusion is implemented. The more important the government’s interest, the greater the intrusion that may be constitutionally tolerated.’\(^ {82}\) The court goes on to state that if the privacy protections are adequate ‘the constitutional scales will tilt in favor of upholding the government’s actions.’\(^ {83}\) If, on the other hand, the protections are inadequate to protect against the risk of error and abuse, the balance will tip toward a judgment of unconstitutionality. Determining whether a program is constitutional or legal, therefore, includes determinations of proportionality. The law requires U.S. intelligence agencies to implement measures to protect privacy. Whether or not a given program is found to be legal or not, however, is based upon whether these measures are deemed to be adequate in light of the government’s interests. If a program is determined to be proportional it is considered constitutional. If it is determined to be overly invasive, it will be found to be unconstitutional.

**Discussion**

This section contains themes that were identified across all three elements of effectiveness, cost, and proportionality and compares the results of this paper with the authors’ previous study on intelligence practitioners.

While this study intentionally covered a pre- and post-Snowden timespan, no significant differences were found in how oversight bodies dealt with the three aspects of overall effectiveness (strict effectiveness, cost, proportionality) in these two different periods.

**Oversight bodies and intelligence practitioners compared**

Comparing the results of this study with the findings regarding intelligence practitioners revealed some notable similarities and differences.

The question of the effectiveness of surveillance technology is rarely treated by oversight bodies and intelligence practitioners alike. In their investigation of two of the NSA’s surveillance programs, the PCLOB specifically raised the question of whether these programs were effective. The Board’s hearings, in which they posed this question to intelligence officials, is the sole instance found of intelligence practitioners specifically addressing this question. The PCLOB’s reports are also only one of two instances found of oversight bodies doing so. Instead, discussions turn around cost or privacy and proportionality issues, or focus on existing or new oversight mechanisms to implement.

Among the measures of effectiveness identified, oversight bodies and intelligence practitioners signal several of the same measures. Thwarting plots, identifying and locating criminal and terrorist suspects, and providing knowledge of the structure and workings of criminal organizations are considered by both to be ways of evaluating effectiveness. The number and quality of reports generated based on information gained from a surveillance program is also considered to be an important measure of effectiveness by both stakeholders.

Both groups also revealed cost as a driver of evaluations of effectiveness. Effectiveness is evaluated not so much out of concern for effectiveness itself, but out of concern for cost. Oversight bodies call for assessments evaluating value to cost, and practitioners evaluate systems because both parties only want money spent on programs that are effective.

Proportionality judgments involving individual human judgment were a theme apparent in both studies. It is individual human beings who ultimately decide what is proportional both within oversight mechanisms – courts, boards, legal offices – and within the intelligence agencies – e.g., directors.
One notable difference is in the interplay between legality and proportionality. Intelligence practitioners state that proportionality is determined by the law: they themselves do not make proportionality judgments – they simply act according to what the law prescribes. The law determines what is proportional, and this is enforced by oversight. Our research findings on oversight bodies seem to reveal the reverse of this logic. That is, that legality is determined, in part, by what is proportionate. The law lays down certain procedures (e.g., FISC guidelines), but there is still room for, and it is even necessary to have, judgments of what is proportional. These proportionality judgments are part of what determine legality.

Dependency

Although many oversight elements are independent of intelligence agencies and therefore their work is independent, our research revealed that they are heavily dependent on the intelligence community for the documents and testimony necessary to carry out their oversight. Likewise, they are dependent on intelligence agencies to determine the effectiveness of surveillance technology. The technical expertise necessary to perform these evaluations lies largely within the intelligence agencies. While some oversight bodies draw on outside technical support (e.g., the PCLOB held a public forum which included a panel of technology experts), these experts do not have access to classified material and therefore are not advising specifically on the surveillance systems in question. A notable exception is the IOCCO, which includes technical experts as part of its inspection team. However, the IOCCO does not explicitly evaluate effectiveness.

The Senate Committee’s conducting of its oversight of NSA electronic surveillance was assisted by briefings by the NSA and access to court documents. The PCLOB’s reports on NSA surveillance programs relied on testimony, hearings, and evidence received from the members of the intelligence community. Likewise, the ISC concluded that media allegations that GCHQ circumvented British law were false based on evidence given by GCHQ.

This inter-relatedness is also evident in the measures of effectiveness unearthed in our two studies. At least part of the reason both practitioners and oversight bodies come up with similar measures of effectiveness is that the oversight bodies are relying on intelligence officials to indicate how to evaluate the effectiveness of surveillance programs.

This dependency does not equate to oversight bodies giving the agencies a green pass at every turn. Examples abound of oversight bodies finding fault with surveillance programs: the audit on TRAILBLAZER, the PCLOB 215 report, FISC and IPT judgments, and compliance issues raised by the FISC and the U.K. Commissioners. It is, however, an interesting point that, in order to conduct their oversight (including any conclusions on effectiveness), oversight bodies must rely on the intelligence agencies themselves for testimony, documentation, error reporting, and the like. This dependency seems inevitable. It is the members of the intelligence agencies who carry out the surveillance actions and use the technology. Arguably, therefore, they know best the functioning of the systems, what actions they have taken, errors they have made, and the subsequent documentation. This inter-reliance, however, explains why certain groups and individuals claim that oversight bodies no longer serve their purpose and have been co-opted (e.g., Greenwald). It also points to an issue of trust.

Trilemma

Many oversight bodies are given a mandate that focuses on one of the three elements of effectiveness, cost, and legality/proportionality. For example, the IPT was established to handle complaints regarding unlawful or disproportionate actions by the intelligence agencies. The NSA General Counsel is charged with providing legal advice. Consequently, their activity focuses on the given element. Any given report tackles only one and occasionally two of these elements together: the U.K. Commissioners’ reports address compliance; the ISC annually reviews the intelligence agencies’ spending; the PCLOB reports on NSA surveillance programs analyze effectiveness and legality; and the Inspector General’s
audit on TRAILBLAZER and THINTHREAD reviews the cost and effectiveness of the programs. The Bulk Powers Review, although not produced by an oversight body, was launched to investigate effectiveness and specifically excluded proportionality. A final example is that of the Senate Committee initiating an in-depth review of the legality and cost-effectiveness of intelligence collection programs. This example is particularly interestingly because it is a review specifically of surveillance programs, and it focuses on cost and legality, but not on strict effectiveness. In the documents studied, oversight bodies were never found to address all three elements of effectiveness, cost, and proportionality simultaneously. Mechanisms designed to deal with all three, like the Senate and House committees, never evaluated all three at once.

The fact that no agency or oversight body addresses all three elements together reminds us of a well-established theory in macroeconomics – the impossible trinity, or trilemma. This theory states that policymakers in open economies must choose two out of three conflicting, yet desirable goals: monetary independence, exchange rate stability, and financial integration. Because it is impossible to have all three, policymakers must decide which one they will give up. No such formal framework exists in the security realm, however, the same reality is present. While oversight bodies (and intelligence practitioners) speak of simultaneously delivering effective surveillance, in a cost-efficient manner, while maintaining proportionality, these are, in fact, conflicting goals.

Both domains contain three conflicting goals, and in practice, stakeholders address only two of the three goals simultaneously. This trilemma concept points to why many oversight bodies are tasked with performing only one of these activities, such as overseeing the protection of privacy or of legal compliance. Others are tasked with two or all three missions, but alternately perform them one (or possibly two) at a time. This allows oversight mechanisms to successfully treat the issue at hand without having to enter into the impossible task of successfully addressing all three elements.

Conclusion

As digital data have become increasingly important to society, so too has it become central to the work of intelligence agencies. As their surveillance of this data increases, so does the importance of the work of oversight bodies. Consequently, how oversight bodies consider and evaluate the overall effectiveness (including effectiveness, cost, and proportionality) of surveillance technology is a crucial question.

Oversight bodies were found to minimally treat the question of strict effectiveness. Instead they rely on the intelligence agencies to perform evaluations of effectiveness. Measures of effectiveness that oversight bodies were found to value are thwarted plots, knowledge gained, reports, validating and prioritizing information, and speed. These are similar to the measures identified for intelligence practitioners, pointing to a dependency of oversight bodies on intelligence officials to indicate how to evaluate surveillance programs. Oversight bodies are equally dependent on the agencies for the documentation and testimony necessary to perform their oversight.

Overseeing spending is the specific mandate of certain oversight bodies. In this context, they speak of the value of surveillance programs in relation to their cost. Evaluations of surveillance technology focus on cost to value considerations. This is another similarity found with intelligence practitioners.

Ensuring agencies and their surveillance technology stay within the bounds of the law is an important and central function of oversight. In addition to ensuring the legality of surveillance programs, oversight mechanisms report on compliance, enumerating, and investigating errors made within legal programs. Oversight bodies ultimately rely on judgments of proportionality to help determine legality, while intelligence practitioners demonstrate the reverse, relying on the law to determine proportionality.

Oversight mechanisms typically have a mandate concerned with one of the three elements of effectiveness, cost, and proportionality. If their mandate includes more than one of these three, in any given report they only evaluate one or two of the three, but never all three simultaneously. Successfully addressing all three is an impossible trilemma.

The results of this study, along with those of the authors' previous paper, are an important component of the ongoing discussion surrounding surveillance technology. Understanding how intelligence
practitioners and oversight bodies treat questions of effectiveness, cost, and proportionality, and weigh these elements against one another is crucial to creating meaningful dialogue between these groups and others, such as the public and privacy advocate groups. Such dialogue is necessary to build trust, without which the use of surveillance technology might undermine the democratic culture it is meant to protect.

Notes

1. Inspector General of Dept. of Defense, “Audit of TRAILBLAZER.”
5. PCLOB, “Section 215” and “Section 702.”
8. Ekblom, SIs framework.
11. Tsvetovat and Carley, “Effectiveness of Wiretap Programs.”
19. Born et al., Who’s Watching the Spies?
20. Born and Leigh, Making Intelligence Accountable; Born and Caparini, Democratic Control of Intelligence Services; Willis, Understanding Intelligence Oversight; Born et al., International Intelligence Cooperation; Willis and Vermeulen, Parliamentary Oversight of Intelligence; Born and Willis, Overseeing Intelligence Services; Born et al., Making International Intelligence Cooperation Accountable; Gill, Intelligence Governance and Democratisation.
21. Willis and Vermeulen, “Parliamentary Oversight of Intelligence.”
22. Davies, “Intelligence Culture and Intelligence Failure” and “Ideas of Intelligence;” O’Connell, “Thinking About Intelligence Comparatively;” de Graaff and Nyce, Handbook of European Intelligence Cultures; Krieger, “Oversight of Intelligence.”
24. Hizjen, “More than a Ritual Dance.”
29. IPT website.
30. PCLOB, “Section 215” and “Section 702.”
31. Ibid, 27.
38. PCLOB, “Section 702,” 2.
The Dutch oversight body mentioned in the introduction – CTIVD – states specifically on its website that it does not evaluate effectiveness. It is interesting that its function so definitely does not include this, and that it so pointedly wants to distance itself from any expectations of effectiveness evaluations.

TRAILBLAZER was eventually abandoned after over $1 billion had been spent on the program. (Mayer, “The Secret Sharer”).

Disclosure statement

No potential conflict of interest was reported by the authors.
Notes on contributors

Michelle Cayford is a PhD candidate, whose research is focused on the evaluation of the effectiveness of surveillance technology in intelligence work. She obtained her Master’s degree in International Security from Sciences Po in Paris, and her Bachelor’s degree in History and French from the University of Washington in Seattle. Her professional experience includes work at NATO in the Secretary General’s Policy Planning Unit, and as a lead criminal intelligence analyst in a U.S. government counter-narcotic trafficking unit.

Wolter Pieters is an associate professor of cyber risk at Delft University of Technology, Faculty of Technology, Policy and Management. He has MSc degrees in computer science and philosophy of technology from the University of Twente, and a PhD in information security from Radboud University Nijmegen. His research focuses on cyber security risk management and decision-making in complex systems. He published widely on electronic voting, verification of security properties, cyber risk management, human factors in security, and philosophy and ethics of security.

Constant Hijzen is an assistant professor in Intelligence Studies at the Institute of Security and Global Affairs and the Institute for History at Leiden University (the Netherlands). In his dissertation, he focused on the political, bureaucratic, and societal context of the Dutch security services. His postdoctoral research focuses on intelligence cultures and the specific role of paradigm shifts from a comparative perspective.

Bibliography


Assistant Attorney General, NSA Deputy Director, and General Counsel ODNI. Joint Statement Before the Permanent Select Committee on Intelligence House of Representatives at Hearing on “FISA Amendments Act Reauthorization” (2011).

Assistant Attorney General, NSA Deputy Director, and General Counsel ODNI. Joint Statement Before the Senate Select Committee on Intelligence at a Hearing on “FISA Amendments Act Reauthorization” (2012).

Attorney General. “Minimization Procedures Used by the NSA in Connection with Aquisitions of Foreign Intelligence Information Pursuant to Section 702 of FISA.” September 21, 2016.

Attorney General. “Minimization Procedures Used by the NSA in Connection with Aquisitions of Foreign Intelligence Information Pursuant to Section 702 of FISA.” March 24, 2017.

Attorney General. “Procedures Used by the NSA for Targeting Non-U.S. Persons Outside the U.S. to Acquire Foreign Intelligence Information Pursuant to Section 702 of FISA.” July 10, 2015.

Attorney General. “Procedures Used by the NSA for Targeting Non-U.S. Persons Outside the U.S. to Acquire Foreign Intelligence Information Pursuant to Section 702 of FISA.” March 29, 2017.


Brand, Rachel L. “Testimony of Rachel L. Brand, Member of the Privacy and Civil Liberties Oversight Board, before the United States Senate Committee on the Judiciary.” (2016).


Davies, P. H. J. “Intelligence Culture and Intelligence Failure in Britain and the United States.” *Cambridge Review of International Affairs* 17, no. 3 (October 2004): 495–520.


Eskens, Sarah, Ot van Daalen, and Nico van Eijk. Ten *Standards for Oversight and Transparency of National Intelligence Services*. Institute for Information Law, University of Amsterdam, 2015.


Intelligence and Security Committee."Access to Communications Data by the Intelligence and Security Agencies."February 2013.
Intelligence and Security Committee."Statement on GCHQ's Alleged Interception of Communications under the US PRISM Programme."n.d.


Judge, Foreign Intelligence Surveillance Court. "Amendment to Order for Purposes of Querying the Metadata Archive, BR 07-10." May 31, 2007.

Judge, Foreign Intelligence Surveillance Court. "Amendment to Primary Order, BR 11-07;" February 10, 2011.


Judge, Foreign Intelligence Surveillance Court. "Order, BR 06-08." August 18, 2006.

Judge, Foreign Intelligence Surveillance Court. "Order, BR 06-12." November 15, 2006.


Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 07-16;" October 18, 2007.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 08-01;" January 2008.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 08-04;" April 3, 2008.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 08-07;" June 26, 2008.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 08-08;" August 19, 2008.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 08-13;" December 12, 2008.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 09-01;" March 5, 2009.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 09-06;" May 29, 2009.

Judge, Foreign Intelligence Surveillance Court. "Primary Order, BR 10-10;" February 26, 2010.


Judge, Foreign Intelligence Surveillance Court. “Primary Order, BR 11-57.” April 13, 2011.
NSA Inspector General, NSA General Counsel, and NSA Director. “Report to the Intelligence Oversight Board on NSA Activities - 4Q 2011.” April 5, 2012.


Senate Select Committee on Intelligence. “Central Intelligence Agency’s Detention and Interrogation Program,” December 9, 2014.


Senate Select Committee on Intelligence. “Postwar Findings About Iraq’s WMD Programs and Links to Terrorism and How They Compare with Prewar Assessments.” September 8, 2006.


