Reduced Analytic Uncertainty through Increased Analytic Rigour: Effects of Using Structured Analytic Techniques in Estimative Intelligence

A thesis submitted for the degree of

Doctor of Philosophy

By

Lars C. Borg

Brunel Centre for Intelligence and Security Studies

Brunel University, London

31 May 2022

Abstract

The post-9/11 intelligence reform's endorsement of Structured Analytic Techniques (SATs) as a remedy to intelligence failures has been increasingly criticized, with a main argument being that SATs do not eliminate bias or increase judgement accuracy. However, previous research on SATs has predominately only focused on one technique, Analysis of Competing Hypotheses (ACH), without procedural understanding of its use in the intelligence process. This research sets forth a comprehensive SATs-model for improving the estimative process in intelligence and investigates analysts' experienced effects of using this SATs-methodology in the intelligence process. The main finding is that analysts experience that using SATs in a comprehensive, layered, and iterative manner in the estimative intelligence process increases analytic rigour and their ability to assess uncertainty. Using a set of interconnected SATs in a creative, critical, and sensemaking logic-process increases analyst's analytic objectivity and integrity and thereby making them more confident in their key judgments and aware of the attached uncertainties. The effects furthermore remained years after having received SATs training. However, agency culture plays a vital role, both for the use of SATs and for the use of community standards for communication of uncertainty. Hence, leadership endorsement is important for SATs to be used by more than the dedicated few who have SATs as part of their analytical DNA.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACH	Analysis of Competing Hypotheses
AMAN	*Agaf HaModi'in* (Israeli military intelligence directorate)
BJP	*Bharatiya Janata* Party (India)
CIA	Central Intelligence Agency
COA	Course of Action
DCI	Director of Central Intelligence
DI	Defence Intelligence (United Kingdom)
DIA	Defense Intelligence Agency (United States)
EFA	Exploratory Factor Analysis
GBN	Global Business Network
ICA	Intelligence Community Assessment
ICBM	Intercontinental ballistic missile
ICD	Intelligence Community Directive
IntBn	Intelligence Battalion (Norway)
IRTPA	Intelligence Reform and Terrorism Prevention Act
ISTAR	Intelligence, surveillance, target acquisition, and reconnaissance
JIC	Joint Intelligence Committee (United Kingdom)
JIO	Joint Intelligence Organisation (United Kingdom)
NATO	North Atlantic Treaty Organization
NCO	Non-commissioned officer
NCW	Network Centric Warfare
NIE	National Intelligence Estimate
NIS	Norwegian Intelligence Service

NORDIS	Norwegian Defence Intelligence School
ODNI	Office of the Director of National Intelligence
PCA	Principal Component Analysis
PHIA	Professional Head of Intelligence Analysis
PST	Police Security Service (*Politiets sikkerhetstjeneste,* Norway)
SATs	Structured Analytical Techniques
SME	Subject Matter Expert
SWOT/TOWS	Strengths, weaknesses, opportunities, threats
WEP	Words of Estimative Probability
WMD	Weapons of Mass Destruction

Acknowledgements

To my family:

Thank you for putting up with me, sitting in the office for hours on end and reminding me that there is more to life than writing a PhD.

To the Norwegian Defence University College:

Thank you for funding this PhD, granting me the opportunity to become a Norwegian “Doctoral Officer”.

To my supervisors, Dr Gustafson, Dr Pickering, Professor Pedersen:

Thank you for your guidance, insight, and endless support.

To all my fellow officers:

‘The nation that will insist upon drawing a broad line of demarcation between the fighting man and the thinking man is liable to find its fighting done by fools and its thinking by cowards.’
--Maj Gen Sir William Francis Butler, from his biography of Charles George Gordon, 1889

Introduction

Intelligence is wittily talked about as the world's second oldest profession. Despite its longevity, there is still a lack of a common intelligence theory, let alone a universally agreed upon definition of intelligence.[1] A main reason is that intelligence can be interpreted either in a narrow sense, as secretly collected information, or in a wide sense, as all-source intelligence estimates intended to improve policymaking.[2] In conjunction with the definition debate of intelligence there has been a debate as to whether intelligence analysis should be perceived as being art or science.[3] In this debate, intelligence analysis as an art is generally attributed to the complexity of the intelligence problems with numerous different variables attached, the often uncertain and incomplete information, and the free will of humans, making it difficult to communicate an estimate with the degree of precision wanted by policymakers.[4] Conversely, intelligence as a science is associated with applying methods from the social sciences, often modelled after natural sciences, while admitting that these methods do not have the same level of precision as natural sciences.[5] The lack of a unified understanding of intelligence and intelligence analysis has led to a development of many competing intelligence theories but the either–or debates have arguably impeded the theoretical development since intelligence constitutes of both secret collection and all-source analysis, using elements from both art and science.

This thesis attempts to bring new knowledge to the body of intelligence analysis theory and thereby also bring the unproductive art versus science debate closer to termination. The research explores intelligence analysts' experienced effects of using a comprehensive Structured Analytic Techniques (SATs) methodology in estimative intelligence with a focus on analytic rigour and the ability to assess uncertainty. The term Structured Analytic Techniques (SATs) was first introduced in 2005 but the concept of what was previously

---

[1] Gill and Phythian, "Developing Intelligence Theory."
[2] Davies, *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 2: Evolution of the UK Intelligence Community*.
[3] Richards, *The Art and Science of Intelligence Analysis*; Marrin, "Is Intelligence Analysis an Art or a Science?"
[4] Richards, *The Art and Science of Intelligence Analysis*, chap. 5.
[5] Kent, *Strategic Intelligence for American World Policy*; Lee, "A Scientific Methodology for MIS Case Studies."

known as Alternative Analysis can be traced back to the CIA in the 1970s and 1980s.[6] The strive towards more rigorous, objective, and scientific orientated intelligence estimates is nonetheless even older, going back at least to President Wilson's The Inquiry.[7] Several American committees, both governmental and within the intelligence community, have since addressed the need to adapt and reform the tradecraft of intelligence analysis to prevent intelligence failures.[8] A game-changer was, however, the committee reviews and critique of the intelligence effort voiced after the 9/11 terror attacks and the infamous estimates of Iraq's WMD-programme.[9] The latter resulted also in an uncommon public critical review of British intelligence.[10] The consequence of these committee recommendations was an intelligence community reform that has been more wide-reaching and with a larger impact than previously witnessed. A core reason is that unlike previous reforms, this time the changes for the American intelligence community became enshrined in law though the Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004.[11] The new legislation started a renewed community-wide focus on two enduring topics in estimative intelligence; how to increase analytic rigour through improved analytic tradecraft and how to improve communication of uncertainty in intelligence estimates. This effort spread to Europe, much due to the international effort of fighting Islamic terrorism.

Both increased analytic rigour and improved communication of uncertainty are central elements of a new intelligence failure theory: discourse failure. Opposed to various other intelligence failure theories, discourse failure takes a more comprehensive standpoint, focusing on the policymaker-intelligence interface when arguing that intelligence fails due to a ‘failure of comprehension: the constriction of the language and vocabulary to identify,

---

[6] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020, 5–7; Heuer, “Adapting Academic Methods and Models to Governmental Needs.”
[7] Gelfand, *The Inquiry: American Preparations for Peace, 1917-1919*.
[8] Two examples are: Jeremiah, “Recommendations of the Jeremiah Report”; CIA Senior Review Panel, “Report on a Study of Intelligence Judgments Preceding Significant Historical Failures: The Hazards of Single-Outcome Forecasting.”
[9] National Commission on Terrorist Attacks Upon the United States, “The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States”; Silberman and Robb, “The Commission on the Intelligence Capabilities Weapons of Mass Destruction: Report to the President of the United States.”
[10] Lord Butler, “Review of Intelligence on Weapons of Mass Destruction.”
[11] United States Congress, Intelligence reform and terrorism prevention act of 2004.

analyze, and accept that a significant threat exists.'[12] The theory of discourse failure therefore changes our understanding of the role of intelligence. Traditionally, the essence of intelligence has been understood as how to improve decision-making by making uncertain estimates less uncertain.[13] The latest intelligence reform, however, requests analysts to assess uncertainty in estimative intelligence products; through an objective analysis of alternatives present both probabilities for these alternatives as well as a level of analytic confidence in the judgements.[14] This new approach contains three intertwined parts connected to analytic judgements: the use of probabilistic expressions, statements of analytic confidence, and the use of an intelligence analysis methodology, helping analysts to identify alternatives and assess the associated uncertainty.

The latest intelligence reform led to a renewed academic focus on analytic tradecraft and uncertainty handling in the intelligence community. Following the reform, SATs have become the *de facto* "science" of intelligence analysis, which again prompted an increased focus on researching the validity of SATs. To this day the research has mainly had a two-fold focus, either on SATs as debiasing techniques or SATs' effect on increasing accuracy of analytic judgements. The research results have yielded mixed results, especially concerning SATs' effect on reducing bias. Much of the previous research has unfortunately been underwhelming of several reasons. First, previous research has only reviewed single SATs, predominately Analysis of Competing Hypotheses (ACH), in which the main problem lies in how hypotheses have been generated, how much evidence has been used, and the participants' experience. Second, the previous research has had an approach often disconnected from a working intelligence environment. Instead of investigating how ACH and other SATs are being used as part of the estimative process in an intelligence community, the research has had a narrow and limited approach, disengaged from intelligence analysis as part of the overall intelligence process. Third, contrary to what some claim, to prevent bias using SATs alone is an impossible task. Heuristics are inescapable and consequently most often also the biases derived from these heuristics. The tendency of

[12] Neumann and Smith, "Missing the Plot? Intelligence and Discourse Failure." 96
[13] Clark, *Intelligence Analysis: A Target-Centric Approach*, 8–9; Omand, "The Future of Intelligence: What Are the Threats, the Challenges and the Opportunities?," 14.
[14] United States Congress, Intelligence reform and terrorism prevention act of 2004. Sec. 1019

biases to impede judgements can be reduced by harnessing the psychological mechanisms that lead to them, but the use of single SATs alone is not enough. Fourth, arguing for SATs to provide more accuracy can be the result of being victim of hindsight bias. There exists no methodology that with certainty can say what will happen in the future. Moreover, most probability theories belong to natural sciences and mathematics, which makes little sense in estimative intelligence that is mostly concerned with complex problems in the social world. Lastly, there is a missing dimension of the previous research since it has assumed that SATs are taught and used the same way across different nations and different agencies, something which is far from the truth. The bottom line is that much of the criticism towards SATs is unwarranted, being a result of conjecturing and theorizing about SATs, often based upon the art versus science debate, yet without a sufficiently thorough study of the sources and the material on the actual uses of SATs. This thesis, however, argues that the art versus science debate is a false dichotomy that has led intelligence analysis theory astray and that a different approach is needed when conducting research on intelligence analysis tradecraft.

This thesis shifts the paradigm on researching the effects of SATs. First, it breaks new ground by demonstrating that there is a large variance between different NATO nations on the approach to teaching and using SATs. Second, instead of looking at only one SAT, this thesis investigates the key effects of using a comprehensive structured analytic methodology, built around a layered and iterative use of several different SATs, taught and practised in a two-week SATs-course where practical use in a simulation exercise is at the centre. Third, this research is more extensive than previous studies as the findings in this thesis are based upon a mixed research approach using both quantitative and qualitative data from participants from the three main Norwegian intelligence agencies. Fourth, as opposed to previous research that has taken a snapshot review, this thesis has also researched the long-time experienced effect of using SATs. Lastly, the expertise of the author of this thesis resides not only in the academic work but also in the expertise as a professional intelligence analysis practitioner, contributing a real-life procedural understanding while simultaneously providing a more scholarly than anecdotal character to the results. Together this produces the most comprehensive and far-reaching research on SATs to this date.

This research has primarily investigated the experienced effect that the SATs-methodology taught at the Norwegian Defence Intelligence School (NORDIS) has on analytic rigour and on analysts' ability to assess uncertainty in estimative intelligence. The research on these topics within the field of intelligence theory is in its dawn, although there have lately been a few interesting separate results about the use of verbal probabilities, the concept of analytic confidence, and the value of using certain Structured Analytic Techniques. Unfortunately, these topics have been treated as single working silos. There are, however, and as this thesis demonstrates, interlinked issues of uncertainty, analytic rigour, and the use of intelligence analysis methodology. The novelty of this thesis lay in the combination of these issues, particularly that analysts experience and are of the opinion that the use of a comprehensive SATs-methodology increases analytic rigour through improved objectivity and integrity and thereby increasing their ability to assess uncertainty.

The thesis is divided into three main parts. Part one is dedicated to the first chapter, which covers previous research on SATs, the research design and methodology of this thesis, and a literature review. First, the chapter will present and discusses the current research on the effects of SATs, arguing that much of the criticism is based upon the pretences of having investigated a true analytic process. In fact, however, previous research has only investigated one SAT, Analysis of Competing Hypotheses (ACH), and not in a uniform manner or as part of a process resembling most real-life usages. Hence, previous research has for the most part lacked procedural understanding and researched subsidiary issues. Next, the novel approach of this research is presented, describing why a mixed methods approach was best suited to answer the research questions that are centred around analysts' experiences and opinions of using a comprehensive SATs-methodology. Due to its ground-breaking focus, the research has also been an exploration for discovery. Quantitative data has been collected from participants attending the two-week NORDIS SATs-course in addition to data collected from a follow-up questionnaire issued to intelligence analysts in three different Norwegian intelligence agencies. The analyses of the quantitative data derived from questionnaires have been used to find significant relationships between different variables representing operationalized phenomena. Subsequently, a series of semi-structured interviews has been conducted among intelligence analysts from the same three Norwegian intelligence agencies. The qualitative data from the interviews has then been used to find and understand any underlying reasons for the quantitative results. The chapter finishes with a literature review, explaining why it has been necessary to look at several different academic fields to present a more comprehensive picture and to add new knowledge to the current body of intelligence theory literature.

Part two presents and discusses the current theoretical viewpoints concerning intelligence failure, analytic rigour, uncertainty, and SATs. Chapter 2 briefly discusses reasons for intelligence failures, where in addition to the traditional explanations, discourse failure as a new and more comprehensive theory of intelligence failure is introduced. The nature of cognitive limitations has led to a pessimistic view of intelligence failures being inevitable, where Betts for instance argue that 'cognition cannot be altered by legislation.'[15] The chapter will therefore discuss the impact the American Intelligence Reform and Terrorism Prevention Act (IRTPA) and a few other critical committee reviews have had on the community-wide intelligence reform. As discourse failure concerns a failure of communication and comprehension that a threat exists or can emerge, the chapter argues that the new legislature and subsequent reform is aimed at remedying discourse failure through adherence to new analytic tradecraft standards. The promulgation of these analytic standards has generated a renewed focus on two core areas of estimative intelligence. The first is to increase analytic rigour by means of improving intelligence analysis tradecraft. The second is to improve the handling and communication of uncertainties in intelligence assessments. Both are core elements of estimative intelligence.

Chapter 3 addresses the lack of understanding that intelligence cannot be understood as a uniform construct but rather that there are several different typologies of intelligence, where estimative intelligence more and more has become the benchmark for decision support. Estimative intelligence, as opposed to other types of intelligence, involves both knowledge and understanding and can be defined as '[looking] at what might be or what might happen.'[16] The chapter gives an overview of different knowledge claims, problem types, and their impact on estimative intelligence. Estimative intelligence mainly concerns intelligence problems typically labelled mysteries and complexities, problem types riddled

[15] Betts, "Analysis, War, and Decision: Why Intelligence Failures Are Inevitable."
[16] Central Intelligence Agency, "CIA: Who We Are and What We Do," 2.

with uncertainty and alternative outcomes as opposed to secrets where there exist definite answers. There are therefore both epistemological and ontological issues that impact on key judgements. The chapter concludes that knowledge claims in estimative intelligence are fallible and thus in need of an improved intelligence analysis methodology.

The fourth chapter presents the term analytic rigour and discusses what constitutes a rigorous estimative process for intelligence analysts. Increasing analytic rigour has bearing on the ontological and epistemological nature of intelligence analysis where different problem types warrant different approaches due to the problem of heuristics and biases. It will be argued that analytic rigour can be improved by employing different elements of logic, creativity, and critical thinking in a comprehensive manner that will increase analysts' objectivity and integrity, both in the estimative process and the resulting products. The main argument of the chapter is that analysts must cultivate analytic rigour through a compound mode of reasoning and sensemaking together with creative and critical thinking. The significance of this argument is that the art versus science debate is a false dichotomy, and that estimative intelligence must use principles from both to increase analytic rigour.

Chapter 5 addresses the fact that producing intelligence estimates usually involves making probabilistic statements about the future. Despite the centrality of probability and uncertainty in estimative intelligence, the intelligence community lacks a common understanding of central theories concerning the field of uncertainty. It is furthermore evident that the intelligence community's intermittent focus on improving uncertainty handling in intelligence assessments and its lack of theoretical understanding not only has resulted in an inability to agree on a common usage or standard, but it has also fuelled discourse failure. Research within other fields has demonstrated that verbal probabilities are highly ambiguous and that how uncertainty is expressed affects the recipients' understanding of the probabilistic message. The chapter therefore argues that assessing uncertainty in estimative intelligence is better handled two-folded by presenting both the relative probability of the alternative outcomes as well as the analytical confidence attached to the judgements, indicating the basis for the judgements and what it may take to change.

The sixth chapter is ground-breaking as, for the first time, different nations' approaches to Structured Analytic Techniques (SATs) are compared. Separate starting points and developments have led different nations to disparate approaches to teaching and using SATs. The American approach to teaching and using SATs is one of stand-alone techniques, while the general European approach, in comparison, treats SATs as a part of an integrated social science approach to intelligence analysis. Here the Norwegian approach stands out as more layered and iterative than most other approaches, combining a collaborative use of different SATs with creativity, critical thinking, and a compound mode of reasoning and sensemaking. By integrating a variety of SATs in a comprehensive, layered, and iterative forecasting methodology, analysts can combine horizon scanning and strategic outlooks with tactical indications and warnings. The key argument is that this comprehensive approach to teaching and using SATs lay a foundation for increased analytic rigour through analytic objectivity and integrity and thereby make it easier for analysts to assess uncertainty. The desired outcome is to present more rigorous and comprehensible intelligence estimates to policymakers, seeking to prevent discourse failure.

Part 3 is dedicated to presenting and discussing the results of the data on the effects of SATs, collected from Norwegian intelligence professionals. This research has taken a novel approach, instead of answering an impossible question of either bias mitigation or judgment accuracy, the research questions and answers are importantly different when investigating SATs' effect on analytic rigour and on analysts' ability to assess uncertainty. Chapter 7 presents and discusses the results of survey 1, which was dedicated to investigating the effects on analytic rigour and ability to assess uncertainty after having attended the two-week long NORDIS SATs-course. Chapter 8 presents and discusses the results of survey 2, which was investigating the more long-term effects of SATs training and use. This chapter also compares the results of the two surveys. Chapter 9 analyses the qualitative interview data collected from key intelligence professionals in three different Norwegian intelligence agencies. The last chapter takes on an overall discussion of the effects using SATs have on analytic rigour and on analysts' ability to assess uncertainty. The main finding of this research is that when using SATs in a comprehensive, layered, and iterative manner, analysts' experience increased analytic rigour and that it is easier to assess uncertainty. By increasing analytic objectivity and integrity, a comprehensive use of SATs improves the quality of the analytic process and thereby reduces analytic uncertainty. There are nevertheless cultural differences between the different Norwegian intelligence agencies that impact on the use of SATs and how uncertainty is assessed and communicated in estimative products.

The results of this research have direct relevance for the intelligence community due to the sector-overarching nature of contemporary intelligence problems where the intelligence community needs to look at several interlinked factors to make better forecasts about the future. Not only can the findings have a large impact on future training of intelligence analysts, but they may also impact on cultural and organisational issues as well. Most importantly, however, is the potential impact the key findings have on limiting the possibility for future intelligence failures, especially with reference to the renewed focus on strategic warning, intelligence estimates that no longer are limited to the military sphere.

Part 1: Research on Structured Analytic Techniques (SATs): the need for a new approach

The quest to prevent intelligence failures by improving intelligence analysis tradecraft has been an enduring one, reaching back for at least a hundred years. Different official committees and review boards have intermittently addressed the need for intelligence tradecraft reforms. The call for researching the effects of intelligence tradecraft reforms is a more recent development, arguably a result of an increased focus on intelligence theory as an academic field. Research focusing on the post-9/11 intelligence reform is a case in point, where there still are many aspects in need of scientific discovery. The current research on the effect of Structured Analytic Techniques (SATs) is both limited and narrow, having yielded mixed results for several reasons. This first part will present reasons for why most of the previous research on SATs has been underwhelming and then present the research design and methodology of the novel research of this thesis. Part 1 closes with a literature review, explaining the need to look at literature from different academic disciplines to bring new knowledge to the field of intelligence analysis theory. The main argument of part 1 is that instead of investigating to what degree SATs help reach unattainable goals by researching techniques as one-offs without procedural understanding, a new and different approach is needed, one that researches the effects of SATs as taught and used by intelligence professionals as part of the overall intelligence process.

Chapter 1: Research design and methodology

Improving the tradecraft of intelligence analysis has had an intermittent focus, especially in the United States, ever since Sherman Kent's days. The United States Congress Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 was, as Chapter 2 will discuss, a game changer, instigating mandatory community wide changes to the tradecraft of intelligence analysis to prevent discourse failure. A result is that the use of Structured Analytic Techniques (SATs) has started to become the norm in the Western intelligence community. This focus on SATs as a universal solution has met increased criticism. As an example, the NATO System Analysis and Studies Panel (SAS)-114 found what was 'important conceptual shortcomings in the overall approach to analytic tradecraft' and 'significant deficiencies in [tested] methods.'[1] While some of the criticism is justified, one can question other parts, especially since Analysis of Competing Hypotheses (ACH) is the only technique that has been tested repeatedly and the research has produced diverse results.

This first chapter of the thesis will start by discussing the validity of the criticism towards SATs, underlining the divergent research designs and lack of procedural understanding of estimative intelligence analysis. Subsequently, the research design and methodology of this thesis will be presented, emphasizing that when researching intelligence analysis tradecraft, it is more advantageous to pursue an exploratory mixed methods approach to cover more aspects of interest. Lastly, the current literature on intelligence analysis theory will be reviewed, addressing the problem of academic working silos. The key point is that this research will bring new knowledge to intelligence analysis theory by taking a more comprehensive viewpoint of different factors that impact on estimative intelligence, especially intelligence failures, analytic rigour, uncertainty handling, and the teaching of SATs.

[1] Mandel, D., *Briefing Note to NATO Office of Chief Scientist on NATO Science & Technology Organization Activity SAS-114*, April 2018. The only method scientifically tested was ACH, see Mandel, "Communicating Uncertainty, Assessing Information Quality and Risk, and Using Structured Techniques in Intelligence Analysis."

The validity of SATs – the emperor's new clothes?

The implementation and use of Structured Analytic Techniques (SATs), previously known as Alternative Analysis, has been advocated and endorsed by academics, intelligence review boards, and large intelligence agencies alike. The main arguments among the advocates for changes have been that a more structured approach to the estimative intelligence process should not only harness intuition and reduce biases, but also increase analytic rigour in the process, resulting in more accurate estimates.[2] The steadily increased focus on SATs has nonetheless not come without criticism, from academia as well as within intelligence agencies. One recurring argument is that using SATs is too time-consuming.[3] Another argument is that SATs are inferior when dealing with complex problems that have infinite variables like when dealing with the unpredictability of the social world.[4] A third argument is linked to naive falsification, that when dealing with too many hypotheses and not having information to support any of them, a continuous mindfulness is more suitable than occasionally employing a set of Structured Analytic Techniques.[5] However, research show that our intuition can only deal properly with a few variables at the time and relying on mindfulness alone may result in inconsistent and biased analysis due to heuristics.[6] Time-pressure is generally also not a reason analyst themselves bring up for not using SATs.[7]

SATs are also criticised for not increasing the accuracy of intelligence assessments.[8] This criticism must be understood through an American lens, here using the words of James

---

[2] Heuer, *Psychology of Intelligence Analysis*; Folker Jr, "Intelligence Analysis in Theater Joint Intellgience Centers: An Experiment in Applying Structured Methods"; Sherman Kent School Kent Center for Analytic Tradecraft, "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis"; George, "Fixing the Problem of Analytical Mind-Sets: Alternative Analysis"; Office of the Director of National Intelligence, "Intelligence Community Directive 203: Analytic Standards," 2007; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2010.

[3] Coulthart, "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency"; Folker Jr, "Intelligence Analysis in Theater Joint Intellgience Centers: An Experiment in Applying Structured Methods"; Khalsa, "The Intelligence Community Debate over Intuition versus Structured Technique: Implications for Improving Intelligence Warning."

[4] Khalsa, "The Intelligence Community Debate over Intuition versus Structured Technique: Implications for Improving Intelligence Warning."

[5] Fishbein and Treverton, "Making Sense of Transnational Threaths."

[6] See Chapter 4

[7] Coulthart, "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency."

[8] Mandel, Karvetski, and Dhami, "Boosting Intelligence Analysts' Judgment Accuracy: What Works, What Fails?"; Friedman, *War and Chance: Assessing Uncertainty in International Politics*, 32; Rieber, "Intelligence Analysis and Judgmental Calibration."

Wirtz: ‘Today’s interest in devising measures of effectiveness for intelligence stems from America’s pathological fascination with metrics.’[9] An example of this is Shrage's Washington Post comment “What Percent Is 'Slam Dunk '? Give Us Odds on Those Estimates.”[10] The call for measuring accuracy is also noticeable among several academics, although arguably also having a North-American bias.[11] There is little research available describing any preference for probability estimates from European policymakers, although one article indicates that Norwegian military commanders would prefer, if time permitted, to be presented with a more refined picture of possible future scenarios and their respective estimated uncertainties.[12]

Accuracy of judgements is, however, not a key objective of the post-9/11 intelligence reform. When the IRTPA describes the Director of National Intelligence’s responsibility ‘to ensure the most accurate analysis of intelligence’, it describes implementation of procedures of sound analytic methods that maximises analysis of all available data, enabling competitive analysis that allows for differences in analytic judgements.[13] The subsequent official analytic standards describe that accuracy may only be established retrospectively and that key assumptions, information gaps and indicators for change are more important.[14] Although accuracy is related to reducing ambiguity, analysts should not avoid ‘difficult judgements in order to minimize the risk of being wrong.’[15] The original CIA Tradecraft Primer furthermore stated that SATs alone are ‘no guarantee of analytic precision or accuracy of judgments’, but rather that using SATs ‘improve the sophistication and credibility of intelligence assessments’, making the intelligence more useful for policymakers.[16] Consequently, the

---

[9] James Wirtz, in Treverton et al., *Towar. a Theory Intell. Work. Rep.*, 26.
[10] Schrage, “What Percent Is ’Slam Dunk ’? Give Us Odds on Those Estimates.”
[11] Rieber, “Intelligence Analysis and Judgmental Calibration”; Mandel and Barnes, “Accuracy of Forecasts in Strategic Intelligence”; Chang and Tetlock, “Rethinking the Training of Intelligence Analysts”; Chang, Chen, and Mellers, “Developing Expert Political Judgment: The Impact of Training and Practice on Judgmental Accuracy in Geopolitical Forecasting Tournaments”; Dhami and Mandel, “Words or Numbers? Communicating Probability in Intelligence Analysis.”
[12] M. Isaksen and McNaught, “Uncertainty Handling in Estimative Intelligence – Challenges and Requirements from Both Analyst and Consumer Perspectives.”
[13] United States Congress, Intelligence reform and terrorism prevention act of 2004, secs. 1011, 102A h.
[14] Office of the Director of National Intelligence, “Intelligence Community Directive 203: Analytic Standards,” 2007.
[15] Office of the Director of National Intelligence, “Intelligence Community Directive 203: Analytic Standards,” 2015, 4.
[16] Sherman Kent School Kent Center for Analytic Tradecraft, “A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis.”

objective of using SATs has never been accurate predictions but more precisely an integral use of different techniques to advance the analysis in situations of change and uncertainty.

Since estimative intelligence analysis is notoriously uncertain, given its future-oriented scope, accuracy can be hard to prove because it presupposes access to a reality that we not usually have access to. Many of the accuracy proponents tend to forget that intelligence estimates are prone to change when the assessments impact policy-making processes, in many cases invalidating any predictions made. One can therefore argue that the accuracy debate is part of a larger and rather pointless art versus science debate in intelligence theory.[17] Although some research on SATs has addressed that different SATs serve different purposes, the research has not taken the nature of different intelligence problems into consideration. Measuring accuracy of assessments can theoretically be achievable when dealing with intelligence problems classified as secrets, given the intelligence agency has access to the necessary information and that the analysts are able to take denial and deception efforts into consideration.[18] For mysteries and complexities, however, one single truth just waiting to be unveiled does not exist. These forecasts are uncertain estimates about what may happen in the future where absolute foreknowledge is impossible. These predictions can also change due to policymakers' actions, or due to changes in driving forces, altering the course of events. Instead, the best one can aim for is searching for the 'best truth' given the circumstances and assumptions of the future.[19] Estimative intelligence analysis is therefore, just as social science, based on inferences where the conclusion is imperfect, but can be rigorous with correct use of methodology.[20] A consequence is that accuracy in intelligence forecasts is just as unachievable as being able to accurately predict the weather several months ahead or how new variants of a virus will impact a pandemic.[21] Long-term estimates are no better than the data available, the assumptions made, and the models used. As such, researching the effect of using SATs on analysts' ability to assess

[17] Tang, "How Do We Know? What Intelligence Analysis Can Learn from the Sociology of Science."
[18] See Chapter 2 for the discussion on the taxonomy of intelligence problem types. Additionally, ACH was first designed to mitigate denial and deception, see Heuer, *Psychology of Intelligence Analysis*.
[19] Berkowitz and Goodman, *Best Truth: Intelligence in the Information Age*, ix.
[20] Phythian, "Intelligence Analysis and Social Science Methods: Exploring the Potential for and Possible Limits of Mutual Learning."
[21] Hammer, "How Accurate Are Weather Forecasts?" NIHP, "Coronavirus modelling at the NIPH", https://www.fhi.no/en/id/infectious-diseases/coronavirus/coronavirus-modelling-at-the-niph-fhi/

uncertainty, as described in Chapter 5, is arguably more valuable for making estimative intelligence more useful for policymakers and to avoid discourse failure.

The main critique of SATs is nevertheless that there is a lack of empirical proof of which techniques improve analytic performance, especially that SATs do not reduce biases as promised.[22] Some of this criticism make theoretically valid claims, like Chang et. al. arguing that SATs have a problem with inconsistent handling of information, leading to bias bipolarity and noise neglect.[23] They furthermore state that:

> Ultimately, SATs rely on subjective analytic inputs (which extends even to interpreting relatively objective scientific-signature and remote-sensing data). Without well-defined rules that reliably improve interpretation of inputs and sharpen analytic outputs, SATs may serve only as a vehicle transporting subjectivity from one end of the process to the other. The SAT process becomes an end in itself, dressing up subjective judgments in a cloak of objectivity.[24]

There is a concern with this criticism that needs to be addressed. Of the little research conducted on SATs, the vast majority has only tested one technique: Analysis of Competing Hypotheses (ACH).[25] ACH is the "granddaddy" of SATs and is specifically designed to lay emphasis on the principle of falsification. When using ACH as described by Heuer, analysts are to brainstorm a set of hypotheses and make a list of evidence for and against each hypothesis, including the absence of evidence. After having put hypotheses and evidence in a matrix, each piece of evidence is analysed with regards to its diagnostic value towards each hypothesis. Evidence that is not inconsistent with one or more hypotheses has no diagnostic value and should be deleted. Thereafter, by focusing on refutation instead of confirmation, the most likely hypothesis should be the one with the least amount of inconsistent evidence.

---

[22] Chang et al., "Restructuring Structured Analytic Techniques in Intelligence"; Chang and Tetlock, "Rethinking the Training of Intelligence Analysts"; Artner, Girven, and Bruce, "Assessing the Value of Structured Analytic Techniques in the U.S. Intelligence Community"; Dhami et al., "Improving Intelligence Analysis With Decision Science"; Dhami, Belton, and Careless, "Critical Review of Analytic Techniques."
[23] Chang et al., "Restructuring Structured Analytic Techniques in Intelligence."
[24] Chang et al., 344.
[25] Convertino et al., "The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis"; Kretz, Simpson, and Graham, "A Game-Based Experimental Protocol for Identifying and Overcoming Judgment Biases in Forensic Decision Analysis"; Kretz and Granderson, "An Interdisciplinary Approach to Studying and Improving Terrorism Analysis"; Folker Jr, "Intelligence Analysis in Theater Joint Intellgience Centers: An Experiment in Applying Structured Methods"; Brasfield, "Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses"; Lehner et al., "Confirmation Bias in Complex Analyses"; Dhami, Belton, and Mandel, "The 'Analysis of Competing Hypotheses' in Intelligence Analysis"; Whitesmith, "The Efficacy of ACH in Mitigating Serial Position Effects and Confirmation Bias in an Intelligence Analysis Scenario."

The tentative conclusion can, however, rely on a few critical pieces of evidence, and a critical review of the arguments and evidence should be conducted to identify any omitted factors.[26] ACH has much in common with the hypothetico-deductive scientific method. At the same time, while the original scientific method relies on deductive reasoning, resulting in true or false results, ACH works better when using abductive reasoning, especially since most indicators in intelligence are of derivative and not causal nature.[27] ACH usage in intelligence analysis has consequently more in common with how diagnostics is done within the field of medicine.[28]

There are several potential pitfalls when using ACH where the technique's reliability and validity can be questioned. Common mistakes are not having mutually exclusive hypotheses, not identified all possible hypotheses, having an unbalanced set of evidence, diminishing returns, and not taking definitive evidence into account.[29] One could therefore argue that using ACH is pseudo-science, but several of these pitfalls are linked to either dogmatic or naive falsification.[30] Heuer himself argued that ACH is a dual process of both art and science and 'requires BOTH expert judgement and analytic methodology.'[31] Heuer's point is not only important to have in mind when assessing the value of ACH, but an equally important aspect of the value of SATs in general.

Despite having been repeatedly tested, the research results on ACH are not conclusive, having for different reasons yielded mixed results. Folker found that ACH helped analysts reaching the right conclusion only when analysing a simple scenario, not a complex one.[32] Lehner et al. found ACH significantly reducing confirmation bias, but only for inexperienced participants, perhaps because the experienced participants were less biased to begin with. Furthermore, they found ACH did not reduce anchoring bias, whereby concluding that ACH is

[26] Heuer, *Psychology of Intelligence Analysis*, chap. 8.
[27] Pope and Jøsang, "Analysis of Competing Hypotheses Using Subjective Logic," 8-10.
[28] Walter, "The Question of Judgment: Intelligence and Medicine,"; Marrin and Clemente, "Improving Intelligence Analysis by Looking to the Medical Profession."
[29] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020, 165–67.
[30] See Chapter 4 for discussion.
[31] Richards Heuer in van Gelder, Heuer, and Marrin, "Van Gelder's 'How to Free Your Mind' Re ACH."
[32] Folker Jr, "Intelligence Analysis in Theater Joint Intellgience Centers: An Experiment in Applying Structured Methods."

not a robust debiasing method.[33] Convertino et al. also got a mixed result, individuals in the heterogenous group debiased strongly while individuals in the homogenous group did not.[34] Dhami, Belton & Mandel found that ACH made analysts better at evidence diagnostics, but not any better at hypothesis generation or correctly choosing the most likely hypothesis.[35] Both Kretz & Granderson and Whitesmith did not find ACH to prevent confirmation bias.[36] Brasfield, on the other hand, found that ACH mitigated confirmation bias in addition to increase the amount of evidence used.[37] As this was research for a master's thesis, this result nevertheless carries less weight. Lastly, although not measuring bias, Karvetski et al. found ACH to be ineffective at handling correlated evidence, thereby decreasing the coherence of probability judgments.[38]

Although the current research points towards ACH being unfit for improving intelligence analysis, there are several shortcomings of that research that need to be addressed. Several of the studies have had few participants.[39] Many studies also had solely participants without a background from intelligence analysis.[40] The most profound shortcoming is nevertheless how ACH has been tested. Some studies have provided the participant with pre-defined hypotheses, others have made the participants identify the hypotheses themselves. Moreover, most studies have used simple scenarios, assessing past events like the reason for the USS Iowa explosion or whether Iraq has weapons of mass destruction (WMDs) or not,

---

[33] Lehner et al., "Confirmation Bias in Complex Analyses."
[34] Convertino et al., "The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis."
[35] Dhami, Belton, and Mandel, "The 'Analysis of Competing Hypotheses' in Intelligence Analysis."
[36] Kretz and Granderson, "An Interdisciplinary Approach to Studying and Improving Terrorism Analysis"; Whitesmith, "The Efficacy of ACH in Mitigating Serial Position Effects and Confirmation Bias in an Intelligence Analysis Scenario."
[37] Brasfield, "Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses."
[38] Karvetski et al., "Structuring and Analyzing Competing Hypotheses with {Bayesian} Networks for Intelligence Analysis."
[39] Lehner et al., "Confirmation Bias in Complex Analyses"; Kretz, Simpson, and Graham, "A Game-Based Experimental Protocol for Identifying and Overcoming Judgment Biases in Forensic Decision Analysis"; Kretz and Granderson, "An Interdisciplinary Approach to Studying and Improving Terrorism Analysis"; Convertino et al., "The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis."
[40] Whitesmith, "The Efficacy of ACH in Mitigating Serial Position Effects and Confirmation Bias in an Intelligence Analysis Scenario"; Kretz, Simpson, and Graham, "A Game-Based Experimental Protocol for Identifying and Overcoming Judgment Biases in Forensic Decision Analysis"; Kretz and Granderson, "An Interdisciplinary Approach to Studying and Improving Terrorism Analysis"; Brasfield, "Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses"; Karvetski, Mandel, and Irwin, "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation."

problems typically labelled as secret within the field of intelligence.[41] Whitesmith had two pre-defined binary hypotheses (yes/no) on the existence of WMDs and only six intelligence reports, making it difficult to know whether the participants used the cognitive system 1 or system 2 regardless of using ACH or not.[42] Dahmi, Belton & Mandel used only 12 pieces of evidence in a scenario where ACH was not a very suitable technique, assessing tribal membership based on interviews.[43] Karvetski et al. made the participant assess the toxicity of mushrooms, using 3 hypotheses and 22 pieces of evidence.[44] Brasfield and Folker were the only experiments involving predictions or forecasts, while Kretz & Granderson was the only study where the participants were given the evidence in phases.[45] As a general rule, the experiments involving intelligence professionals have used simple scenarios with few hypotheses and few reports. Those experiments using complex intelligence related scenarios often have had small sample sizes of unskilled non-analysts. The main problem seems to be that much of this research has suffered from a type III error by having asked the wrong questions to begin with. By having focused on ACH as a stand-alone technique instead of researching the effect of using ACH as part of the intelligence process, the current research has lacked procedural understanding. Summing up, since none of the experiments have used the same methods, scenarios, amount and type of evidence, or type of participants, both the reliability and validity of the ACH research to this date can be questioned.

When little research has been conducted within a field, taking a wider perspective can also be helpful. The principles behind ACH are not new. In 1939-40, the French intelligence agency, *Deuxième Bureau*, used a scientific method based upon on obtaining and judging

---

[41] Dhami, Belton, and Mandel, "The 'Analysis of Competing Hypotheses' in Intelligence Analysis"; Whitesmith, "The Efficacy of ACH in Mitigating Serial Position Effects and Confirmation Bias in an Intelligence Analysis Scenario"; Lehner et al., "Confirmation Bias in Complex Analyses"; Karvetski, Mandel, and Irwin, "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation"; Convertino et al., "The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis."

[42] Whitesmith, "The Efficacy of ACH in Mitigating Serial Position Effects and Confirmation Bias in an Intelligence Analysis Scenario"; Miller, "The Magical Number Seven, Plus Minus Two: Some Limits on Our Capacity for Processing Information"; Halford et al., "How Many Variables Can Humans Process?"

[43] Dhami, Belton, and Mandel, "The 'Analysis of Competing Hypotheses' in Intelligence Analysis."

[44] Karvetski, Mandel, and Irwin, "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation."

[45] Brasfield, "Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses"; Folker Jr, "Intelligence Analysis in Theater Joint Intellgience Centers: An Experiment in Applying Structured Methods"; Kretz and Granderson, "An Interdisciplinary Approach to Studying and Improving Terrorism Analysis."

reporting against a series of hypotheses, an approach resembling today's ACH.[46] Hence, instead of asking if ACH mitigates bias or improves accuracy, one should rather ask 'For what situations (if any) and what types of people (if any) and under which conditions (if any) does a particular approach to ACH improve analysts' expert judgments?'[47] The same approach can be taken towards other SATs or even SATs in general. Coulthart, for instance, found through an evidence-based evaluation of the techniques described in CIA's "Tradecraft Primer" that there are credible reasons to believe that especially Devil's Advocacy, ACH, and structured brainstorming improves analysis.[48] Although not subject to peer-review or ethical approval, a previous Norwegian survey found that the vast majority of the participants believed SATs improved the quality of their analyses.[49] There are consequently some indications that SATs have more than face value, but more research is needed.

A different take on researching the effects of SATs

The call to investigate the effects of Structured Analytic Techniques (SATs) has not only come from critical voices, also key proponents of SATs have been equally keen on finding out to what degree the use of SATs improves the intelligence process. Heuer and Pherson dedicated a whole chapter in the first edition of *Structured Analytic Techniques for Intelligence Analysis* to this.[50] The previous section has showed that two principal approaches have typically been used to test SATs, logical reasoning and empirical research. Both approaches have limitations, especially when either the participants or the conditions, or sometimes both, are different from real-life ones. Moreover, given the issues with previous research, intelligence analysis theory will benefit from researching the best way to apply relevant scientific concepts to a non-scientific world rather than investigating effects of unrealistic goals. Heuer and Pherson therefore proposed that a broader approach should be used, combining structured interviews, observations, surveys, and experiments under life-like conditions.[51] By extension, researchers therefore also need to recognize that there is

[46] Strong, *Intelligence at the Top: The Recollections of an Intellgience Officer*, 57–58.
[47] National Research Council, *Field Evaluation in the Intelligence and Counterintelligence Context: Workshop Summary*, 19.
[48] Coulthart, "An Evidence-Based Evaluation of 12 Core Structured Analytic Techniques."
[49] Stenslie, "The Necessity of Experts."
[50] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2010, chap. 13.
[51] Heuer and Pherson, chap. 13.

not a single, universal approach to SATs and that an evaluation must include how SATs are taught and used. Unlike previous research, this project has set out to investigate the effect of SATs in a condition that is similar to the intelligence analysis process analysts face on a daily basis. Hence, instead of a focus on a single technique, the research is based upon the comprehensive and iterative NORDIS SATs-methodology, a very different approach to SATs than how SATs are taught by the United States Defence Intelligence or the United Kingdom Professional Head of Intelligence Analysis (PHIA).[52] Moreover, the NORDIS SATs-course focuses on learning to use the techniques together in small analytical teams through a simulation exercise. The scenario using for the SATs-course concerns the piracy situation at the Horn of Africa and the Coast of Somalia in 2010 and includes 41 pages of background material together with a total of 115 reports, where 60 of these were issued during the last two days of the simulation exercise of the course. The other reports were issued as inputs for the analytical process when learning the different techniques.

The choice of research methodology to use when looking into the effects of using SATs in estimative intelligence analysis can depend upon the type of effect one is interested in evaluating. If accuracy of judgements is the key measurement, a quantitative approach is advisable.[53] The same arguably goes for measuring the effect on bias mitigation. However, rather than measuring the accuracy or the bias-mitigating effect of using SATs this research has explored analysts' experiences and opinions, i.e. the perceived effects, of using SATs on analytic rigour and ability to assess uncertainty. Furthermore, since the body of theory regarding the value and effect of using SATs is extremely limited, this research has for the large part been conducted as an exploration for discovery, the reason for pursuing an overall exploratory approach. The project has therefore evolved from a narrow focus investigating the effect the NORDIS SATs-methodology has on analytic confidence to a broader investigation of analysts' use of SATs and their experiences of how SATs affect the estimative intelligence process. The wider focus was nevertheless narrowed down to the key effects called for in the Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004, analytic

[52] See Chapter 6 for details.
[53] Mandel, Barnes, and Richards, "A Quantitative Assessment of the Quality of Strategic Intelligence Forecasts"; Mandel and Barnes, "Accuracy of Forecasts in Strategic Intelligence"; Marchio, "'How Good Is Your Batting Average?' Early IC Efforts To Assess the Accuracy of Estimates."

rigour, as in analytic objectivity and analytic integrity, and the effect on assessing uncertainty, as in statements of both probability and analytic confidence. This approach has led to one main research question (RQ) and three sub questions:

Main RQ: What are the experienced effects of using a comprehensive Structured Analytic Techniques (SATs) methodology in estimative intelligence?
Sub RQ 1: What is the effect of SATs on analysts' ability to assess uncertainty?
Sub RQ 2: What is the effect of SATs on analytic rigour?
Sub RQ 3: What is the main impact on analysts' usage of SATs?

Both quantitative and qualitative methods can be used answering the research questions and both have their strengths and weaknesses. Quantitative methods are very well suited to find statistical data explaining the what, where, when and who, and can find unbiased relationships between key variables like education or experience, and the experienced value of using SAT. Qualitative methods, on the other hand, are better suited when addressing why and how, such as looking at norms, beliefs and attitudes among individuals or a group of people, all relevant factors for finding the underlying reasons for the perceived values of using SATs in an intelligence agency. The direction of the research questions suggests a qualitative approach more than a quantitative one. However, more and more researchers have concluded that a research design does not need to be an either/or approach to the research questions. Instead, 'methods should be mixed in a way that has complementary strengths and nonoverlapping weaknesses.'[54] Therefore, combining a quantitative approach with a qualitative one will often provide an opportunity to assess the research questions from different angles. Moreover, a recent key principle of mixed methods research is one of methodological eclecticism, meaning that the researcher becomes a *connoisseur of methods*, selecting the most suitable methods to answer the research questions as the study evolves over time.[55] As a consequence, the research questions in this thesis have been better answered using a mixed methods research approach.

[54] Johnson and Turner, "Data Collection Strategies in Mixed Methods Research," 199.
[55] Teddlie and Tashakkori, "Overview of Contemporary Issues in Mixed Methods Research."

The data collection and subsequent analysis has for the above-mentioned reasons been a mix of relevant quantitative and qualitative techniques aimed at developing theory from the data. Quantitative data was collected from intelligence personnel on two occasions. Survey 1 was administered on the first and the last day of the two-week long intelligence course in the use of the NORDIS SATs-methodology.[56] A subsequent cross-sectional survey (survey 2) was issued to analysts in three different Norwegian intelligence units.[57] Following the analysis of the survey data, qualitative data has been collected through semi-structured interviews with selected personnel from the same three intelligence units as well as with representatives of institutions delivering SATs-courses in Norway, the Netherlands, and the United Kingdom.

Both surveys were piloted on a representative group of intelligence professionals, survey 1 on intelligence professionals attending the NORDIS SATs-course and survey 2 on intelligence professionals attending an advanced intelligence analysis course hosted by NORDIS. Both surveys were thus piloted on a group of people equal to the respondent sample of the two surveys and the results of the pilots showed that a full-scale survey would be worthwhile. After piloting survey 1, a decision was made to use a 5-point Likert scale instead of a 7-point one. Although there are arguments for using a larger response scale, research has found that using a scale with more than 5 points yields little added value.[58] Moreover, more recent research has also found that using a smaller range of meaningful response alternatives results in reduced response heuristics and leads to greater validity.[59] Furthermore, for both surveys, minor changes were made to the wording of the different statements to reduce ambiguity for the respondents. Lastly, additional four statements were added to survey 2 for better granularity of the responses.

Respondents in survey 2 and the semi-structured interviews all work or recently have worked as intelligence analysts and therefore deal with intelligence analysis every day, making the results very relevant for the intelligence community. In survey 1, however, the

[56] Appendix 2
[57] Appendix 3
[58] Lissitz and Green, “Effect of the Number of Scale Points on Reliability: A Monte Carlo Approach.”
[59] Weathers, Sharma, and Niedrich, “The Impact of the Number of Scale Points, Dispositional Factors, and the Status Quo Decision Heuristic on Scale Reliability and Response Accuracy.”

respondents' job descriptions are more diverse, where a few respondents only take part in multi-discipline/all-source intelligence analysis on an intermittent basis. This difference in populations could be argued to cause a problem of noncomparable successive samples. Nevertheless, due to the survey 1 participant's professional background from either the military or the national intelligence community, the survey 1 results have far more relevance for comparison with the survey 2 results than what identical research on a general population or university students aiming for a degree in intelligence and/or security would have had. Consequently, all participants in this research can be said to be representative of the Norwegian intelligence community population, making it possible to generalize the findings.

Doing research as part of a PhD leads one onto a journey involving some known but perhaps even more emerging limitations and shortcomings. A known limitation was that conducting research on a population consisting of intelligence professionals required a strict obedience to certain ethical and security regulations to receive approval, not only from Brunel University, but also from the Norwegian Intelligence Service and from the Norwegian Defence Intelligence School, an approval that was granted from all relevant authorities before commencing the data gathering.[60] As a part of this approval, security regulations based upon the Norwegian National Security Act of 2018 (*Lov om nasjonal sikkerhet*) implied a need for full anonymity of the participants, also towards the researcher. This made a longitudinal study impossible for this research. The security regulations also prohibited the disclosure of the total attendance of the SATs-courses as well as how many were asked to participate in survey 2. Consequently, reporting a response rate would equally be a breach of Norwegian security regulations. What constitutes an adequate response rate is debated. For many years, a high response rate was a measurement of survey accuracy where Rea and Parker argue that a response rate below 50% stand the chance of being biased.[61] Others argue there is not necessarily a correlation between high response rates and survey accuracy, and that demographic representativeness is more important.[62] By using Dillman's

[60] Appendix 1
[61] Rea and Parker, *Designing and Conducting Survey Research : A Comprehensive Guide*, 196.
[62] Holbrook, Krosnick, and Pfent, "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms."

formula for calculating adequate sample sizes, the sample size for both surveys has nevertheless been assessed to be sufficient to generalize the results to the population of Norwegian intelligence analysts.[63] Moreover, by running the surveys on Norwegian intelligence professionals from all Norwegian intelligence units, demographic representativeness was ensured.

Of the emerging shortcomings discovered as part of the research journey, three stand out as the most important ones. First, the research project started out with a narrow view, focusing on the effect using SATs have on analytic confidence alone. During the research, it was realized that the implications of the research were bigger, addressing more aspects of the tradecraft reform following the IRTPA, and therefore moved from analytic confidence alone to SATs' effect on analytic rigour and analysts' ability to assess uncertainty. The research did collect data on factors impacting analytic confidence as well, but the data analysis revealed that most of this data had no direct impact on the research questions. The specific findings on other factors impacting analytic confidence are still interesting and can be reported in a later article. One result of the initial narrow view is nevertheless that the data on SATs' effect on analytic rigour is missing the pre-course dimension of survey 1, making the findings on analytic rigour less conclusive than the findings on SATs' effect on analysts' ability to assess uncertainty. Second, despite being representative for the same population, there are different participants in the two surveys, making this part of the research a form of successive independent samples study instead of a longitudinal study. Hence, the data analysis of the long-term effects of using SATs have produced indications only since there exists no survey 1 data for the survey 2 population. Third, the two surveys did not contain the exact same items, resulting in minor differences in the two uncertainty components, survey 1 post-course SATs' effect on ability to assess uncertainty and survey 2 SATs' effect on the ability to assess uncertainty. The two components did, however, share enough equally worded items that a comparison of the results would give a very good indication of any similarities or differences of opinion of the effect using SATs have on assessing uncertainty between respondents of the two surveys. Notwithstanding these shortcomings, this research

[63] Dillman, *Mail Internet Surv. Tailored Des. Method, 2nd Ed.*, 206–7.

has produced importantly different results than previous research on SATs, making it the most novel research on the effect of SATs to this date.

Survey methodology

Questionnaire designs and measures

To ensure granularity of the collected data, several demographic variables were recorded in both surveys: gender, age, service background (military/civilian), place of work (unit), education, years of service, and years of intelligence experience. Survey 2 also included variables recording what SATs-course the respondents had attended and how long ago they had conducted the SATs-training. Using IBM SPSS Statistics version 26, the demographic variables were coded either as ordinal or nominal according to the type of data.[64] For instance gender, service and unit were set as nominal data, while education, experience and age were set as ordinal data since the different responses were divided into ordered categories.

In survey 1, the respondents were asked to mark their level of agreement with 15 statements in the pre-course questionnaire and with 18 statements in the post-course questionnaire, using a 5-point Likert scale ranging from agree (5) to disagree (1). In addition, they were asked to rank four factors according to their importance for deciding analytic confidence, where 1 was most important and 4 least important. Finally, participants had the opportunity to fill in additional information in two open-ended questions.

The preliminary analysis of the data from survey 1 was used to develop the items for survey 2, resulting in a total of 18 questions and statements. The statements concerning the perceived effect SATs have on assessing uncertainty and the perceived effect SATs have on analytic rigour were measured using a 5-point Likert scale ranging from agree (5) to disagree (1), while the items measuring to what extent SATs are used, had a 5-point Likert scale ranging from almost always (5) to never (1). In addition, one question asked the participants to mark one or several reasons for not using SATs, and at the end the participants were

[64] Pallant, *SPSS Survival Manual*, 12.

asked to rank four factors according to their importance of impact on analytic confidence, using the same ranking as for survey 1. It was also possible for the respondents to add free-text comments to be used in the qualitative analysis. All items in both surveys were in both English and Norwegian to prevent misunderstandings in meaning.

In several sciences, Likert scale variables have traditionally been treated as ordinal data since it often is difficult to create a scale with numerically uniform increments but rather a scale with only partial semantic autonomy.[65] However, Likert scale data can be treated as interval variable data as long as one can assume that the interval between the different choices on the scale is equal.[66] Respondents typically create equal subdivisions between the different alternatives in the continuum, resulting in quasi-interval variables that can be treated as interval data by applying sequential numerical values to the alternatives.[67] Furthermore, assigning numbers to Likert scale categories and treating them as interval data can often be advantageous to being limited to ordinal techniques and usually only resulting in small errors.[68] There are issues to be aware of by treating Likert scale variables as continuous data, like over-dimensionalisation when conducting exploratory factor analysis.[69] Nonetheless, by showing awareness to the potential pitfalls, this thesis follows a much used tradition in both social science and psychometrics by having analysed Likert scale items in both surveys as interval data. Variables involved ranking, on the other hand, have been analysed as ordinal data.

Data collection procedures

Survey 1 data collection was conducted between August 2017 and January 2019, where personnel from the Norwegian Armed Forces and the Norwegian Intelligence Service attending a two-week long intelligence course in the use of the NORDIS SATs-methodology were asked to participate in a survey investigating the effects of using a comprehensive SATs-methodology in estimative intelligence. The SATs-course participants were on the first

[65] Corbetta, *Social Research: Theory, Methods and Techniques*, 166–67.
[66] Field, *Discovering Statistics Using IBM SPSS Statistics*, 10.
[67] Corbetta, *Social Research: Theory, Methods and Techniques*, 167–68.
[68] Labovitz, “Some Observations on Measurement and Statistics.”
[69] van der Eijk and Rose, “Risky Business: Factor Analysis of Survey Data – Assessing the Probability of Incorrect Dimensionalisation.”

day of the course informed of the scope of the research and were asked to take part by filling out a questionnaire. The content of the Participant Information Sheet (PIS) was explained to them, especially that any participation was voluntary, the right to withdraw, and the confidentiality of the responses ensuring absolute anonymity of the respondents. Moreover, a hard copy of the PIS was distributed together with the pre-course questionnaire. The post-course questionnaire was handed out right after the last lecture on the last day of the course and collected before the participants left the school.

Survey 2 was conducted in January and February 2019, distributed to intelligence analysts in three different units: the Norwegian Intelligence Service (NIS), the Intelligence Battalion in the Norwegian Army (IntBn), and the Norwegian Police Security Service/*Politiets sikkerhetstjeneste* (PST). The participants received a copy of the PIS that was issued together with the survey. The NIS personnel were given the survey electronically, using a survey software already in use on the IT network, while the personnel from the IntBn and the PST were given a paper-copy. Although a longitude study would have been preferable for investigating long-term effects, it was decided that investigating the usage of SATs and the experienced effects of using SATs among full time analysts would produce results more valuable for the intelligence community. This allowed for a higher response rate where analysts that had not been part of survey 1 could participate. Although survey 2 was a separate study *per se*, the survey design made it possible to treat certain elements of surveys 1 and 2 as a successive independent samples study. Furthermore, in survey 2, the initial idea was to use a different population than personnel from the NIS and the IntBn as a control group. However, finding a control group of intelligence professionals without any SATs-training is today virtually impossible. The PST conduct their own intelligence analysis training and although the PST training does include the use of some SATs, especially ACH and Scenario Generation, the PST analysts have not been subject to analysis training using the NORDIS SATs-methodology. As such, the PST was assessed to be a useful comparison although not necessarily function as a control group.

Data analysis procedures

The data analysis aimed to investigate the relationship between the use of SATs and the experienced effect on analytic rigour and ability to assess uncertainty, both the experience in general as well as investigating any differences between demographic variables, with the intent to make a pertinent contribution to the understanding of the value of using SATs in estimative intelligence analysis. The survey data was analysed using IBM SPSS statistical software version 26.

First, preliminary analysis was conducted using the Explore, Codebook, Frequencies, Missing Values Analysis, and reliability analysis functions in SPSS with the aim to gain descriptive statistics and to check for any violations of statistical assumptions to choose the right statistical tests. One thing is that missing data can affect the generalizability of results, although the pattern of missing data is more important than how much data is missing. Missing data can be divided into missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR), where the latter cannot be ignored.[70] When entering the data in SPSS, missing data was coded as 99 and set as missing. Contrary to the survey 1 participants, the respondents in Survey 2 had the opportunity of marking 'Not relevant' on the Likert scale variables. In order not to interfere with the overall quantitative data analysis, 'Not relevant' was coded as 66 and missing. 'Not relevant' was, however, taken into consideration in the descriptive analysis of the data. Another important aspect of the data analysis is to check for the reliability of a scale to check for its internal consistency. A measurement of Cronbach's alpha coefficient is the most used indicator whereas a rule-of-thumb the result should be above .7.[71] Before modifying the items for further analysis, the data was explored using descriptive statistics and graphs to check for any violations of statistical assumptions, something that could bias the results depending on the tests used.[72]

Second, in both surveys Exploratory Factor Analysis was run to explore the relationship between the variables. There are three main usages for EFA; identifying the structure of latent variables for things that cannot be directly measured, questionnaire or survey design, and data reduction by creating fewer variables that retain as much of the information from

[70] Tabachnick and Fidell, *Using Multivariate Statistics: Pearson New International Edition*, 96.
[71] Pallant, *SPSS Survival Manual*, 101.
[72] Field, *Discovering Statistics Using IBM SPSS Statistics*, chap. 5.

the original variables as possible.[73] EFA in general is used to identify structures, or clusters, of variables that correlate highly with each other but at the same time form independent subsets.[74] Two statistical techniques are commonly used, factor analysis (FA) that produces factors and principal component analysis (PCA) that produces components, although most textbooks treat both outputs as factors. There are, nevertheless, differences to the techniques. While FA tries to identify maximum common variance in a correlation matrix through a minimum of factors that "cause" the variables, PCA tries to explain maximum total variance through linear components that are "caused" by the variables.[75] In addition to extraction method, factor rotation method is also widely discussed in the literature. While an oblique rotation method allows for factors to correlate, an orthogonal rotation is suitable for independent factors.[76] Choosing methods to use mainly depends on the wanted outcome and the type of data one is working with. In the end, PCA as extraction method combined with direct oblimin as rotation method was used since the goal was to reduce the number of items to a more manageable set of components to be used in the subsequent analysis rather than a theoretical exploration of underlying factors. Additionally, using survey data gave theoretical grounds to suggest that the components would correlate.[77]

There are several potential pitfalls to EFA that a researcher needs to be aware of, ranging from suitability of the data set for factor analysis to testing the factors for validity. Since almost all data can be factor analysed, EFA can be used to rescue poor research by presenting order out of chaos. Determining the suitability of the data set is a result of several issues, like sample size, missing data, outliers, linearity, multicollinearity and singularity, and the number of sizeable correlations. Normality of the data, however, is not mandatory as EFA can still produce adequate results with non-normal data.[78]

Third, the reliability of the different component scales was tested as well as the relationship between the different components running both parametric and non-parametric correlation

[73] Field, 666.
[74] Tabachnick and Fidell, *Using Multivariate Statistics: Pearson New International Edition*, 660.
[75] Field, *Discovering Statistics Using IBM SPSS Statistics*, 667. See also Tabachnick and Fidell, 662-3
[76] Field, 681.
[77] Field, 676,681.
[78] Tabachnick and Fidell, Using Multivariate Statistics: Pearson New International Edition, 665–68.

tests. Thereafter, a series of relevant parametric or non-parametric tests were employed, dependent upon the degree of normality of the data. These were different analysis of variance tests, like t-tests, one-way between groups, and one-way repeated measures ANOVA with the different non-parametric alternatives (Mann-Whitney U test, Wilcoxon Signed Rank test, Kruskal-Wallis' test, and Friedman's ANOVA). The chi-square test was also employed wherever suitable.

Since there was a predominance of non-normal data, non-parametric, or assumption free, tests would be the preferred option. However, taken into consideration that the scales qualify as both ordinal scales and interval scales, it would be correct to employ parametric tests if the distribution of the data is not outside the border of what would constitute a normal distribution. Non-parametric tests replace the original scores by ranking the data and thereafter analysing the ranking instead of the original scores.[79] In both surveys, using 5-point Likert scales, such transformation would lead to minimal changes as all recorded variables had a range of 4, from 1 to 5. Despite that the non-normal data was on the threshold of normal distribution, the data often violated other assumptions needed for bias-free parametric tests. Consequently, mainly results from the non-parametric tests were reported unless there were differences between the two. Then the results from both tests were reported to check for type I or type II error.

Semi-structured interviews design, measures, and analysis

A limited number of semi-structured interviews were conducted from late 2019 to mid 2020 for a qualitative investigation and follow up of some of the key findings of the two surveys. In addition to the interviews with Norwegian Analysts regarding the use of SATs and the experienced effect on the analytical process, there were discussions in person, through Microsoft Teams, and e-mails, with five SATs instructors, two from the United Kingdom and three from the Netherlands. These conversations focused on the types and taxonomy of techniques taught and how these were taught. Since the topics were not directly comparable with the questions asked the Norwegian analysts, the findings from these conversations have been incorporated in chapter 6 to shed light on the different approaches

[79] Field, *Discovering Statistics Using IBM SPSS Statistics*, 214.

to SATs across a few NATO nations. Although not official interviews, they were asked and gave permission for the use of the essence of the conversation in this thesis as well as any use of quotes.

In-depth interviewing has the advantage of accessing data for social research often not available through questionnaires, such as experiences, interpretations, reflections, and perceptions of the topic being researched.[80] Interviews can also add context and raise issues not previously thought of. Interviews can thereby add the why to the typical what findings from quantitative research methods. Typical weaknesses of using interviews as a research method is that the collected data usually cannot be generalized to a population since the data is based upon personal experience and it is usually difficult to draw a random sample of subjects to interview.[81] Interviews do, nevertheless, produce unique knowledge of the human situation for the interviewees, creating an understanding not otherwise accessible.[82]

There are several stages for conducting semi-structured interviews, from developing an interview guide to analysing and reporting the findings.[83] Being a mixed-methods approach, thematizing and developing an interview guide was started upon together with the questionnaires, although there were some minor changes to the levels of details asked from the initial interview guide to the one used after analysing the questionnaire data. A central part to all these stages is the ethical issues of doing interviews, from obtaining informed consent via ensuring confidentiality, to the integrity of the researcher.[84] A total of 7 intelligence analysts across the three units NIS, IntBn and PST agreed to take part in interviews. The selection of interviewees was to a large degree based upon finding intelligence analysts who could provide rich data to the subject of the effect of using SATs as well as be representative of the population researched.[85] Furthermore, the researcher's personal history and experience has also influence the recruitment of participants, as the interviewees were asked to take part based upon the researcher's knowledge of the past

[80] Morris, *A Practical Introduction to In-Depth Interviewing*, 5.
[81] Morris, 7.
[82] Brinkmann and Kvale, *Doing Interviews*, 10.
[83] Brinkmann and Kvale, 40–41; Morris, *A Practical Introduction to In-Depth Interviewing*.
[84] Brinkmann and Kvale, *Doing Interviews*, 37.
[85] Morris, *A Practical Introduction to In-Depth Interviewing*, 53–54.

and current experience of the interviewees' background as intelligence analysts. A pitfall of such an approach can easily be the challenge of confirmation bias, a problem not easily avoidable. However, the idea of the neutral interviewer has largely been discarded and the dominant view today is that both parties use the interview as a stage for a two-way conversation where both parties express their views.[86] Furthermore, a familiarity of the issue at hand together with the ability to create a good rapport with the interviewees are important issues.[87] In an attempt to mitigate any biases, there was an attempt to recruit interviewees with different backgrounds, such as age, military/civilian and years of experience, although there was a certain homogeneity regarding gender and level of education.

Before commencing the interviews, the interviewees were given the Participant Information Sheet to read and thereafter were asked to sign the consent form. By signing, they agreed to take part in the study at their own free will understanding that their confidentiality was ensured, but also that they could withdraw before the completion of this study. The interviews were from 30 to 45 minutes long, using the interview guide as a basis for how to progress.[88] The sensitivity of their work requested that the interviews had to be conducted at their workplace in case one happened to touch upon classified issues during the interview. Because of limitations in the Norwegian Security Act, recording and later transcribing the interviews was not an option. Although not optimal, the lack of recordings was mitigated by the taking of extensive notes, seeking clarifications when needed, and asking for permission to use specific notable quotes.

There are different modes of analysing the interviews, focusing mainly on meaning or language, doing a bricolage, a more ad hoc approach combining different approaches, or doing interview analysis as a theoretical reading.[89] The mode taken in this thesis has been one of an ad hoc approach, with an emphasis on noting patterns of responses to the thematic subdivision of the interviews and comparing these patterns with the key findings

---

[86] Morris, 91.
[87] Brinkmann and Kvale, *Doing Interviews*, 43.
[88] Appendix 4
[89] Brinkmann and Kvale, *Doing Interviews*, 119–36.

from the two surveys. Lastly, a coherent understanding of the interview data was attempted by noting relations between these and the survey results.

Literature review

When doing research in a specific field of intelligence, such as investigating the effects of using SATs in estimative intelligence, finding relevant and representative literature can be a difficult task. A main reason is that intelligence studies is a rather young academic field where the literature is predominately from the anglosphere. For this literature there has historically been a British tradition of taking a historic viewpoint while the North American literature has more often ventured further into current intelligence practises. Consequently, a large body of the intelligence literature on intelligence analysis stems from North America, although with occasional voices from other parts of the world.

Intelligence theory literature received a boost in 2005 after a workshop on the topic.[90] One of the results has been a steadily growing body of intelligence literature from all over the world. Despite this growth, or perhaps because of it, there is still no unified intelligence theory, but rather a collection of theories.[91] So is also the case for literature on intelligence analysis. In the last decade there has been a growing number of books and articles from countries outside North America, especially European ones. A shortcoming is that some of this literature fall into the trap of just describing current state of play in different areas instead of aiming towards unifying aspects. Marrin points towards weaknesses in this current state of play, especially the lack of new contributions and original ideas, and that there is therefore a need to start looking outside the intelligence studies literature to develop new theoretical perspectives.[92]

As described in the start of this chapter, the research on the validity of using SATs in intelligence analysis has predominately been concentrated towards either bias mitigation or accuracy improvement, an approach that has produced mixed results. The main problem with the current research is that is has taken a limited view, disconnected from the process

---

[90] Treverton et al., *Towar. a Theory Intell. Work. Rep.*
[91] Marrin, "Evaluating Intelligence Theories: Current State of Play."
[92] Marrin.

of intelligence analysis and how this fits into the general intelligence process. To understand the gaps in the literature on intelligence analysis in general, and SATs in particular, one needs to start with a historic perspective. Sherman Kent is often pointed at as the father of intelligence analysis.[93] He certainly pioneered much of the differentiation between basic, current, and estimative intelligence, and his seminal work on the use and misuse of verbal probability is just as relevant today. The literature on verbal probabilities in intelligence has often been combined with literature advocating the use of a more scientific, or structured, approach to intelligence analysis. The result has been a long been a debate about intelligence being an art or a science, both within the intelligence community and in the academic discourse around intelligence theory.[94] Although a specific starting point is hard to find, one of the earliest voices in this debate was Sherman Kent in his renowned work *Strategic Intelligence for American World Policy.*[95] A much lesser-known voice was the British officer Kenneth Strong, who in his memoirs discarded the scientific approach to intelligence analysis in the French Army Staff prior to the outbreak of World War II on the account of being nonsense.[96] Other early voices were Abbot Smith, who was an early advocate for alternative analysis, and David Wark who together with Kent attempted to create a meaningful scale of verbal probabilities to be used in estimative intelligence.[97] The either-or debate of intelligence analysis is nonetheless arguably a false dichotomy, as argued by contemporary authors like Marrin, Gill, and Phytian.[98] This dissertation will enhance this argument by showing that intelligence analysis needs both to increase analytic rigour. To understand why, there is a need to look at all the aspects that constitute a rigorous analytic process, and thereby why analysis failure provides a too narrow point of view as a general theory of intelligence failure.

---

93 Davis, "Sherman Kent and the Profession of Intelligence Analysis."
94 Richards, *The Art and Science of Intelligence Analysis*.
95 Kent, *Strategic Intelligence for American World Policy*.
96 Strong, *Intelligence at the Top: The Recollections of an Intellgience Officer*.
97 Wark, "The Definitions of Some Estimative Expressions"; Kent, "Words of Estimative Probability"; Kent, "A Crucial Estimate Relived"; Smith, "On the Accuracy of National Intelligence Estimates."
98 Marrin, "Is Intelligence Analysis an Art or a Science?"; Marrin, "Analytic Objectivity and Science: Evaluating the US Intelligence Community's Approach to Applied Epistemology"; Gill and Phythian, *Intelligence in an Insecure World*; Phythian, "Intelligence Analysis and Social Science Methods: Exploring the Potential for and Possible Limits of Mutual Learning."

Analysis failure has been the traditional explanation for intelligence failure and Wohlstetter is an early example when she argues that intelligence fails due to signals being buried in 'noise'.[99] Others, like Ben-Israel, Kuhns and Herbert, later write about intelligence analysis' relationship to epistemology and that analysts have to become better thinkers.[100] The cognitive aspects of intelligence analysis became a focus area after the Yom Kippur War, when for instance Bar-Joseph addressed the problem of cognitive closure.[101] The focus spread to the United States, where Heuer was an early proponent for communicating the psychological aspects of intelligence analysis, especially the problems related to heuristics and biases.[102]

One can thus argue that some writers on intelligence analysis have indeed done as Marrin suggests, to look at other fields. The problem is that the current literature on intelligence analysis to a large degree falls in the trap of academic working silos, not addressing overarching principles. An early example is Heuer's description of CIA's attempt to incorporate ideas from the behavioural revolution of political science into intelligence analysis tradecraft.[103] Later, others within the CIA made early attempts to increase analytic rigour through the endorsement of an adapted form of social science methodology, especially as Jack Davis' Alternative Analysis and Douglas MacEachin's Linchpin Analysis became the backbone of changes made to analytic training in the CIA.[104] The outcome of this effort was the start of a number of publications focused on fast and frugal problem solving techniques aimed at increasing critical thinking and reliability of intelligence analysis, where Heuer and Pherson's handbook of Structured Analytic Techniques (SATs) has attained almost a biblical status.[105] Neither of these publications address the linkage between logic, creative

---

[99] Wohlstetter, *Pearl Harbor: Warning and Decision*.
[100] Kuhns, "Intelligence Failures: Forecasting and the Lessons of Epistemology"; Herbert, "The Intelligence Analyst as Epistemologist"; Ben-Israel, "Philosophy and Methodology of Intelligence: The Logic of Estimate Process."
[101] Bar-Joseph, "Intelligence Failure and the Need for Cognitive Closure: The Case of Yom Kippur."
[102] Heuer, *Psychology of Intelligence Analysis*.
[103] Heuer, "Adapting Academic Methods and Models to Governmental Needs."
[104] MacEachin, "The Tradecraft of Analysis: Challenge and Change in the CIA's Directorate of Intelligence"; Davis, "Improving CIA Analytic Performance : Strategic Warning."
[105] Central Intelligence Agency, "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis"; Jones, *The Thinker's Toolkit: Fourteen Powerful Techniques for Problem Solving*; Pherson and Pherson, *Critical Thinking for Strategic Intelligence*; George, "Fixing the Problem of Analytical Mind-Sets: Alternative Analysis"; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2010.

and critical thinking, uncertainty handling, and intelligence analysis methodology in a comprehensive manner. The few publications that make a few of these links, such as "Quick Wins for Busy Analysts" or the Netherlands Defence Intelligence and Security Institute's "Analysis Handbook" are typically training manuals, doctrines, or handbooks with no or limited public disclosure.[106]

There is therefore still more to learn from other disciplines. Gill and Phythian argue that the reasons for intelligence failures are multi-faceted.[107] Hence, the solutions must consider several factors, both inside and outside traditional areas of intelligence theory. Dahl's argument that tactical warning is the solution to mitigate failure is therefore too simplistic. By ciphering in collection failure for analytical shortfalls, his theory leapfrogs the need to cultivate both current and long-term analysis.[108] In comparison, Neuman and Smith have to a larger degree brought the field of intelligence failure theory forward with their new theory of discourse failure.[109] In this theory they address the problem of communicating the possibilities of new and emerging threats. As such, discourse failure interlinks with a central aspect of intelligence analysis theory: the need for alternative analysis, now known as SATs, and SATs' effect on analytic rigour and uncertainty handling.

To avoid falling into the trap of conducting research within academic working silos that lack procedural understanding, this dissertation has investigated several other academic fields for relevant research and literature that affects the process of intelligence analysis. Philosophy, psychology, and decision science are such fields that have relevant input for the process leading to increased analytic rigour. Here it has been necessary to investigate the issues of epistemology and modes of reasoning. Of special interest is how issues that raise concerns can be mitigated, by comparing and contrasting writings on induction, deduction, and abduction with critical rationalism as well as critical and creative thinking. Tversky and Kahneman's seminal research on heuristics and biases has been contrasted to Gigerenzer's

[106] Professional Head of Defence Intelligence Analysis, "Quick Wins for Busy Analysts"; DISI Analysis Instruction Group, *Analysis Handbook: Theory and Methodology in Intellgience Analysis (HB 30-31)*.
[107] Gill and Phythian, *Intelligence in an Insecure World*.
[108] Dahl, *Intelligence and Surprise Attack*; Gustafson, "You Can't Yell Loud Enough: Intelligence and Warning."
[109] Neumann and Smith, "Missing the Plot? Intelligence and Discourse Failure."

idea of our cognitive adaptive toolbox.[110] Given the problem of induction, as stated by Hume, as well as what Hintikka calls the scandal of deduction, Pierce's theory of abduction combined with Popper's critical rationalism is investigated as a more viable way forward for intelligence analysts.[111] But improved logic is only part of the solution, analysts need also to improve what Guilford calls divergent thinking as well as Facione's ideas of critical thinking.[112] When discussing these factors that affect analytic rigour, this dissertation combines issues that previously have been treated separately, thus providing new knowledge to intelligence theory.

The intelligence literature on probability and confidence is also limited, despite Friedman's effort to increase the understanding of the topic with the differentiation between probability and analytic confidence.[113] Moreover, most of the intelligence literature on the topic has had a unilateral and North American focus on improving judgement accuracy.[114] Key concepts are nevertheless still missing, and it has been necessary to look at other academic fields for inputs, especially since uncertainty handling differs between natural and social sciences. Gillies and Hacking nicely sum up the theories of probability, despite that their writing is mostly focusing on the natural sciences.[115] As for communication of uncertainty and confidence, there is a vast literature within psychology, decision science and climate studies, fields that can give valuable input to improve uncertainty handling in intelligence analysis. First, Teigen and Kahneman & Tversky have a different explanation of

---

[110] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases"; Gigerenzer, "The Adaptive Toolbox."

[111] Hume, *An Enquiry Concerning Human Understanding*; Hintikka, *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*; Pierce, "On the Natural Classification of Arguments"; Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*.

[112] Guilford, *The Nature of Human Intelligence*; Facione, "Critical Thinking: A Statement of Expert Consuncus for Purposes of Educational Assessment and Instruction."

[113] Friedman, "Probability and Confidence"; Friedman, *War and Chance: Assessing Uncertainty in International Politics*; Friedman and Zeckhauser, "Analytic Confidence and Political Decision-Making: Theoretical Principles and Experimental Evidence From National Security Professionals."

[114] Mandel and Barnes, "Accuracy of Forecasts in Strategic Intelligence"; Barnes, "Making Intelligence Analysis More Intelligent: Using Numeric Probabilities"; Friedman et al., "The Value of Precision in Probability Assessment: Evidence from a Large-Scale Geopolitical Forecasting Tournament"; Mellers et al., "The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics."; M. Isaksen and McNaught, "Uncertainty Handling in Estimative Intelligence – Challenges and Requirements from Both Analyst and Consumer Perspectives."

[115] Gillies, *Philosophical Theories of Probability*; Hacking, *An Introduction to Probability and Inductive Logic*.

the different variants of uncertainty that is a better fit to intelligence analysis.[116] Second, Budescu, Wallsten, Teigen, and Brun explain how different factors like wording, framing, and context impact on communication and understanding.[117] Moreover, Zlotnick's research on Bayesian probability and Telock's concept of superforecasters are useful as a backdrop into key factors that can affect uncertainty handling in intelligence.[118] Although this dissertation provides a comprehensive picture of these issues, it is still safe to say that the intelligence community has to increase the effort to increase the knowledge of uncertainty handling within in the intelligence domain.

What this thesis does differently than previous literature is to put the relevant bits and pieces together to present a more comprehensive picture of factors that impact a rigorous use of SATs. As a preface, the thesis will present a more multifaceted understanding of intelligence failures, arguing that the latest intelligence analysis reform attempts to address the problem of discourse failure. Mitigating discourse failure in intelligence relies on two factors, an improved understanding of the epistemological and ontological process behind a more rigorous intelligence analysis and an improved understanding of uncertainty, probability, and confidence. This thesis is a first attempt to combine these issues where the aim is to create new knowledge of the factors that impact analytic rigour and uncertainty, and also to present a more comprehensive view on analysts' experienced effects of using SATs. A central point previously not addressed is that there are different approaches to teaching and using SATs, where this research investigates the effect of the NORDIS SATs-methodology, a more comprehensive and iterative approach that fits better to the intelligence process than the typical "single tools in a toolbox" approach of other nations.

[116] Kahneman and Tversky, "Variants of Uncertainty"; Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases."; Fischhoff and MacGregor, "Subjective Confidence in Forecasts."
[117] Olson and Budescu, "Patterns of Preference for Numerical and Verbal Probabilities"; Wallsten and Budescu, "A Review of Human Linguistic Probability Processing - General-Principles and Empirical-Evidence"; Budescu, Karelitz, and Wallsten, "Predicting the Directionality of Probability Words from Their Membership Functions"; Budescu, Broomell, and Por, "Improving Communication of Uncertainty in the Reports of the Intergovernmental Panel on Climate Change."; Teigen and Brun, "Verbal Probabilities: A Question of Frame?"; Løhre and Teigen, "There Is a 60% Probability, but I Am 70% Certain: Communicative Consequences of External and Internal Expressions of Uncertainty"; Teigen and Brun, "The Directionality of Verbal Probability Expressions: Effects on Decisions, Predictions, and Probabilistic Reasoning."
[118] Tetlock and Gardner, *Superforecasting: The Art and Science of Prediction*; Turner et al., "Forecast Aggregation via Recalibration"; Zlotnick, "Bayes' Theorem for Intelligence Analysis."

The methodology used to create this new knowledge has been a mix of quantitative and qualitative methods through research for discovery, which is the reason an overall exploratory approach has been pursued. Social research methodology literature and especially mixed methods literature presented by SAGE has then been central in designing an overall research strategy.[119] Thereafter, literature on survey design has been consulted since a large part of the data is derived from questionnaires.[120] For quantitative data analysis, the two SPSS guidebooks by Field and Pallant have been invaluable.[121] For the qualitative interview design and data analysis, the books by Morris and Brinkman and Kvale have been key.[122]

Summary

Previous research on intelligence analysis theory has for the most part had a limited view, often characterized by academic working silos. Most of the research on SATs is no different where Coulthart is one of few one who has taken a broader approach. The other research has predominately investigated the effect ACH has on either bias mitigation or judgement accuracy. The ACH research has furthermore been conducted with different research designs, particularly using different populations and scenarios, and without procedural understanding of the process of intelligence analysis. Consequently, it can be argued that much of this research suffers from a type III error, having reached statistically correct answers, but by having asked wrong questions. The research conducted for this thesis is more comprehensive, investigating overarching principles and factors that affect analytic rigour and uncertainty handling in estimative intelligence, and thereafter testing to what degree using a comprehensive and iterative SATs-methodology that adhere to these principles produces the wanted effects stated in the 2004 IRTPA. The literature used for this

[119] Corbetta, *Social Research: Theory, Methods and Techniques*; Tashakkori and Teddlie, *SAGE Handbook of Mixed Methods in Social & Behavioral Research*; Plano Clark and Ivankova, “Mixed Methods Research: A Guide to the Field”; Johnson and Turner, “Data Collection Strategies in Mixed Methods Research.”

[120] Lissitz and Green, “Effect of the Number of Scale Points on Reliability: A Monte Carlo Approach”; Weathers, Sharma, and Niedrich, “The Impact of the Number of Scale Points, Dispositional Factors, and the Status Quo Decision Heuristic on Scale Reliability and Response Accuracy”; Dillman, *Mail Internet Surv. Tailored Des. Method, 2nd Ed.*; Rea and Parker, *Designing and Conducting Survey Research : A Comprehensive Guide*; Holbrook, Krosnick, and Pfent, “The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms.”

[121] Field, *Discovering Statistics Using IBM SPSS Statistics*; Pallant, *SPSS Survival Manual*.

[122] Brinkmann and Kvale, *Doing Interviews*.

thesis is therefore also more interdisciplinary, combining key elements from philosophy, psychology, and decision science with intelligence theory. Together, the literature forms the theoretical background for part 2 that discusses intelligence failure, analytic rigour, uncertainty handling, and the development of SATs in different nations. Together, the research of this thesis is the most novel research on SATs to this day.

Part 2: Theoretical aspects of discourse failure, analytic rigour, uncertainty, and the teaching of SATs in estimative intelligence

Intelligence theory can, like many other fields, come across as suffering from academic working silos. Intelligence failure theory has for long been characterised by competing schools where suggested mitigations subsequently are limited to the school in thought.[1] A consequence is that several key aspects related to the tradecraft of intelligence analysis are treated separately, that being epistemology and logic, cognitive limitations, understanding of uncertainty, and their impact on intelligence analysis methodology. Moreover, new terms are sometimes introduced in official reports, like analytic rigour and analytic confidence were in the post-9/11 reform, yet without having any clear theoretical basis. This part of the thesis presents and discusses the current theoretical viewpoints concerning intelligence failure, logic, analytic rigour, uncertainty, and SATs. The different chapters will present the current theoretical standings on the different issues and suggest new and improved ways to understand them in a setting of intelligence theory. The key take-away is that to mitigate causes for intelligence failure, the different aspects must be addressed in a more comprehensive manner since there are interlinked facets impacting on the totality of the process of estimative intelligence, not least the role SATs should play.

[1] Dahl, *Intelligence and Surprise Attack*, chap. 1.

Chapter 2: Post-9/11 intelligence reform and the evolution of a new theory of intelligence failure

Sherman Kent stated already in 1949 that 'Intelligence is bound to make mistakes.'[1] Intelligence failure has since developed to become one of the most researched fields within an ever-growing body of intelligence theory literature. As a result, there are several different theories of what causes intelligence failures, from the classical theory of analysis failure to the new theory of discourse failure in the intelligence-policymaker interface. Where the cause of failure has been placed has to a large degree been mimicked in subsequent intelligence reforms aimed at mitigations. For the topic of estimative intelligence, the post-9/11 reform addresses the core elements of the new theory of discourse failure by attempting to improve how to identify, analyse, and communicate possible current and future threats. Being a community-wide reform across several nations, this reform has been more wide-reaching and has had a larger impact on the intelligence community than previous reforms. Within estimative intelligence analysis, the impact can be observed in the strong focus on improving analytic rigour through improved analytic tradecraft and through attempts to improve the communication of uncertainty to policymakers.

This chapter will first give an overview of the different intelligence failures theories in which the latest, discourse failure, is far more comprehensive and multi-causal than previous theories. The current literature on discourse failure is very limited and key aspects have yet to be addressed. This chapter will give a more nuanced picture, arguing that discourse failure is not limited to the intelligence-policymaker interface but is just as prevalent in the different sub-processes of the intelligence cycle. For the intelligence community, addressing the reasons for discourse failure will therefore have an impact on everything from how to identify that a potential threat may exists to how intelligence assessments about possible threats are presented to policymakers.

[1] Kent, *Strategic Intelligence for American World Policy*, 194.

The second part of the chapter will look at the latest intelligence community reform that largely was a result of American legislature post-9/11. Instead of a strict compartmentalization of remedies, the post-9/11 intelligence reform is with the issuing of new directives and standards, amongst other issues, seemingly also attempting to mitigate discourse failure. The main argument of this chapter is that the new legislation has started a renewed community-wide focus on two enduring topics in estimative intelligence; how to increase analytic rigour through improved analytic tradecraft and how to improve expressions of uncertainty. Both issues are core elements of how to identify, analyse and communicate the existence of possible threats, issues that can be ameliorated with an improved intelligence analysis methodology.

Traditional intelligence failure theories

Reasons for intelligence failures have often been found within the intelligence community. The conventional wisdom put blame on intelligence analysis, especially due to analysts' cognitive limitations, a part of what Dahl calls the traditional school of intelligence failure theory.[2] This school of thought can be traced back to Roberta Wohlstetter's seminal work on the Japanese attack on Pearl Harbor in which she argued that the main reason for intelligence failures and surprise attacks is that analysts tend to look for confirmatory signals and that signals pointing towards alternative explanations are being buried in "noise".[3] Her point of view has since been advocated by many others where perhaps Richards Heuer's *Psychology of Intelligence Analysis* has become one of the most central pieces in intelligence theory on cognitive limitations, a book that lends most of its arguments from research done in the field of heuristics and biases, by people like Amos Tversky, Daniel Kahneman, and others.

Another reason for analysis failure that is closely related to cognitive limitations is analysts' preference for inductive logic.[4] The problem of induction will be discussed in chapter 4, but the core argument is that historic and current events cannot say anything certain about the

---

[2] Dahl, *Intelligence and Surprise Attack*, 9–11.
[3] Wohlstetter, *Pearl Harbor: Warning and Decision*.
[4] Kuhns, "Intelligence Failures: Forecasting and the Lessons of Epistemology"; Hatlebrekke, *The Problem of Secret Intelligence*.

future, there will always be possibility for surprises due to chance and free will. The tradition of inductive reasoning in intelligence analysis suggests that there is only one truth out there ready to be uncovered if the analysts can connect the dots, but that is usually not the case. Clark, amongst others, thus argues that analysts' lack of objectivity in handling collected material is one of three main reasons for intelligence failure.[5]

Advocates of analysis failure also find credence in several official intelligence failure reports. The Agranat Commission found that 'the "concept" ... had become AMAN's conclusion in advance', the Jeremiah Commission found that 'both the intelligence and the policy communities had an underlying mindset going into these tests that the BJP would behave as we behave', and the 9/11 report blamed analysts' lack of imagination.[6] The official reports have also been echoed internal reviews. In 1983, a CIA Senior Review Panel surveyed 12 major intelligence failures between 1945 and 1978 where the key recommendation was a stern warning of the hazards of single outcome forecasting.[7] Subsequent intelligence reforms have therefore to a large degree been aimed at improving intelligence analysis methodology, although the results are debatable.

While it can be difficult to alter human cognition, it is possible to regulate intelligence organisations through reforms and legislation. In this reformist school, intelligence failures are explained by bureaucratic and organizational issues, especially unnecessary secrecy and lack of coordination and intelligence sharing.[8] Wilensky attributes this view to the argument that '…intelligence failures are built into complex organizations.'[9] Historically, most intelligence organisations have either separate collection services or organised themselves vertically for different collection disciplines. From a shared knowledge point of view, such compartmentalization arguably results in communication failure, ineffective collection, and diffusion of responsibility.

---

[5] Clark, *Intelligence Analysis: A Target-Centric Approach*, 4–5.
[6] National Commission on Terrorist Attacks Upon the United States, "The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States"; Bar-Joseph, "Israel's 1973 Intelligence Failure." See also 'Jeremiah News Conference', 2 June 1998, https://www.cia.gov/news-information/press-releases-statements/press-release-archive-1998/jeremiah.html, accessed 8 October 2020.
[7] CIA Senior Review Panel, "Report on a Study of Intelligence Judgments Preceding Significant Historical Failures: The Hazards of Single-Outcome Forecasting."
[8] Dahl, *Intelligence and Surprise Attack*, 11–13; Clark, *Intelligence Analysis: A Target-Centric Approach*, 3–4.
[9] Wilensky, *Organizational Intelligence: Knowledge and Policy InGovernment and Industry*, 179.

Zegart consequently puts emphasis on organizational adaptation failure as a core reason for intelligence failures, arguing that changes to organizational culture are needed to prevent future failures.[10] Considering organisational theory, it is particularly relevant to pay attention to bureaucratic structures and challenges in an intelligence organisation, the coordination between different departments, and the interaction with other government agencies with or without overlapping functions and roles. Restructuring the intelligence processes could then provide more targeted results.[11] Kindsvater furthermore argues that 'If [policymakers have] an intelligence question, they should be able to do one-stop shopping based on the issues, not based on how intelligence was collected or analysed.'[12] Kinsvater's point may be the result of, for most Europeans, an American peculiarity of the many more or less competing intelligence agencies, whereas in the United Kingdom, policymakers would go to the Joint Intelligence Organisation since the impact of the question usually goes beyond the remit of their ministry or department.

The call for reorganisation echoes in several intelligence post-mortem reviews, like in the Aspin-Brown Committee report, the 9/11 Commission Report, the Butler report, and the American report of The Commission on the Intelligence Capabilities Weapons of Mass Destruction, where the latter recommended 'a stronger and more centralized management of the Intelligence Community, and, in general, the creation of a genuinely integrated community instead of a loose confederation of independent agencies.'[13] Intelligence sharing and closer cooperation can nevertheless be hindered by bureaucratic conditions. The Chilcot inquiry is therefore interesting as it looks at the whole-of-government nature of the failure by principally investigating government decision making and only points at the intelligence community secondarily.[14] Moreover, in many countries, like Norway, there are political and legislative traditions for a sharp distinction between the legal and administrative frameworks for the security services and the intelligence services respectively. Even organisational

---

[10] Zegart, *Spying Blind: The CIA, the FBI, and the Origins of 9/11*, 11–13.
[11] Clark, *Intelligence Analysis: A Target-Centric Approach*, 34–38.
[12] Kindsvater, "The Need to Reorganize the Inteffigence Community," 35.
[13] Silberman and Robb, "The Commission on the Intelligence Capabilities Weapons of Mass Destruction: Report to the President of the United States."
[14] Chilcot, "The Report of the Iraq Inquiry: Executive Summary."

innovations such as joint intelligence centres suffer from lack of revision of the legislation. Consequently, reorganisation may have a limited effect unless other factors play an equally important role in achieving real change.

There is a third school of thought that has a contrarian view of the reason for intelligence failures. Instead of putting blame on either cognitive or organizational factors, the argument is that intelligence failures stem from poor intelligence collection.[15] More specifically, the central issue of collection failure is lack of access to relevant information.[16] In many cases, an intelligence agency will not have access to the information, either due to active denial and/or deception, but more often because the information requirements have not been prioritised.[17] To the extent that the intelligence agency has access, the collected information is often incomplete, inaccurate, or misleading and does not cover what is needed to provide an accurate judgement to the intelligence requirement.[18] The consequence is that both the information and the assessments derived from the collection will have uncertainties attached.

Post-9/11, collection failure has received its fair share of attention, especially the lack of relevant and timely human intelligence.[19] More specifically, it can be argued that collection failure stems from a tendency to collect the information that was available at the time, often based upon old habits, not necessarily the information the analysts have had the most need for in their analyses.[20] It compares to Kaplan's 'principle of the of the drunkard's search', the drunk man looking for his keys under the street lamp because it is lighter there, even though he may have lost them further away.[21] Cogan therefore argues that the intelligence community should hunt for the relevant information, not gather all that is possible.[22] Another issue is the focus on current intelligence at the expense of more long-term analysis, hereunder a lack of a more comprehensive focus on other factors than threat actors alone,

---

[15] Dahl, *Intelligence and Surprise Attack*, 13–14.
[16] Schmitt, "Truth to Power? Rethinking Intelligence Analysis"; Hendrickson, "Critical Thinking in Intelligence Analysis."
[17] Abdalla and Davies, "Intelligence, Policy, and the Mandate: A Third Form of Strategic Failure."
[18] Gill and Phythian, *Intelligence in an Insecure World*, 146–47.
[19] Dahl, *Intelligence and Surprise Attack*, 13–14.
[20] Dahl, 13–14; Gill and Phythian, *Intelligence in an Insecure World*, 146–47.
[21] Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science*, 11.
[22] Cogan, "Hunters Not Gatherers: Intelligence in the Twenty-First Century."

such as information concerning social, political, and cultural issues.[23] More and better information may, however, in itself not remedy failures. As Hilsman points out: 'Facts cannot in themselves contain self-evident or obvious answers.'[24] To achieve better-focused collection, the role of analysts and intelligence managers is just as central as sensors and sources. Although collection failure is a relevant theory, it can in many cases be difficult to prevent as it is often difficult to know beforehand exactly what information will be relevant and there is always a need to prioritise collation due to limited resources.

The abovementioned intelligence failure theories all argue that the main fault for intelligence failures is to be found within the intelligence process. Betts and others, on the other hand, argue that it is rather the policymakers who need to carry the most blame for intelligence failures.[25] The traditional direction of this theory blames the decision-makers for lack of responsiveness to intelligence warnings. Here Betts asserts that there are different qualities needed for accuracy and for influence, and whereas analysts want to be as objective and nuanced as possible, decision-makers in general want clear-cut answers.[26] Policymaker failure is also closely related to the issue of politicization of intelligence, such as direct pressure, following the house line, cherry picking, question framing or having a shared mindset that does not consider alternatives.[27] Lock Johnson follows up on this topic, characterizing decision-makers' disregard of objective intelligence one of the seven sins of strategic intelligence.[28] The issue of what threat Iraq constituted to the United States and the United Kingdom back in 2003 is a case in point. Both countries' post-war inquiries that investigated the respective governments' use of pre-war intelligence found evidence of governmental misuse of intelligence, presenting uncertain judgements as facts to the public to create a pretext to go to war.[29]

---

[23] Lord Butler, "Review of Intelligence on Weapons of Mass Destruction."
[24] Hilsman, "Intelligence and Policy-Making in Foreign Affairs."
[25] Betts, "Analysis, War, and Decision: Why Intelligence Failures Are Inevitable"; Clark, *Intelligence Analysis: A Target-Centric Approach*; Gentry, "Intelligence Failure Reframed."
[26] Betts, "Policy-makers and Intelligence Analysts: Love, Hate or Indifference?"
[27] Gill and Phythian, *Intelligence in an Insecure World*, 153–55.
[28] Johnson, *America's Secret Power: The CIA in a Democratic Society*, 63–64.
[29] Chilcot, "The Report of the Iraq Inquiry: Executive Summary"; United States Senate Select Committee on Intelligence, "Senate Report 110- Report on Whether Public Statements Regarding Iraq by U.S. Government Officials Were Substanciated by Intelligence Information." See also press conference with Sir John Chilcot, 6 July 2016, https://www.bbc.com/news/uk-politics-36721645, and press statement from John Rockefeller, 5

Still, there are those claiming that the intelligence community is to blame for a faulty policymaker-intelligence interface. Davies argues that intelligence failure is 'a failure to provide warning or the provision of a significantly inaccurate assessment.'[30] The Robb-Silberman Commission, for instance, found the pre-war intelligence on Iraq's WMD program to be 'dead wrong in almost all of its pre-war judgments about Iraq's weapons of mass destruction.'[31] Omand has a slightly different take on it, defining intelligence failure as a situation for which an operational warning or tactical warning could be expected, but this notice was omitted.[32] Dahl follows up on this, claiming that intelligence fails due to low policymaker receptivity to strategic intelligence because it seldom is actionable and that what is needed is more precise tactical warning to foil attacks.[33] Dahl's view can find support in the 9/11 Commission's finding that the Clinton and Bush administrations 'were not served properly by the intelligence agencies.'[34] However, it takes two to tango and pointing towards the intelligence community alone for policymaker failure implies policymakers just sit and wait for intelligence to guide policy instead of recognizing that policymakers can take a pro-active part irrespective of what intelligence they receive. Walter Laqueur neatly summarizes a major dilemma for any intelligence service:

> It could be argued that, almost by definition, intelligence is always bound to fail. If it correctly predicts the political or military initiative of another country, and if as a result, countermeasures are taken and the initiative does not take place, it will be blamed for making false predictions.[35]

Instead of the traditionally treatment of intelligence failures as competing theories, it has recently become more common to argue that the reasons of failure are not mutually exclusive but can rather be overlapping.[36] The reasons are often wide-ranging regarding where in the process the failure occurs, but narrow ranging regarding severity when the

---

June 2008, https://www.intelligence.senate.gov/press/senate-intelligence-committee-unveils-final-phase-ii-reports-prewar-iraq-intelligence, accessed 22 October 2020.

[30] Davies, "Intelligence Culture and Intelligence Failure in Britain and the United States," 3.

[31] Silberman and Robb, "The Commission on the Intelligence Capabilities Weapons of Mass Destruction: Report to the President of the United States," 1.

[32] Omand, *Securing the State*, 237.

[33] Dahl, *Intelligence and Surprise Attack*, 21–26.

[34] Kean, Thomas, interview by John Shovelan, ABC Radio, 23 July 2004, http://www.abc.net.au/am/content/2004/s1160100.htm, accessed 22 October 2020

[35] Laquer, *World of Secrets: The Uses and Limits of Intelligence*, 4.

[36] Gill and Phythian, *Intelligence in an Insecure World*, 169.

failure is understood in a context of national importance. Intelligence failure is therefore often not about the fact that a warning has been given, but rather how the warning was interpreted by policymakers. Gentry clarifies such policy failure through his argument that 'intelligence fails if a state does not adequately collect and interpret intelligence information, make sound policy based on the intelligence (and other factors), and effectively act.'[37] The 9/11 Commission Report not only pointed a finger towards the American intelligence community for a failure of imagination, capabilities, and management, the committee also blamed the US Government for policy failure.[38] A similar British viewpoint, from the Chilcot inquiry, was that the Iraq War constituted a governmental strategic failure because '…no substantial change in approach was ever implemented.'[39] The Chilcot Inquiry also provided a more compound picture, blaming the Joint Intelligence Committee for expressing unwarranted certainty in judgements as well as blaming the government for deciding to go to war before all other options had been exhausted.[40] Shulsky and Schmitt emphasise the latter point when arguing that intelligence failure is 'a misunderstanding of a situation that leads a government to take actions that are inappropriate and counterproductive to its own interests.'[41] This idea introduces a relatively new viewpoint, namely that the failure can lie in the communication process that takes place between policymakers and the intelligence community. This theory for intelligence failure has now become more pronounced; discourse failure due certain pathologies in the policymaker-intelligence interface.[42]

Discourse failure – the new theory of intelligence failure

Discourse failure was introduced as a new theory of intelligence failure in the wake of the 9/11 terrorist attack and the intelligence regarding Iraqi weapons of mass destruction (WMD) leading up to the 2003 Iraq War. Instead of putting the blame solely on either the

[37] Gentry, "Intelligence Failure Reframed."
[38] National Commission on Terrorist Attacks Upon the United States, "The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States."
[39] Chilcot, "The Report of the Iraq Inquiry: Executive Summary," 1:109.
[40] Chilcot, "The Report of the Iraq Inquiry: Executive Summary."
[41] Shulsky and Schmitt, *Silent Warfare: Understanding the World of Intelligence, 2nd Ed*, 63.
[42] Neumann and Smith, "Missing the Plot? Intelligence and Discourse Failure"; Hatlebrekke and Smith, "Towards a New Theory of Intelligence Failure? The Impact of Cognitive Closure and Discourse Failure"; Hatlebrekke, *The Problem of Secret Intelligence*.

intelligence community or the policymakers, the argument is that the main cause of failure stems from the intelligence dialogue, the communication process that takes place between the intelligence functions and the policymakers. More precisely, discourse failure is due to psychological, societal, and communicative pathologies within the policymaker - intelligence interface. Neuman and Smith defines discourse failure as a 'failure of comprehension: the constriction of the language and vocabulary to identify, analyze, and accept that a significant threat exists.'[43] They furthermore call discourse failure the missing dimension of intelligence failure and argue that many factors which influence political perceptions are not considered in the traditional intelligence failure theories. Neumann and Smith especially points towards ideological preferences and the ideas we have of the world as the most important issues affecting priorities and resources, leading to mistaken threat perceptions.[44] Hatlebrekke and Smith amplifies that these ideological prisms harm the epistemic process between intelligence producers and policymakers, where especially the need for cognitive closure fuels discourse failure.[45] Hatlebrekke furthermore argues that in addition to the issues of excessive secrecy and the problem of an intelligence tribal language, discourse failure is closely related to the problem of inductive reasoning, and that combined, these elements reduce the level of precision and scope of intelligence products.[46]

One can always argue that discourse failure is not a new theory, but rather that it is an integral part of policymaker failure. In political theory, discourse failure is a result of people's misconceptions and biases of how societies work and the political incentives to further spread vivid and popular explanations.[47] Instead of promoting critical thinking and careful deliberations through the brain's system 2, dissenting views are discouraged, leaving the society with ingrained erroneous theories. Not being limited to policymakers or intelligence alone, but rather with the societal, political, and psychological background, discourse failure offers a more multi-faceted explanation of intelligence failure than the traditional competing theories. Simultaneously, it can be argued that the current literature on discourse failure has

[43] Neumann and Smith, "Missing the Plot? Intelligence and Discourse Failure." 96
[44] Neumann and Smith.
[45] Hatlebrekke and Smith, "Towards a New Theory of Intelligence Failure? The Impact of Cognitive Closure and Discourse Failure." 156-157
[46] Hatlebrekke, *The Problem of Secret Intelligence*, 8.
[47] Pincione and Tesón, *Rational Choice and Democratic Deliberation: A Theory of Discourse Failure*, chap. 2.

had a limited perspective. Neuman and Smith blames discourse failure on the argument that dominant political ideologies in a society limit the scope of intelligence products.[48] This may be true in polarized societies like the United States as in the Republican narrative that the 2020 Presidential Election was fraudulently stolen by the Democrats, but not necessarily so in a more multi-political landscape which is typical for many European countries. Hatlebrekke and Smith limit discourse failure to the argument that analysts' use of inductive logic combined with cognitive closure inhibits imagination of how a threat can change. Instead of being multi-faceted, this argument narrows discourse failure down to a failure of analysis. Pincione and Tesón's theory assumes that all part of society works the same manner, yet one role of intelligence is arguably to speak truth to power. Moreover, since the intelligence process concerns more than the intelligence – policymaker interface alone, the current scope misses a key aspect. Discourse failure does not only affect the intelligence – policymaker interface, it can equally affect the interface between the different functions of the intelligence process. The theory of discourse failure therefore encompasses everything from having asked and prioritised the right question about a potential threat to begin with, via having identified, analysed, and communicated the threat in a clear and comprehensible manner, to how the policymakers have acted when being presented with the intelligence about the threat.

The normative intelligence process starts with a policymaker stating one or several intelligence requirements, which again dictate the direction and scope for the subsequent product. One could therefore logically presume there is a correlation between accurate, timely, and relevant intelligence products and accurate, timely, and relevant intelligence requirements. However, policymakers seldom have the time or knowledge to state the question accurately or nuanced enough to begin with.[49] The variety of intelligence problem types, from secrets to complexities, and how they may interact is not considered. An improper and narrow definition of the Iraq WMD problem set was arguably one of the main problems for the 2002 National Intelligence Estimate.[50] Moreover, it may be that the intelligence requirements are stated too late for an intelligence agency to change focus in

[48] Neumann and Smith, "Missing the Plot? Intelligence and Discourse Failure."
[49] Clark, *Intelligence Analysis: A Target-Centric Approach*, 115–22.
[50] Clark, 426–27.

time to gain access to what would be relevant. Robert Kennedy pointed out: ‘Unfortunately, policymakers frequently don’t seem to know what information they need until they actually need it.’[51] Hence, intelligence agencies are frequently asked to shed light on what just has happened instead of what is likely to happen in the future. Often, this is the result of issues having been given a low priority to start with, only to suddenly come as a surprise in the face of the policymakers, such as in the build-up to the Falklands War.[52] With an imprecise starting point, there is a risk of a framing effect where the initial intelligence requirement is accepted as first presented despite the chance that a reframing of the question would lead to diverse lines of inquiry.[53] A result of this discourse failure is the risk of collection assets being tasked to collect data and information that only to some extent corresponds to what is really needed for the analytical task. As Vandepeer states, ‘Poor communication between a requestor and an analyst risks misunderstanding the problem, wasting resources and delivering irrelevant analysis.’[54] Consequently, we can say that in intelligence, it is not always the customer who is right, but that the intelligence agency must guide the policymakers to formulate precise and relevant intelligence requirements.

Discourse failure is not a result solely based upon faults when stating and prioritising intelligence requirements, but just as well a result from a lack of long-term direction, feedback, and review within an intelligence agency itself. The intelligence communities’ traditions for stove-piping and secrecy have led to failures in sharing information also between the collection and analysis functions. Intelligence analysis also leads to new questions and analysts should therefore have an influence on information requirements prioritized for collection.[55] Without this internal dialogue, analysts may have to replace irrelevant collected information with assumptions to produce intelligence needed for policymaking processes. Analysts can therefore seemingly find themselves at the centre of discourse failure. One the one hand they depend on the collection of relevant data and information where they can provide inputs. On the other hand, analysts can be prone to look

[51] Kennedy, *Of Knowledge and Power: The Complexities of National Intelligence*, 15.
[52] Abdalla and Davies, “Intelligence, Policy, and the Mandate: A Third Form of Strategic Failure.”
[53] Herbert, “The Intelligence Analyst as Epistemologist.”
[54] Vandepeer, *Applied Thinking for Intelligence Analysis: A Guide for Practitioners*, 11.
[55] Gill and Phythian, “From Intelligence Cycle to Web of Intelligence: Complexity and the Conceptualisation of Intelligence,” 27.

for the wrong information due to cognitive limitations and lack of analytic rigour. The consequence is increased risk of producing intelligence estimates with flawed conclusions. The 1983 CIA Senior Review Panel found that the failed intelligence estimates did not present any analysis of possible, albeit less likely alternative developments and outcomes, but were instead 'reinforced [by] some of the worst analytical hazards – status quo bias and a prejudice towards continuity of previous trends.'[56] There had seemingly been little progress for two decades when the 9/11 Commission accused analysts of lacking imagination. The suggested answer, as voiced by the Silberman-Robb Committee, was almost the same as almost 20 years prior, for the intelligence community to recognize 'the importance of fostering a culture of alternative analysis throughout the Intelligence Community.'[57] Such a culture could arguably also foster improved collection of relevant information.

The problem for intelligence is that policymakers have not necessarily been concerned about a lack of alternative analysis. Discourse failures may therefore also appear when policymakers receive intelligence they perceive not to be relevant, timely, or accurate enough, regardless of how the initial intelligence requirement was stated.[58] A reason can be, as Marrin argues, that intelligence estimates are only a supplement to policy assessments and that policymakers will choose their own analysis over intelligence estimates when the two differ.[59] Furthermore, lack of relevance can often be the result of policymakers being presented with intelligence estimates about a subject that only in retrospect proved to be important. Lack of timeliness can either be a result of the intelligence product having arrived before the policymaker has acknowledged that the issue is a problem that needs to be addressed; or that the intelligence reaches policymakers too late, at a time when what they really want is an intelligence product that supports an already chosen course of action. Intelligence can thus have more impact on policymakers when an issue has gotten their attention, but before they have developed a strong sense of what should be done.[60] Lack of

---

[56] CIA Senior Review Panel, "Report on a Study of Intelligence Judgments Preceding Significant Historical Failures: The Hazards of Single-Outcome Forecasting."
[57] Silberman and Robb, "The Commission on the Intelligence Capabilities Weapons of Mass Destruction: Report to the President of the United States.", 406.
[58] Dahl, *Intelligence and Surprise Attack*, 16–24; Gill and Phythian, *Intelligence in an Insecure World*, 153–56.
[59] Marrin, "Why Strategic Intelligence Analysis Has Limited Influence on American Foreign Policy."
[60] Gardiner, "Squaring the Circle: Dealing with Intelligence-policy Breakdowns," 146–47.

accuracy is generally a result of the policymaker perceiving the product to be ambiguous or vague and therefore not precise enough for action.[61] An aspect of this accuracy debate is how uncertainties and probabilities in key judgements are being conveyed to the recipients. A relevant Norwegian example is a 2014 terror threat warning issued by the Norwegian Security Police (PST). In the warning the PST stated that within the next 12 months, it was likely Islamic terrorist would attempt to target Norwegian police, military or state officials.[62] When the press started to investigate what likely constituted, they reported that there was up to a 90% probability for a terrorist attack in Norway.[63] The subsequent media debate was centred around the use of vague expressions that did not function as proper decision support. Consequently, how the intelligence communities report back on the policymakers' intelligence requirements have great impact on policymaker responsiveness and what actions they take, making communication of uncertainties in key judgements a central element of discourse failure.

Summing up, the cause of intelligence failure cannot be found in one specific part of the intelligence process, neither can the fault be pinned down on the intelligence community or policymakers alone. The relationship between intelligence processes and policymaker processes imply that intelligence failure not only is multi-causal but is in essence a result of discourse failure within and between these processes. While not stated explicitly, the intelligence community reforms following the 9/11 terror attacks and the Iraqi WMD controversy not only point towards a multi-causal explanation, but just as much towards discourse failure being an overarching reason.

Post-9/11 intelligence reviews and intelligence reform

Intelligence reforms have to a large degree followed the key findings of intelligence failure reviews. And although the relationship between policymakers and intelligence producers have been debated by many scholars over the years, there has been insufficient effort within

[61] Dahl, *Intelligence and Surprise Attack*, 25.
[62] NRK, "Ny Trusselvurdering Fra PST", 5 November 2014, https://www.nrk.no/rogaland/ny-trusselvurdering-fra-pst_-1.12026501, accessed 27 October 2020.
[63] Lysberg, M. and S. Tallaksen, "Slik spår de terrorfaren", Klassekampen, 29 January 2015, https://arkiv.klassekampen.no/article/20150129/ARTICLE/150129943/62052

the intelligence community to look at ways to improve this relationship. Traditionally, there has also been a noteworthy difference in American and British reform efforts. This difference can be attributed to a different understanding of what constitutes intelligence, which has had large organizational and processual implications for how to inform policymakers. One the American side there have been competing intelligence agencies with tailored products to different governmental departments. On the British side there has been a consensus-based assessment process in the Joint Intelligence Organisation (JIO)/Assessment Staff and the Joint Intelligence Committee (JIC), making intelligence functioning as an integral part of the governmental bureaucracy. Consequently, reforms in the United States have traditionally attempted to mitigate lack of analytic rigour and lack of cooperation and unity of effort. Traditional British efforts, in comparison, have for the most part focussed on collection agencies and only to a limited degree on the organisation of the JIO or the JIC. Even the strong recommendations of the Butler Review that the Assessment Branch needed its due attention for reform were ignored until the issue was forced.

The post-9/11 reforms changed to a large degree this duality of thinking, instigating changes to analytic tradecraft on both sides of the Atlantic. Despite having had almost half a century head start, the American intelligence community's emphasis on analytic tradecraft reforms to remedy analysis failure has only had an intermittent emphasis despite its enduring value. CIA's focus on how to improve the analytic effort through improved methodology has historically coincided with the personal interest and attention given from key persons, like Sherman Kent, Robert Gates, Douglas MacEachin and John McLaughlin.[64] Why did the effort not create lasting change? For the United States, Warner and McDonald found that only the reviews that met certain characteristics for both the process and the substance had lasting effects, especially reviews with influential political backing in conjunction with a war, led by future DCIs, and with the ability to create a shift of power in the American intelligence community.[65] Durbin, on the other hand, makes a two-factor argument; high foreign policy consensus and an ability to overcome the intelligence agencies' information advantages.[66] In

[64] Marchio, "Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis."
[65] Michael Warner and J. Kenneth Mcdonald, "US Intelligence Community Reform Studies Since 1947," *Studies in Intelligence*, 41-43.
[66] Durbin, *The CIA and the Politics of US Intelligence Reform*.

the United Kingdom, radical changes seem to have been watered out, arguably due to the consensus nature of the British intelligence community.

Although there are many who advocate that the reforms following the 9/11 terror attack and the 2003 Iraq War will fail just as most of the previous attempts, some claim that these latest reforms have made a more lasting change.[67] This raises the question of why the intelligence community in some areas has undergone more radical changes in the last 10-15 years than in the 60 previous years combined? Following the findings of the aforementioned studies, the two main reasons are arguably political consensus across the Atlantic, and reforms conducted during a time of war. The so-called War on Terror and the subsequent counterinsurgency operations forced intelligence agencies to make changes. More importantly, however, seems to be the passing of legislature with direct implications for intelligence analysis tradecraft. The Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 in the United States Senate led to the establishment of the Office of the Director of National Intelligence (ODNI), which also has been granted new authorities and responsibilities, enabling a more authoritative intelligence community reform in the United States.[68] In the United Kingdom, the British Government's response to the Butler Review led to the establishment of the Professional Head of Intelligence Analysis (PHIA) as a part of the Cabinet Office, showing some of the same political will for intelligence reform. The role of PHIA is ''to advise on analytical methodology across the intelligence community; and to develop more substantial training than hitherto on a cross-Government basis for all analysts working in these fields.'[69] In both cases, the result is that there now exists an overarching national agency with the authority to implement changes to analytic standards that affect intelligence analysis tradecraft across the intelligence community.

The new legislation has affected intelligence analysis tradecraft in two major ways. First, the IRTPA affirmed that improved analytic tradecraft was needed to increase analytic rigour,

---

[67] Neary, "Intelligence Reform, 2001–2009: Requiesat in Peace?"; Durbin, *The CIA and the Politics of US Intelligence Reform*; Cardillo, "A Cultural Evolution."
[68] United States Senate Committe on Governmental Affairs, "Summary of Intelligence Reform and Terrorism Prevention Act of 2004," chap. 1.
[69] Secretary of State for Foreign and Commonwealth Affairs, "Review of Intelligence on Weapons of Mass Destruction: Implementations of Its Conclusions," 10.

especially analytic objectivity and integrity. IRTPA then placed the responsibility at ODNI to ensure that all elements of the American intelligence community are to '…employ the standards of proper analytic tradecraft.'[70] The subsequent result was a renewed American focus on alternative analysis, what is known today as Structured Analytic Techniques (SATs). The British intelligence community, however, stated relatively empty handed. Due to having a head start, ODNI could turn towards the CIA Sherman Kent school for remedies while the British, and most other European intelligence communities, had to turn to the Americans for inspiration and a helping hand to improve analytic tradecraft.

Second, both the Silberman-Robb report and the Butler Review addressed the issue of how uncertainties in the intelligence assessments were presented to the readers. This led to the section in the IRTPA stating that intelligence products need to '…properly caveat and express uncertainties or confidence in analytic judgments.' [71] In the United Kingdom, changes were based on the recommendations in the Butler Review, where 'The Chief of the Assessments Staff has reviewed and re-issued guidance to his staff on language to be used in assessments.'[72] Consequently, the other main impact of the intelligence reform has been attempts of standardization of a probabilistic language and expressions of uncertainty.

The major question is: have the reforms had any impact? There are especially two areas in which the American and British intelligence communities have focused on implementing reforms into practice. First, in the United States, ODNI started to issue a set of Intelligence Community Directives (ICDs) as a consequence of the responsibilities stated in the IRTPA.[73] Of these directives, ranging from ICD 101 to ICD 906, the "ICD 203 Analytic Standards", first issued in 2007, is the one with the greatest impact on estimative intelligence analysis. The purpose of this directive is to guide analysis and analytic production, especially aiming for '…achieving analytic rigor and excellence, and for personal integrity in analytic practice.'[74]

---

[70] United States Congress, Intelligence reform and terrorism prevention act of 2004.
[71] U.S. Congress, "Intelligence Reform and Terrorism Prevention Act of 2004," sec. 1019/b/2/A.
[72] Secretary of State for Foreign and Commonwealth Affairs, "Review of Intelligence on Weapons of Mass Destruction: Implementations of Its Conclusions," para. 29.
[73] Office of the Director of National Intelligence, "Intelligence Community Directives," https://www.dni.gov/index.php/who-we-are/organizations/ps/ps-related-menus/ps-related-links/policy-division/intelligence-community-directives?highlight=WyJpY2QiLCJpY2RzIl0, accessed 27 Feb 2018.
[74] Office of the Director of National Intelligence, "Intelligence Community Directive 203: Analytic Standards," 2015, 2.

Not until 2019 did PHIA issue a British equivalent, “Professional Development Framework for All-source Intelligence Assessment”, describing both key analytic skills competencies and common analytical standards. In this framework, the standards have been promoted to ‘…be developed, and expanded upon, by individual assessment organisations to reflect their requirements and best procedures.’[75] Through these standards, both directives call for the use of analytic methods and techniques that demonstrate a rigorous, objective, and auditable comprehensive process to ensure clear, independent, and relevant intelligence assessments. In essence, the new standards mandate the change towards a much more structured approach to intelligence analysis although, as presented in chapter 6, there are national and international differences to the approach.

Second, the new British and American standards for intelligence analysis also focused on improved expressions of uncertainty in key judgements. Expressions of uncertainty in the British directive is limited to the PHIA Probability Yardstick, a probability scale divided into seven ranges.[76] The latest American ICD 203, on the other hand, divides uncertainty into both a seven-levels probability scale as well as introducing analytic confidence levels as a measurement scale on its own right.[77] The new directives have thereby clearly taken several of the issues regarding discourse failure into consideration. The downside, as discussed in chapter 5, is nevertheless that the effort has created different outcomes, both internally in the United States, as well as between nations.

The new analytic standards were seemingly immediately put into use in intelligence products. British intelligence assessments are almost always classified, severely limiting any possibility for careful examination. Russia’s invasion of Ukraine somewhat changed this, where daily Defence Intelligence assessments were published on Twitter and Facebook. In these, Defence Intelligence has often showed adherence to the Probability Yardstick.[78] There

[75] Professional Head of Intelligence Assessment, “Professional Development Framework for All-Source Intellgience Assessment,” 26.
[76] Professional Head of Intelligence Assessment, 29. The Probability Yardstick has been adapted from the original UK Ministry of Defence's Uncertainty Yardstick.
[77] Office of the Director of National Intelligence, “Intelligence Community Directive 203: Analytic Standards,” 2015, 3.
[78] As an example, see Ministry of Defence Facebook post 31 March 2022, https://www.facebook.com/photo/?fbid=344336534403487&set=pcb.344337004403440, accessed 22 March 2022

are, however, a few American National Intelligence Estimates (NIEs) published to the public in the years after the Iraq WMD missteps.[79] From 2005 onwards, page 5 in the NIEs not only states how the readers ought to understand the use of the verbal probabilities in the product but in addition describes the levels of analytic confidence given to the assessments.[80] The post-IRTPA NIEs furthermore state in the introduction pages that alternative analysis has been used, although not specifying any techniques.

Looking at the NIEs that are supposed to follow the new analytic standards with more scrutiny, several shortcomings become evident. First, none of these NIEs present alternative, testable scenarios like the ones presented in "NIE 15-90 Yugoslavia Transformed" or the "NIE 11-18-91 Implications of Alternative Soviet Futures".[81] The reader must instead put good faith in the words of the producer that SATs have been employed. Second, none of the intelligence products made available to the public in the period between 2005 and 2017 *actually* used the new standards of expressing uncertainty set forth on page 5, as in both levels of probability as well as levels of analytic confidence. Instead, these products all use the same poetic language as before the reform, where the products are filled with the same non-standard expressions, especially 'we assess', 'we judge' or 'it cannot be ruled out'.

An obvious explanation for the American lack of adherence to standards would be that the initial ICD 203 from 2007 did not contain any table of probability levels; such a table was not presented until the revised version in 2015. Yet, intelligence products ranging from the 2007 NIE "The Terrorist Threat to the US Homeland and Prospects for Iraq's Stability: A Challenging Road Ahead" to the "ICA 2017-01D: Assessing Russian Activities and Intentions in Recent US Elections" do contain a written statement of how to understand the use of the products' estimative language, including both a chart showing how the set of approved probabilistic terms relate to each other as well as an explanation of the three levels of

[79] National Intelligence Council, "National Intelligence Estimate January 2007 - Prospects for Iraq's Stability: A Challenging Road Ahead"; National Intelligence Council, "National Intelligence Estimate November 2007 - Iran: Nuclear Intentions and Capabilities"; Office of the Director of National Intelligence, "Intelligence Community Assessment 2017-01D: Assessing Russian Activities and Intentions in Recent US Elections"; National Intelligence Council, "National Intelligence Estimate July 2007 - The Terrorist Threat to the US Homeland."
[80] Wheaton, "The Revolution Begins on Page Five: The Changing Nature of NIEs."
[81] Director of Central Intelligence, "National Intelligence Estimate 15-90: Yugoslavia Transformed"; Director of Central Intelligence, "National Intelligence Estimate 11-18-91: Implications of Alternative Soviet Futures."

analytic confidence. Even so, the terms are left out in the actual language used in the products. One can therefore question to what degree the intelligence reform has had any genuine impact in the United States.

Intelligence reforms are not only limited to the USA or the United Kingdom, sometimes they do occur in smaller nations, although perhaps in a far more limited degree and for different reasons. Norway's central position to the Soviet Union resulted in a Cold War focus on data collection that had no culture for comprehensive, multi-discipline, predictive intelligence.[82] Post-Cold War, the pursuit of a more structured approach to intelligence analysis spread to Norway and several other smaller NATO countries without being linked to neither 9/11 nor the Iraq WMD controversy, but more as a bare necessity of changing tasks for the intelligence services and the armed forces. The Norwegian Army Intelligence Battalion (IntBn) was established in 2002-4 to be the backbone the intelligence, surveillance, target acquisition, and reconnaissance (ISTAR) system in Norway.[83] Cooperation with the Netherlands, taking part of the 1st German-Netherland Corps as it was put on stand-by as the Land Component Command of the NATO Response Force 4 in 2005, lay down the foundation for a common framework for teaching and using SATs in the countries' respective ISTAR units.[84] While the initial idea of the ISTAR system was to establish information superiority in support of the network centric warfare (NCW) concept, the reality the countries forces met on the ground was counterinsurgency and stabilization operations. For the ISTAR units, this led to a shift of focus from a current and target-centric recognized intelligence picture of a conventional enemy to produce more comprehensive and predictive intelligence estimates and assessments. This bottom-up approach has resulted in changes in NATO military intelligence doctrines and procedures, but also noticeable differences towards analytic tradecraft between different intelligence agencies.

---

[82] Riste and Moland, *Strengt Hemmelig: Norsk Etterretningstjeneste 1945-1970*, 345.

[83] Norwegian Ministry of Defence, 'St.prop. 42 (2003-2004): Den videre moderniseringen av Forsvaret i perioden 2005–2008', https://www.regjeringen.no/contentassets/4648088bb28649bc8458f1484d9cbe06/no/pdfs/stp200320040042000dddpdfs.pdf

[84] Course material from the Dutch Defence Institute for Security and Intelligence course at Setermoen, Norway, 18 August 2007

Furthermore, in wake of the 2014 terror warning in Norway, the Norwegian intelligence community was tasked to review how key judgements were to be communicated to policymakers. The main result was the issuing of a new community standard for verbal probabilities that was incorporated into new intelligence doctrines.[85] Despite the presence of a new a community standard, The Norwegian Intelligence Service later departed from this by creating their own standard, one they seldom adhere to in publicly available estimates.[86] Obviously, the intelligence professionals Sherman Kent named poets have proved difficult to change habits. A main cause is arguably based upon organizational culture, but other factors may as well be to blame for why using Structured Analytic Techniques and using standardized uncertainty expressions in estimative intelligence products seem to be a difficult task. A perhaps more prominent explanation is two-folded. First, analysts may have limited training and teaching in the use of SATs and can therefore have difficulty in seeing that many of the same principles ruling academic rigour should also rule analytic rigour in estimative intelligence products. Second, uncertainty as in probability and confidence, is a difficult concept, especially when there are no standard theories that analysts are taught or told to use. Both uncertainty handling and communication and analytic rigour are therefore concepts that need to be elucidated. Third, analytic tradecraft and uncertainty handling are seemingly treated as separate issues.

Summary

Intelligence failures are as old as the concept of intelligence itself. Historically, the different theories of intelligence failures have put the most blame on either analysis, collection, organisation, or policymakers respectively. To simplify, one can place the reasons for intelligence failure into two main categories: the intelligence community's inability to produce relevant intelligence to policymakers, or policymakers' failure to appropriately act on intelligence. This either-or blame game explanation is an over-simplification of reality. Contemporary writing argues instead that intelligence failures are multi-causal, that both processual and structural factors impact. Consequently, intelligence failures do not only deal

[85] Norwegian Security Police (PST), *Etterretningsdoktrine for PST*, 2017.
[86] Forsvaret, "Etterretningsdoktrinen," 2021; Etterretningstjenesten, "FOKUS 2022: Etterretningstjenestens Vurdering Av Aktuelle Sikkerhetsutfordringer."

with the fact that a warning has been given, but just as much how the warning has been interpreted. This duality has led towards a new theory of intelligence failure, discourse failure, where intelligence failure is a result of restrictions in the language and misunderstood threat perceptions. Discourse failure is furthermore not limited to the intelligence-policymaker interface but is just as evident in the rest of the intelligence process.

The theory of discourse failure shed new light to the post-9/11 intelligence reform that followed the IRTPA of 2004. For estimative intelligence, two areas were identified. First, to increase analytic rigour by implementing an analytic tradecraft based upon Alternative Analysis, thereby ensuring analytic objectivity and integrity. Second, to improve how to communicate uncertainty in estimative intelligence products. While many of the previous intelligence community reforms have been met with organizational inertia, the post-9/11 and post-Iraq War reforms have seemingly made a more lasting impact on intelligence in general, and even more so on intelligence analysis tradecraft and the subsequent dissemination of intelligence assessments. The effects of the reform have nevertheless only to a limited degree been previously investigated and not in a comprehensive manner. Consequently, key changes to intelligence analysis tradecraft will be investigated in subsequent chapters, starting by taking a closer look at what constitutes estimative intelligence.

Chapter 3: Knowledge claims, problem types, and the impact on intelligence estimates

The previous chapter investigated some of the reasons why intelligence fails and how the post-9/11 intelligence reform aims to mitigate the key ones. Yet, a limitation of much research is that intelligence is commonly referred to in the singular, thereby providing little understanding of the existence of different typologies of intelligence and how the differences impact on the analytical process. If improved analytic tradecraft is to mitigate key reasons for intelligence failure, one must first address the questions what is intelligence, and can intelligence be explained through a general theory, or do we have a family of theories?[1] There are many different interpretations of what intelligence constitutes, but as a process the term has traditionally focused either on the clandestine collection of others' secrets or on the production of multiple-source assessments to inform policymakers. Lately, more and more proposed definitions focus on the latter, stating that the purpose of intelligence is to provide policymakers with forewarning of other actors' intentions, which brings an estimative focus to intelligence. Even though secret collection and multi-source assessments can be seen as two separate schools of thought, since they answer to different intelligence problem types and therefore also different knowledge claims, it is vital to appreciate how they also complement each other. Previous research on Structured Analytic Techniques (SATs) has given this little or no attention to despite the centrality of this symbiosis.

This chapter will first define estimative intelligence, arguing that it mainly concerns multiple-source assessments concerning possible alternative futures or developments. Thereafter, intelligence as knowledge and understanding will be discussed, where the difference between the two can be said to be between insight and foresight. Moreover, there will be given reasons as to why intelligence must be understood as being fallible and not certain. This interpretation of intelligence has bearing on how to understand the different types of problems intelligence analysts are tasked to solve. The traditional distinction between secrets and mysteries is too coarse and a more refined division creates a better

[1] Gill and Phythian, "Developing Intelligence Theory," 468.

understanding of the possible impediments intelligence analysts can face. The main argument of this chapter is that estimative intelligence concerns answering questions about the future, that being mysteries, complexities, or chaotic problems, but where foresight cannot come without insight, making it equally important to uncover secrets in the process when the overall aim is increased analytic rigour and improved uncertainty handling.

Defining estimative intelligence

The very definition of intelligence has for long been a central debate in intelligence theory. A well-established definition has described intelligence as a trinity: a product, the process leading to that product, and the organizations involved.[2] This is a very all-encompassing definition and can easily lead to arguments like William Agrell's 'When everything is intelligence – nothing is intelligence.'[3] Several new definitions on intelligence have been proposed, and dependent upon the different viewpoints, intelligence ranges from something that can be done in any organization to intelligence being limited to a state-sanctioned process in need of secrecy.[4] Furthermore, in many nations Intelligence has connotation to reconnaissance and security and the word used for intelligence is in several languages better translated to information.[5] Even within the so-called Special Relationship there have historically been two opposite views on what constitutes intelligence. The traditional understanding of the term intelligence in the United Kingdom has been a narrow one where intelligence was understood as a special type of information acquired by clandestine methods and sources.[6] Across the Atlantic, however, there has long been a broad understanding, where an emphasis on the finished, and preferably predictive, all-source product resulting from the analytical process has defined the American concept of intelligence.[7] The British equivalent to the production of national all-source intelligence analysis estimates has been called assessments. To confuse the matter more there is a

---

[2] Lowenthal, *Intelligence: From Secrets to Policy*, 9; Kent, *Strategic Intelligence for American World Policy*, 76.
[3] Agrell, "When Everything Is Intelligence – Nothing Is Intelligence."
[4] Warner, "Wanted: A Definition of 'Intelligence'"; Gill and Phythian, *Intelligence in an Insecure World*, chap. 1.
[5] Davies, *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 1: Evolution of the U.S. Intelligence Community*, 44.
[6] Davies, *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 2: Evolution of the UK Intelligence Community*, 4; Herman, *Intelligence Power in Peace and War*, 114–18.
[7] Shulsky and Schmitt, *Silent Warfare: Understanding the World of Intelligence, 2nd Ed*, 159–64; Herman, *Intelligence Power in Peace and War*, 114–18.

British intragovernmental division, where contrary to the narrow civilian understanding, the military use the wide definition of intelligence.[8] In smaller Western nations the understanding of intelligence can be dependent upon how the intelligence community is organised. In Norway, where the National Intelligence Service is organised as a part of the armed forces, there is today a broad understanding of intelligence, where the definition is close to the agreed definition in NATO.[9] This division between a narrow and a wide definition on the civilian side versus the unity of thought within the military in NATO is a factor that can distort intelligence theory and thereby contribute to intelligence failure. This division furthermore underlines that one should not take for granted that intelligence theory based upon research of civilian agencies can be made fully applicable to nations where the intelligence services are organised differently.

Taking the wide definition into account, several authors agree that there nevertheless are some common aspects of intelligence.[10] First of all, intelligence is about finding relevant information others want to be kept secret. Others can be everything from a person to a nation state if the entity in question is foreign, although friends and allies normally cooperate rather than spy on each other. Second, intelligence is a state-sanctioned activity, meaning that it is the responsibility of some form of governmental run organisation, that being civilian or military. Third, intelligence is about providing recipients with knowledge and understanding to inform and advise on policy- and decision-making. Based on these three pillars, Gill and Phythian offer their definition of intelligence:

> The mainly secret activities – targeting, collection, analysis, dissemination and action – intended to enhance security and/or maintain power relative to competitors by forewarning of threats and opportunities.[11]

The definition of Gill and Phythian is echoed in a key military doctrine issued by NATO, specifying that intelligence concerns the operational environment as well as actors' capabilities and intentions.[12] In the definitions of Gill & Phythian and NATO there are two

[8] Davies, *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 2: Evolution of the UK Intelligence Community*, 6.
[9] Forsvaret, "Etterretningsdoktrinen," 2021.
[10] For a detailed debate, see Warner, "Wanted: A Definition of 'Intelligence'"; Gill and Phythian, *Intelligence in an Insecure World*, chap. 1.
[11] Gill and Phythian, *Intelligence in an Insecure World*, 19.
[12] NATO, "NATO Standard AJP-2.1(B): Allied Joint Doctrine for Intelligence Procedures."

words that stand out respectively: forewarning and intentions. Both imply a futuristic viewpoint riddled with uncertainty and ambiguity. Hence, the word estimate is central, defined in the Oxford English Dictionary as 'an approximate judgement based on considerations of probability.'[13] There is, however, more to intelligence estimates than probability alone, where probability and confidence together is one way of presenting the duality of uncertainty.[14] Uncertainty will also manifest itself differently based upon different types of intelligence products. Sherman Kent divided intelligence into three parts, basic, current and speculative-evaluative, where the latter was defined as estimative intelligence, dealing with capabilities and intentions.[15] Today, when explaining estimative intelligence for kids, the CIA has defined it as 'looking into what might be or what might happen.'[16] The first half is typically current intelligence while the latter half is often called forecasting or predictions.[17] Being speculative-evaluative about the future, there will seldom be only one possible future outcome, but rather several alternative ones with different degrees of uncertainty attached.[18] Omand arguably sums up what intelligence is all about when he states that the essence of intelligence is '…to help improve the quality of decision-making by reducing ignorance.'[19] In other words, intelligence is knowledge and knowledge is power. A key question is then what is knowledge and ways of knowing?

Intelligence as knowledge and understanding

In intelligence data is typically collected by sensors, whereby data progress into information and thereafter to knowledge.[20] Knowledge as a term is defined in different ways, from awareness based on experience to a perception of facts or truth of a subject.[21] In estimative intelligence, knowledge is closely linked the latter, which is rooted in philosophy where there are two central concepts: ontology and epistemology. Ontology can briefly be said to be the

---

[13] Oxford English Dictionary, "Estimate."
[14] See Chapter 5 for discussion on uncertainty
[15] Kent, *Strategic Intelligence for American World Policy*, 8.
[16] Central Intelligence Agency, "CIA: Who We Are and What We Do."
[17] Kuhns, "Intelligence Failures: Forecasting and the Lessons of Epistemology"; Doran, "Why Forecasts Fail: The Limits and Potential of Forecasting in International Relations and Economics."
[18] Smith, "On the Accuracy of National Intelligence Estimates."
[19] Omand, "The Future of Intelligence: What Are the Threats, the Challenges and the Opportunities?," 14.
[20] NATO, *AJP-2(B) Allied Joint Doctrine for Intelligence, Counter-Intelligence and Security*, sec. Annex B.
[21] "Knowledge", in Oxford English Dictionary, https://www.oed.com/view/Entry/104170?rskey=5EZJbm&result=1&isAdvanced=false#eid, accessed 24 November 2021.

study of the nature and categorization of reality and existence, while epistemology is the study of methodologies that examine how one can justify this existence and grouping, which leads to knowledge. From an epistemological point of view, knowledge can be either *a posteriori*, based upon sensory and perceptual information, or *a priori*, based upon an intellectual process.[22] Common in epistemology is the concept of propositional knowledge, consisting of the three components justification, truth, and beliefs. From this trinity we get a definition of knowledge as 'justified true belief'.[23] Gettier points towards a serious flaw in this definition, arguing that justified true belief not necessarily is 'a *sufficient* condition for someone's knowing of a given proposition.'[24] The Gettier cases are knowledge claims that are incorrect because the reasons for a belief are false, despite being justified. In intelligence, denial, deception, and limited access are just some issues that can produce false knowledge claims.

One response to the Gettier problem is that the justification needed for knowledge must rest on infallible premises that necessitate the truth of a proposition.[25] This response is linked to formal logic and deduction, where the problem is that the process does not produce new knowledge.[26] Furthermore, the argument of needing infallible premises treats knowledge as an absolute "thing", 'awaiting discovery through scientific investigation.'[27] However, knowledge is not a "thing", but rather an impermanent and short-lived process of relating.[28] A different proposition is therefore that knowledge is fallible and context-dependent because there will always be possibilities for error.[29] Our knowledge of something can drastically change when presented with new and surprising information. This latter argument rests on an interpretation that knowledge depends on how relevant, supportive, and comprehensive the evidence is and how valid the reasoning is.[30] Consequently, the epistemic nature of estimative intelligence entails an assessment of the

---

[22] Moser, *Oxford Handb. Epistemol.*, 4.
[23] Moser, 4.
[24] Gettier, "Is Justified True Belief Knowledge?," 123.
[25] Kirkham, "Does the Gettier Problem Rest on a Mistake ?," 503.
[26] Hintikka, *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*, 222. See Chapter 4 for further discussion.
[27] Snowden, "Complex Acts of Knowing: Paradox and Descriptive Self-Awareness," 5.
[28] Stacey, *Complex Responsive Processes in Organizations: Learning and Knowledge Creation*, chap. 1.
[29] Lewis, "Elusive Knowledge," 550.
[30] Vrist Rønn, "(Mis-) Informed Decisions? On Epistemic Reasonability of Intelligence Claims."

degree of justification for a belief or a key judgement, or as chapter 5 will discuss, how confident one is.

It follows from the previous paragraph that intelligence should be regarded as fallible knowledge, with its action-oriented scope and future-oriented content.[31] From an ontological viewpoint, however, it makes no sense talking about knowledge of the future '… because the future is *per se* unobservable and thus unknown [but] if we have assessed a specified possible future state of the world, we can monitor and update our assessment as history unfolds.'[32] Accordingly, if the aim of intelligence is to reduce ignorance, knowledge alone may not be enough, there is also a need for understanding. In a decision-making context, understanding can be defined as 'the perception and interpretation of a particular situation to provide the context, insight and foresight required for effective decision-making.'[33] Insight is the knowledge and comprehension that comes from situational awareness and analysis of what and why something has happened. Foresight is adding comprehension and judgements of what may happen in the future, aiming to elucidate implications and consequences of the current circumstances, thereby creating understanding.

The nature of estimative intelligence, providing foresight and understanding, is therefore not only epistemic but also ontological. The impact of intelligence on policy- and decision-making will be impeded if not both are not considered since 'Is there a threat?' and 'How can a threat materialize itself and develop?' have both ontological and epistemic elements. The problem is that there are both ontological and epistemological limits to precise understanding, resulting in that estimative intelligence is fallible and beset with uncertainty. Rescher states that 'the future of the domain at issue is developmentally open - causally undetermined or underdetermined by the existing realities of the present and open to the development of wholly unprecedent patterns owing to the contingencies of choice, chance, and chaos.'[34] Consequently, one also need to take decision-making into consideration, which

[31] Ben-Israel, "Philosophy and Methodology of Intelligence: The Logic of Estimate Process."
[32] Njå, Solberg, and Braut, "Uncertainty-Its Ontological Status and Relation to Safety," 15.
[33] Development Concepts and Doctrine Centre, "Joint Doctrine Publication 04: Understanding and Decision-Making."
[34] Rescher, *Predicting the Future: An Introduction to Theory of Forecasting*, 134.

often will increase the margin of error, leading to more uncertainty.[35] Some epistemological elements of that uncertainty can arguably be lessened by making the analytic process more rigorous, while others, especially ontological elements, will be hard, if not impossible, to eliminate. While this may seem like an unnecessary philosophical discussion, the epistemic and ontological nature of intelligence has bearing on the nature of different intelligence problems and how they can best be answered and communicated to reduce ignorance, foresee threats and opportunities, and thereby improve policy- and decision-making.

The taxonomy of intelligence problem types

A vital aspect of intelligence as insight and foresight is how we can categorize the different types of problems intelligence analysts are tasked to solve. A classical division of knowledge problems within intelligence has been between mysteries and secrets. The first problem type deals with information concerning adversaries' intentions while the latter concerns their capabilities.[36] Robert Gates' dictum was that 'Secrets are things that are potentially knowable', while mysteries have 'no clear-cut answers, often because the other leaders themselves do not know what they are going to do or have not worked out their problems'.[37] Secrets are therefore problems set in the past or the present while mysteries are set in the future. Being set in the past or the present, secrets can theoretically be uncovered given that the intelligence agencies have access to the right information. Secrets can for that reason be compared to what is labelled puzzles in operations research, issues that are explicit, with a known range of opinions, and with a single correct solution.[38] Accordingly, the analysis can be compared to connecting the dots. Intelligence products elucidating secrets are typical basic and current intelligence products aimed at creating insights, what Sherman Kent described as descriptive and reportorial.[39] An example is the typical Cold War secrets of the arsenal of Soviet ICBMs where the degree of certainty one could ascribe judgements was usually high given that an intelligence agency had been able to collect the right information. A different example, proving that uncovering secrets can be

[35] Guillan, "Epistemological Limits to Scientific Prediction: The Problem of Uncertainty," 515.
[36] Handel, *War, Strategy, and Intelligence*, 239.
[37] Herman, *Intelligence Power in Peace and War*, 103.
[38] Pidd, *Tools for Thinking: Modelling in Management Science*, 44.
[39] Kent, *Strategic Intelligence for American World Policy*, 7–8.

a long and difficult job due to lack of access is the hunt for Osama Bin Laden, where there to the very end was disagreement to the certainty of his whereabouts.[40]

While some secrets can be hard to uncover, judgements regarding mysteries are generally much more difficult to make since mysteries have yet to materialise and are therefore by nature not knowable. Mysteries are usually deliberately kept hidden by denial and deception, such as the preparations of the 1998 Indian nuclear test. Furthermore, being future and contingent, mysteries usually contain a degree of randomness, having many possible outcomes that can only be estimated through assessing the cause-and-effect relationship between the most important driving forces.[41] Mysteries are connected to foresight where comprehension and assessments create understanding, what Sherman Kent named speculative-evaluative assessments that deal with the future.[42] Typical products are estimates or assessments with a forecast of the course of action of a threat actor. Mysteries are impossible to predict with certainty and can often produce spontaneous surprising events like the Soviet invasion of Afghanistan, a decision that to the end was opposed by many in the Soviet hierarchy and only came about as a result of several other factors.[43]

What we see today, however, is that a distinction between only two types of intelligence problems is too simplistic. For that reason, several other divisions of intelligence problems have been promoted. Krizan, for instance, divides intelligence problems into five categories, from simplistic problems where the aim is to obtain information and the role of facts is the highest, to indeterminate problems where the aim is to predict future situations and where the role of judgement is the highest.[44] Another delineation of knowledge problems is a spectrum of puzzles, problems and messes, the latter described as an ambiguous, complex and interlinked system of problems with no certainty of an existing solution.[45] Messes are by

[40] Friedman and Zeckhauser, "Handling and Mishandling Estimative Probability: Likelihood, Confidence, and the Search for Bin Laden."
[41] Fishbein and Treverton (2004a), pp. 11-12
[42] Kent, *Strategic Intelligence for American World Policy*, 8.
[43] Westad, "Concerning the Situation in 'A': New Russian Evidence on the Soviet Intervention in Afghanistan."
[44] Krizan, "Intelligence Essentials for Everyone."
[45] Pidd, *Tools for Thinking: Modelling in Management Science*, 46–47.

others labelled complexities or wicked problems.[46] Fishbein and Treverton define complexities as problems involving

> large numbers of small sized actors, fluidity of rules governing behaviour, and the large influence of situational as opposed to internal factors in shaping behaviour. Due to these characteristics, these problems can yield a wide range of *sui generis* outcomes that defy probabilistic prediction.[47]

Wicked problems, however, are harder to define as they are unique complex societal problems without any definite or objective answers, but where there instead is a 'need of an exhaustive inventory of all conceivable solutions ahead of time.'[48] The civil war in Syria that followed the Arab Spring in 2011 is an example of a complexity or wicked problem, with its numerous actors with different desired end states, combined with sociological, cultural, and economic factors.

A stepwise delineation of intelligence problem types may still be to over-simplify the tasks given to intelligence agencies. A more cross-cutting outline of problems is a division between ordered and disordered where known and knowable problems are ordered while complex and chaotic problems are disordered.[49] Ordered problems can be compared to secrets or linkages of secrets, a typical domain of subject matter experts. Complex and especially chaotic problems, on the other hand, require a break-down of previous expert assumptions and an establishment of new patterns to be able to take rational choices between alternatives. Chaotic problems are comparable to Donald Rumsfeld's unknown unknowns or Taleb's Black Swan events, being problems that are unpredictable outliers with an extreme impact.[50] A refinement between the complex and chaotic is still necessary. Too often surprising events are explained as chaotic Black Swan events despite having to some degree been foreseen. The SARS-CoV-2 pandemic is case in point. Not only do pandemics occasionally break out, the experience of a simulation exercise of a similar scenario was presented to the White House already in 2017.[51] It may be impossible to foresee a true

---

[46] For a brief outline, see Vandepeer, *Applied Thinking for Intelligence Analysis*, 18.
[47] Fishbein and Treverton, "Making Sense of Transnational Threaths."
[48] Rittel and Webber, "Dilemmas in a General Theory of Planning."
[49] Snowden, "Complex Acts of Knowing: Paradox and Descriptive Self-Awareness."
[50] Rumsfeld, "DoD News Briefing - Secretary Rumsfeld and Gen. Myers"; Taleb, *The Black Swan: The Impact of the Highly Improbable*.
[51] Toosi, Lippman, and Diamond, "Before Trump's Inauguration, a Warning: 'The Worst Influenza Pandemic since 1918.'"

unknown unknown, but it is possible to depict situations at the border of our imagination without these scenarios being written off as science fiction.

The complexity of estimative intelligence points towards describing intelligence analysis as connecting the dots or being like a jigsaw puzzle as far too simplistic. A more appropriate analogy of estimative intelligence analysis is that it is like having a pile of jigsaw pieces from multiple puzzles, no box covers to guide you, and where several relevant pieces are missing. Although secrets, mysteries, and complexities can be seen as distinct problem sets, there are vital links between them. Answers to secrets function as indicators towards answering mysteries, which again serve as indicators towards unravelling complexities. While a poor analysis of both secrets and mysteries can lead to intelligence failures, the most prominent intelligence failures in the post-Cold War era have arguably been results of complex and chaotic problems. A reason is that while these problems most often are in the analysts' blind spot before the fact, it is in retrospect usually easy to back-track the reasons for their occurrence. This hindsight bias makes surprising events seemingly both explainable and predictable.[52] The inability to see on beforehand what may happen or is going to happen can often be attributed to cognitive limitations of the human brain, what Richard Betts has identified as one of the chief enemies of intelligence.[53] For traditional intelligence analysis, leaning on subject matter expertise, applied theory, and inductive reasoning, trying to predict complex and chaotic problems regularly lead to biased judgements, based upon the effect in Kahneman's words 'What you see is all there is'.[54] Improved understanding of cognitive limitations is one step forward, but to improve estimative intelligence, there is also a need to investigate what makes knowledge and understanding more rigorous and less ambiguous. These aspects will then have to be tied together to improve the process of estimative intelligence, thereby reducing the risk of intelligence failures.

Summary

The concept of intelligence is increasingly debated in intelligence theory. A growing number of definitions have focused on intelligence as a process aimed at informing policymakers of

[52] Taleb, *The Black Swan: The Impact of the Highly Improbable*, xxii.
[53] Betts, *Enemies of Intelligence: Knowledge and Power in American National Security*, 12.
[54] Kahneman, *Thinking Fast and Slow*, 85.

what the future may bring. Instead of focusing on intelligence as either the collection of information from clandestine sources or the production of multiple-source assessments, it is appropriate to understand intelligence as mainly estimative, in need of both. The division has, however, not considered the different types of problems intelligence deals with. Instead of a bisection into secrets and mysteries, a more refined division is between secrets, mysteries, and complexities. Secrets are present and therefore potentially knowable while mysteries and complexities are future and therefore riddled with uncertainties. They do, nonetheless, build upon each other. Secrets are vital inputs to create knowledge, which is needed for understanding of mysteries and complexities. The difference is in other words about insight and foresight. Both are fallible claims, the first typically due to denial, deception, and ambiguous information, the latter typically due to the future being contingent and therefore by nature uncertain. While insight usually has one correct answer, foresight can, and should, contain several possible alternative futures where the unveiling of more secrets can help indicate which is more likely than the others. All this necessitates an improved process of intelligence analysis. While better knowledge of cognitive limitations and how they might be mitigated is an important aspect, other equally important and connected aspects are the concepts of analytic rigour, uncertainty handling, and analytic tradecraft. Both analytic rigour and assessing uncertainty can be different secrets and insight than for mysteries/complexities and foresight, impacting on intelligence analysis methodology. Improved understanding of how these issues are interlinked will contribute to the key goal of the post-9/11 intelligence reform, to prevent intelligence to fail.

Chapter 4: Increasing analytic rigour in intelligence analysis: key building blocks

As discussed in chapter 2, a lot of effort in offsetting perceived causes of failure in intelligence has been placed on implementing new analytic standards. The new standards also introduced the term analytic rigour, being a product of adherence to the new principles of estimative intelligence.[1] A key question is thus, what factors contribute to making analysis more rigorous? Here we see that the Western intelligence communities' lack of a common understanding of intelligence and the issues connected to intelligence analysis methodology impede the intelligence discourse, often leading to confusion instead of elucidation and thereby increasing the possibility for discourse failure. Furthermore, when the Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 made the employment of the new standards of analytic tradecraft a law, the United States Congress involved itself into the old debate of intelligence as art or science. This debate is arguably a false dichotomy since core aspects of both inherently need to be present to better elucidate the important intelligence problems policymakers face. There are nonetheless certain core issues that need to be addressed since artistic intuition can lead to biases and the classical logical approaches of deduction and induction have limitations when creating new knowledge or foreseeing the future. Consequently, the duality of intelligence analysis tradecraft as both art and science implies that to increase analytic rigour one should make use of central scientific principles to harness intuition and reduce the effect of biases while at the same time keeping an open mind to new and former unseen developments.

This chapter ventures deeper than previous published writing on analytic rigour, arguing that to improve analytic rigour the intelligence community needs a more comprehensive understanding of how different factors, from philosophical theories of logic to cognitive limitations, impact how we make inferences. As basic as this may seem, very few authors have discussed these factors from an intelligence analysis point of view. When discussed, the issues have most often been treated as single academic working silos. This thesis breaks new

[1] Office of the Director of National Intelligence, "Intelligence Community Directive 203: Analytic Standards," 2007.

ground by describing the process of intelligence analysis through a European integrated social science school of thinking. By doing so, this is one of the first attempts to tie different factors together in a comprehensive manner, and in chapter 6, tie them into a structured intelligence analysis methodology.

This chapter will first address the intelligence community's problem of not having a common definition of analytic rigour, arguing that the concept in essence concerns analytic objectivity and integrity. Thereafter, the chapter will discuss how inappropriate use of heuristics from intuitive thinking leads to biases and their impact on analytic rigour and knowledge claims. Cognitive aspects are linked to modes of reasoning where there are positive and negative attributes to the different logical theories while a combined approach will make judgments more rigorous. Third, the chapter will argue that knowledge of biases and reasoning is not enough, in addition intelligence analysts need skills and attributes for creative and critical thinking to increase analytic rigour. Lastly, individual skills notwithstanding, collaboration in diverse teams will make the analytic process more rigorous. In sum, this chapter argues that analytic rigour in estimative intelligence is a result of a combination of key principles taken from both art and science, principles that need to be cultivated through a common analytic framework to mitigate discourse failure.

The nature of intelligence analysis and the impact on analytic rigour in estimative intelligence

Just as intelligence has been difficult to define, so has it been for intelligence analysis. This has led to a debate of the nature of intelligence analysis, whether it constitutes an art or a science. A main argument for the artists is that unpredictable circumstances out of control of the analysts make the use of scientific methods in intelligence analysis more "pseudo-science" than science.[2] This viewpoint is echoed in a statement from Charles Allen, a former Associate Director of Central Intelligence for Collection: 'We could have talked about the science of intelligence, but the science of intelligence is yet to be invented'.[3] Others argue that intelligence analysis is an art since it is a mixture of pattern recognition and creative

[2] For a discussion on social science, natural science and pseudo-science, see: Kuhns, "Intelligence Failures: Forecasting and the Lessons of Epistemology," 85; Marrin, "Is Intelligence Analysis an Art or a Science?"
[3] Johnston, "Developing a Taxonomy of Intelligence Analysis Variables."

thinking that comes only from experience and talent.[4] However, intelligence analysis as an art encompasses mainly the cognitive aspects of intelligence analysis, which is problematic due to how heuristics can impede creative and critical thinking, in addition to how we tend to handle and communicate uncertainty.[5]

The intelligence community's efforts to shape a more structured approach of intelligence analysis has often been viewed as a pursuit of a more scientific basis for estimative intelligence rather than improving the art of intelligence analysis. Despite the existence of a vast number of different analytic methods that analysts can utilise, there have until recently existed few methods designed specifically for the intelligence domain. It has therefore been left to the individual intelligence analysts to decide which to use and how, the same individuals 'whose expertise is confined to specific substantive areas and their own domains' heuristics.'[6] This has nevertheless not stopped repeated calls for a more scientific intelligence analysis process. Agrell and Treverton argue that '…intelligence analysis is under enormous pressure to transform from a state of proto-science in order to deliver.'[7] Sherman Kent was an early advocate for applying methods from social sciences in intelligence analysis while admitting that these methods do not have the same level of precision as natural sciences.[8] Despite the lack of precision, several authors have argued that social science should be modelled after natural science with a focus on testing hypotheses.[9] However, the scientific method for social sciences should not be interpreted as a fixed sequence of steps, but rather general principles that need to be mastered to widen the perspective and to improve the product.[10] A rule of general scientific principles is particularly applicable in estimative intelligence as adherence to these will arguably increase analytic rigour.

The continued art versus science debate has therefore brought analytic rigour at the centre of today's debate concerning intelligence analysis theory. Still, analytic rigour has yet to be accurately defined. Several of the new intelligence analysis tradecraft standards issued after

[4] For the discussion, see Marrin, "Is Intelligence Analysis an Art or a Science?"
[5] Richards, *The Art and Science of Intelligence Analysis*, chap. 5.
[6] Johnston, *Analytic Culture in the US Intelligence Community: An Ethnographic Study*, 72.
[7] Agrell and Treverton, *National Intelligence and Science: Beyond the Great Divide in Analysis and Policy*, 7.
[8] Kent, *Strategic Intelligence for American World Policy*, 60–61.
[9] For an overview of some authors, see: Lee, "A Scientific Methodology for MIS Case Studies."
[10] Gauch, *Scientific Method in Practise*, 3.

9/11 argue that adherence to the standards ensures analytic rigour. A reason is that ‘Intelligence typically looks to the methods of social and hard science for rigor, even if it almost never has the opportunity to conduct experiments.’[11] The result is that the different directives do not agree on the standards, listing 30 different top-level standards and 50 subsidiary standards, together making up 13 different main factors.[12] The two most influential directives, Intelligence Community Directive (ICD) 203 and PHIA Professional Development Framework, are at the same time largely in agreement of what constitutes analytic rigour when viewing the standards in a macro-perspective.[13] First, analytic rigour emanates from using methods and techniques that increase the integrity of the analytic product through an auditable all-source evidence base that clearly distinguishes between information, assumptions, and assessments. Second, analytic rigour is a result of analytic objectivity through employment of logical and coherent reasoning that allows for a systematic consideration of alternative hypotheses as well as an unambiguous explanation of the uncertainties associated with the key judgements.

Academia has just started to pick up on the concept of analytic rigour in intelligence, where research done in the United States and Australia is currently at the centre. A prominent view is that analytic rigour is an attribute or quality of the analytical process.[14] Marchio postulates that analytic tradecraft based upon critical thinking is the centre of gravity for analytic rigour.[15] Conversely, Corkill argues that ‘evaluation is the critical juncture from which either good or poor quality intelligence emerges.’[16] Zelik, Patterson and Woods consider analytic rigour as employment of strategies in an information analysis activity aimed at avoiding

[11] Agrell and Treverton, *National Intelligence and Science: Beyond the Great Divide in Analysis and Policy*, 80.
[12] Gelder et al., “Analytic Rigour in Intelligence,” 97–102. The official standards used in for the numbers in this thesis were: Office of the Director of National Intelligence, “Intelligence Community Directive 203: Analytic Standards,” 2015; Professional Head of Intelligence Assessment, “Professional Development Framework for All-Source Intellgience Assessment”; US Air Force and United States Air Force, “Air Force Handbook 14-133-Intelligence Analysis”; Canadian Forces Intelligence Command, “Aide Memoire on Intelligence Analysis Tradecraft”; United States Department of Justice, “Common Competencies for State, Local, and Tribal Intelligence Analysis.”
[13] Office of the Director of National Intelligence, “Intelligence Community Directive 203: Analytic Standards,” 2015; Professional Head of Intelligence Assessment, “Professional Development Framework for All-Source Intellgience Assessment.”
[14] Gelder et al., “Analytic Rigour in Intelligence,” 14.
[15] Marchio, “Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis,” 182.
[16] Corkill, “Evaluation a Critical Point on the Path to Intelligence,” 3.

shallow analysis.[17] Johnston suggests a definition of analytic rigour consisting of completeness, accuracy and integrity, achieved by employing methods and techniques that support objectivity, thoroughness, transparency, credibility, relevance, replicability, reliability, and validity.[18] The key finding of a recent expert panel found analytic rigour to be a result of an analytic process being logical, objective, thorough, stringent, and acute (LOTSA).[19] More precisely, analytic rigour implies observing relevant guidelines and employing principles of good reasoning to ensure an objective and acute analysis that addresses all relevant issues.

Analytic rigour in intelligence analysis can thus be said to be a result of a comprehensive analytical process. Exactly what that process should look like is not unequivocal. The expert panel found that the use of Structured Analytic Techniques (SATs) *per se* was not enough to ensure analytic rigour.[20] The different SATs must both be relevant for the task and be applied in a proper and rigorous manner. Analytic rigour is moreover not limited to applying proper reasoning, it is also related to creative and critical thinking skills. In general, it can be argued that analytic rigour aims at increasing analytic objectivity and integrity by mixing key elements from both art and science. Some of the other objectives, however, can be debated, where for instance truth, credibility, and accuracy are not achievable for most problem types, as discussed in chapter 2. There is therefore an underdeveloped part of the theory concerning analytic rigour, especially that different types of intelligence problems require different approaches, resulting in different types of intelligence products. Consequently, the analytic process skills needed for increased analytic rigour may differ depending on the task. Furthermore, the cognitive processes that help us come to a conclusion or key judgement largely impact on how rigorous the analytic process is.

Cognitive factors influencing analytic rigour: the impact of heuristics and biases

Analytic objectivity and integrity have been identified as key factors to achieve analytic rigour but constitute a standard that is often not easily met. The social world phenomena of

[17] Zelik, Patterson, and Woods, “Measuring Attributes of Rigor in Information Analysis.”
[18] Johnston, “Defining Analytic Rigor for Analysis in the Intelligence Community (Unpublished Report),” 7.
[19] Gelder et al., “Analytic Rigour in Intelligence,” 14.
[20] Gelder et al., 21.

which intelligence analysts are tasked to make sense of are inherently uncertain, not only due to factors like denial, deception or general lack of relevant information, but even more so due to how the human brain naturally works. People sense, analyse, make judgements and decisions continuously, from simple tasks like walking and chewing gum simultaneously to conducting rocket science. Dependent upon the issue at hand, the cognitive processes can be intuitive, conducted automatically with no or little effort and conscious awareness, or based upon reasoning and therefore more controlled, conscious, and effortful, allowing for more complex computations. These two modes of operating are called system 1 and 2 respectively and work in duality where system 1 generates impressions and system 2 conducts judgements based upon system 1's suggested interpretations.[21] This division of tasks would work well if we could make use of all of the brain's capacity, but unfortunately, our short-term processing capacity is limited to four independent variables.[22] When presented with more variables, system 1 runs into difficulties and we have to activate more of system 2, which demands a much greater cognitive effort.

Since the brain prefers cognitive ease to mental hardship, our problems are divided into multiple comprehensible tasks for which the solutions are immediately transferred to our long-term memory.[23] When doing so, our brain intuitively searches for order, creating patterns, relationships, and coherence. These simplified conceptual mental models of the reality are often based on little information in addition to being influenced by culture, education, and experience.[24] The models are stored in our long-term memory to be used for future references, functioning as mental shortcuts so that we can behave rationally without having to invoke system 2 more than necessary. The expertise and experience that form our intuition can therefore in some cases be used to better understand the context of a situation. Gigerenzer has identified ten fast and frugal heuristics, or automatic thinking processes, that are active in our cognitive adaptive toolbox and allow us to act ecologically rational, especially in high uncertainty situations.[25] In extreme cases, these heuristics can be

---

[21] Kahneman, "A Perspective on Judgment and Choice: Mapping Bounded Rationality.," 698–99.
[22] Halford et al., "How Many Variables Can Humans Process?," 76.
[23] Kahneman, *Thinking Fast and Slow*, 37-37,64-65.
[24] Heuer, *Psychology of Intelligence Analysis*, 7.
[25] Gigerenzer, "Why Heuristics Work"; Gigerenzer, "The Adaptive Toolbox."

the difference between life and death, such for medical doctors in an emergency care unit or for soldiers in a chaotic war zone.

While many heuristics are useful when originating from one's own experiences, automatic and intuitive judgements work poorly when assessing uncertain events.[26] The brain's preference for cognitive ease makes system 2 a lazy controller and when facing new or complex problems, intuition and heuristics tend to lead to systematically faulty judgements, better known as biases. The situation at hand is compared to our stored mental models where availability, representativeness, anchoring, and affect heuristics result in a natural tendency for 'satisficing', choosing the first hypothesis that explains the situation satisfactorily instead of finding the optimal one.[27] The probability of an event is consequently often based upon how easily it springs to mind, superficial resemblance, the occurrence of similar, but non-relevant instances, and our emotions. Causal explanations are favoured over randomness and chance, not taking prior probability, sample size, and the law of small numbers into consideration, and forgetting that co-occurring events do not necessitate correlation or causation.[28] Our mental models do not change easily, leading to a subconscious pattern of expectations and a bias in evidence evaluation. We turn a blind eye to contradictory information and give instead more weight to information that confirms an existing mental model. We do not consider the absence of evidence, resulting in 'what you see is all there is.'[29] Insufficient adjustments can lead to an anchoring bias, where 'different starting points yield different estimates, which are biased toward the initial values.'[30] Furthermore, ethnocentric and cultural biases often affect intelligence analysis. Complexities in intelligence are usually concerned with other nations and cultures and although cross-cultural competence of the relevant historic and contemporary political, cultural, economic systems can help to produce a feeling of knowledge of foreign behaviour, our intuition may still fall short. [31] There is a tendency to project our rationality upon foreign actors, thereby underestimating the real driving forces for how a situation may develop.

[26] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases."
[27] Simon, "Rational Choice and the Structure of the Environment"; Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases"; Heuer, *Psychology of Intelligence Analysis*.
[28] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases," 1125–28.
[29] Kahneman, *Thinking Fast and Slow*, 85–88.
[30] Tversky and Kahneman, "Judgment under Uncertainty: Heuristics and Biases," 1128.
[31] Heuer, *Psychology of Intelligence Analysis*, 134–38.

The key takeaway is that the human mind is an associative machine that is inclined to jump to conclusions.[32] Relying on the intuitive system 1 can therefore create an illusion of validity and an unwarranted confidence our ability to predict future events. While intuition serves us well on an everyday basis, this is not the case when making judgments under uncertainty. Cognitive biases derived from faulty intuition and automatic reasoning impede analytic rigour and can lead to intelligence failures, especially when an active, deceiving adversary is involved. Nevertheless, cognitive biases are arguably inescapable, and a better understanding of these is no guaranteed safeguard against judgement failures. At the same time, to improve intelligence analysis, one needs to find a way that can reduce the effect of heuristics and biases.

Cultivating analytic rigour through a compound mode of reasoning and inferences

A starting point for intelligence analysts to reduce the effects of heuristics is to have more conscious thinking about thinking, or in other words, to make better use of the brain's system 2 so that it becomes more than a lazy controller. Analysts need a better understanding of how different types of reasoning create knowledge claims and their respective potential pitfalls. Moreover, analysts should be encouraged to make better use of their rational and controlled reasoning skills to increase analytic rigour. In addition to analytic objectivity, analysts must strive towards increased integrity in how they make inferences. Inferences can be made using different modes of reasoning, which have historically been viewed upon as either inductive or deductive. Inductive reasoning can be defined as 'inference of a generalized conclusion from particular instances.'[33] Deductive reasoning, in contrast, is 'inference in which the conclusion about particulars follows necessarily from general or universal premises.'[34] Induction relies on the correspondence principle of truth while deduction relies on the principle of truth as coherence.[35] These

---

[32] Kahneman, *Thinking Fast and Slow*, 50-58,79-88.
[33] *Merriam-Webster*, http://www.merriam-webster.com/dictionary/induction, 3 May 2020
[34] *Merriam-Webster*, https://www.merriam-webster.com/dictionary/deduction, 3 May 2020
[35] Pedersen and Johansen, "Behavioural Artificial Intelligence: An Agenda for Systematic Empirical Studies of Artificial Inference."

definitions are simplifications and there is more to induction and deduction, but both these modes of inference have weaknesses, whereas a third alternative for logical reasoning is better suited to increase analytic integrity in intelligence assessments.

Analysts' traditional mode of reasoning is, according to Kuhns, one of induction.[36] In intelligence analysis, inductive reasoning can be compared with the classical metaphor of intelligence analysis as connecting the dots. This makes induction comparable to pattern and trend analysis, an analytic technique commonly used to produce both basic and current intelligence products, especially when monitoring armed conflicts. Moreover, inductive reasoning is linked to the intuitive system 1, where the results is general models for further reference based upon specific pieces of information, which then is supposed to be controlled by system 2 with more consciously aware judgements.

For intelligence analysts, a fundamental problem emerges when attempting to make predictions based upon these pattern and trend models. Unlike the past and the present, the future has yet to happen and for that reason one does not have any experience with it. Following the arguments of David Hume, if using inductive reasoning for estimative intelligence, analysts would make judgements where:

> …all arguments concerning existence are founded on the relation of cause and effect; that our knowledge of that relation is derived entirely from experience; and that all our experimental conclusions proceed upon the supposition that the future will be conformable to the past.[37]

Hume consequently argues that inductive reasoning is flawed since it is not possible to logically conclude that past and present observations can be used to make certain judgements about the future. He states: ‘If there be any suspicion that the course of nature may change, and that the past may be no rule for the future, all experience becomes useless, and can give rise to no inference or conclusion’.[38] Sextus Empiricus argued for an equivalent view centuries ago:

> When they propose to establish the universal from the particulars by means of induction, they will effect this by a review of either all or some of the particulars. But

[36] Kuhns, “Intelligence Failures: Forecasting and the Lessons of Epistemology,” 88.
[37] Hume, *An Enquiry Concerning Human Understanding*.
[38] Hume.

> if they review some, the induction will be insecure, since some of the particulars omitted in the induction may contravene the universal; while if they are to review all, they will be toiling at the impossible, since the particulars are infinite and indefinite. Thus on both grounds, as I think, the consequence is that induction is invalidated.[39]

While certainty about the future is not guaranteed, the probability of a certain outcome can be predicted when there already exists a history of different outcomes.[40] The strength of a conclusion based upon inductive inference is therefore a matter of degree. Even though some events can be probabilistically estimated, inductive reasoning alone regularly cause forecasting failure. While it may seem like history repeats itself, it really does not. Similarities notwithstanding, there are always smaller or larger changes to some particulars of the course of events. Doran therefore asserts that inductive forecasts '...ultimately fail because no technique has been developed that allows the forecaster to predict, prior to the event itself, when a nonlinearity will occur.'[41] As an example, counting Russian forces at the Ukrainian border in 2014 would not by itself give proof of an upcoming invasion. Historically, the frequency of Russian invasions in neighbouring countries following such movements are by far lower than the frequency of no invasions.

Then how can intelligence analysts remedy the problem of induction? The historical opposite to induction is deduction, a form of logical reasoning dating back to Aristoteles. Deductive reasoning relies on having valid arguments, meaning that the truth of a conclusion is guaranteed if the premises are true.[42] Deduction is furthermore a top-down approach, starting with a generalisation that explains something particular, compared to the bottom-up approach of inductive reasoning that starts with particulars and ends up with a generalization. Although the notion of reaching an absolute certain conclusion can be fascinating for intelligence analysts and policymakers alike, there is nevertheless a paradox to deductive inference. Cohen and Nagel question to what degree a deductive inference can be informative:

> If in an inference the conclusion is not contained in the premises, it cannot be valid; and if the conclusion is not different from the premises, it is useless; but the

[39] Empericus, *Outlines of Pyrrhonism*, 283.
[40] Hume, *An Enquiry Concerning Human Understanding*.
[41] Doran, "Why Forecasts Fail: The Limits and Potential of Forecasting in International Relations and Economics."
[42] Internet Encyclopedia of Philosophy, https://www.iep.utm.edu/ded-ind/, 7 May 2020

> conclusion cannot be contained in the premises and also possess novelty; hence inference cannot be both valid and useful.[43]

For intelligence analysts, knowing that all the premises are true is in most cases impossible. Even in intelligence problems categorized as secrets, there will always be ambiguous and uncertain information that need to be used to reach a conclusion. The whereabouts of Osama Bin Laden was uncertain until the positive identification of his body after the raid of the Alottabad compound. Moreover, the formalism of deductive logic will make impossible to create new knowledge, what Hintikka calls 'a scandal of deductive logic'.

> […] upon being told that deductive reasoning is "tautological" or "analytical" and that logical truths have no "empirical content" and cannot be used to make "factual assertions": in what other sense, then, does deductive reasoning give us new information? Is it not perfectly obvious there is some such sense, for what point would there otherwise be to logic and mathematics?[44]

The problem of deduction is that it does not rely on an observable reality in the same way that induction does. Instead, a valid deductive argument relies on an internal consistency of the relation between the premises and the conclusion. Deduction can therefore not be used for foresight. Consequently, deductive reasoning will alone not be very helpful when analysing a complex external social world and definitely not so when attempting to predict the future.

When put in the context of intelligence analysis, the problems of both inductive and deductive logic imply that neither mode can alone produce any certain foresight or understanding. Fortunately, intelligence analysts do not limit themselves to either inductive or deductive reasoning, but rather use a mix of the two. It is arguably more correct to say that analysts' preferred mode of reasoning is to use the hypothetico-deductive method, also known as the scientific method.[45] The problem with the original hypothetico-deductive method is that it relies on a logic of confirmation. From a generated hypothesis, deductions are made to predict outcomes and one search inductively for empirical evidence that can verify the hypothesis. In other words, the hypothesis '…is confirmed by its true

[43] Cohen and Nagel, *An Introduction to Logic and Scientific Method*, 173.
[44] Hintikka, *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*, 222.
[45] https://plato.stanford.edu/entries/scientific-method/#HisRevAriMil, accessed 11 May 2020

consequences.'[46] The hazard of this approach is confirmation bias, which has been the root of the problem for several intelligence failures.

Karl Popper rejected the inductivist approach to the original scientific method, arguing that a hypothesis can never be proven. He stated that 'It is easy to obtain confirmations, or verifications, for nearly every theory—if we look for confirmations.'[47] As a preferred alternative, Popper promoted the use of critical rationalism and falsification, arguing that '…the criterion of the scientific status of a theory is its falsifiability, or refutability, or testability'.[48] With his concept of critical rationalism Popper claimed to have solved the problem of induction, meaning that instead of gaining knowledge from evidence alone, scientists get knowledge from substituting refuted hypotheses with better ones.

For intelligence analysts, however, a strict obedience to the principle of refutation can lead to either dogmatic or naive falsificationism.[49] A dogmatic falsification is one for which a hypothesis may still be true despite contradictory information, typically due to the Duhem-Quine problem where one does not take an auxiliary theory into account. A naive falsification, on the other hand, is one when there is more than one hypothesis that is not falsified and there is no way to choose among them when only considering falsifications instead of also taking corroborative information into account. In intelligence analysis, one can be presented with reports containing false information in addition to suffering from lack of information due to not having access to the target. These are two of many sources of ambiguity and uncertainty in intelligence reporting, and falsification as an only principle can therefore do more harm than good.

Is one's judgements confined to either a probable outcome based upon statistics, to necessary inferences, or to the classical scientific method of a main and null hypothesis? Fortunately, there is a third mode of reasoning that can be a viable path out of the predicaments of traditional logic. Charles Pierce was the first to distinguish this third mode

[46] https://plato.stanford.edu/entries/scientific-method/#HisRevAriMil, accessed 11 May 2020
[47] Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge*, 47.
[48] Popper, 48.
[49] Lakatos, *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*; Harding, *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis*.

of reasoning, dividing between induction, deduction, and hypothesis.[50] Later he named the latter mode of logical reasoning for abduction, an inference aiming to form a new hypothesis that can best explain the available information:

> The surprising fact, C, is observed;
>
> But if A were true, C would be a matter of course,
>
> Hence, there is reason to suspect that A is true.[51]

Abductive reasoning is therefore, according to Pierce, inference from particulars to a testable hypothesis from which new knowledge can be formed.

Abductive reasoning has evolved somewhat since the original ideas of Pierce and is by many today understood as 'inference to the best explanation.'[52] Its utility spans from everyday use to scientific discoveries. There is even a claim that abduction, when combined into a comprehensive process with both deduction and induction, is 'the inference that makes science'.[53]Abduction differs from inductive inference by appealing to pragmatic explanatory considerations instead of being limited to frequencies. Furthermore, abductive logic and inference to the best explanation works best when having several hypotheses to compare instead of testing just one. Abductive inference will then be a comparative conclusion producing a pragmatic principle of truth:

> Given evidence E and candidate explanations $H_1,\ldots, H_n$ of E, if H_i explains E better than any of the other hypotheses, infer that H_i is closer to the truth than any of the other hypotheses.[54]

There are objections to the understanding of abduction as inference to the best explanation. Chief among this is when comparing a given set of hypotheses, abductive inference would be limited to find 'the best of a bad lot'.[55] However, we see that when using Kuiper's idea of truth approximation, which is hypothetico-probabilistic instead of hypothetico-deductive, we can deal with both the problems of induction, deduction, and dogmatic and naive

---

[50] Pierce, "On the Natural Classification of Arguments," 284.
[51] Pierce, "Harvard Lectures on Pragmatism," para. 189.
[52] Harman, "The Inference to the Best Explanation"; Walton, *Abductive Reasoning*, 4.
[53] Mcmullin, "The Inference That Makes Science," 184.
[54] https://plato.stanford.edu/entries/abduction/#Bib, accessed 14 May 2020. See also Kuipers, "Inference to the Best Theory, Rather than Inference to the Best Explanation — Kinds of Abduction and Induction."and Kuipers, "Empirical Progress and Truth Approximation by the 'Hypothetico- Probabilistic Method.'"
[55] van Frassen, *Laws and Symmetry*, 143.

falsificationism. Additionally, when considering how medical doctors are doing diagnostics on a patient, the method for reaching a clinical judgement is in essence abductive.[56] Over time several authors have argued that intelligence should look to medicine to improve intelligence analysis.[57] Following this argument, a preferred method for increased analytic rigour in estimative intelligence is abductive logic aiming at inference to the best explanation which is closer to the truth compared with the alternative hypotheses, or in other words: the 'best truth'.[58]

Abduction alone may nevertheless not be key to increase analytic rigour. A reasoning process similar to abduction called intelligence sensemaking has of late been recommended as a way of improving intelligence analysis to make it easier to understand complexities.[59] Sensemaking was first introduced by Karl Weick in the 1970 as a process to shift focus from organizations as entities to the cognitive aspect of organizing as a 'retrospective development of plausible images of what people are doing'.[60] The aim is to create shared meaning, awareness and understanding that can serve as a springboard to action.[61] Sensemaking as a concept has developed since its origin and is no longer confined to organizational studies. Sensemaking was introduced to information science in the 1980s and then later to the military concept of network centric warfare in the early 2000s. Paik and Pirolli argue that sensemaking is a set of processes that both seeks and filters information in addition to constructing representative mental models that best fit the available information and can function as a basis for understanding.[62] In relation to intelligence analysis, sensemaking can be described as a process that 'involves the application of expertise, imagination, and conversation—and the benefit of intuition—… to identify changes in existing patterns or the emergence of new patterns, without systematic consideration of

---

[56] Rejón Altable, "Logic Structure of Clinical Judgment and Its Relation to Medical and Psychiatric Semiology."
[57] Walter, "The Question of Judgment: Intelligence and Medicine,"; Marrin and Clemente, "Improving Intelligence Analysis by Looking to the Medical Profession"; Marrin and Torres, "Improving How to Think in Intelligence Analysis and Medicine."
[58] Berkowitz and Goodman, *Best Truth: Intelligence in the Information Age*, ix.
[59] Sands, "Thinking Differently: Unlocking the Human Domain in Support of the 21st Century Intelligence Mission"; Fishbein and Treverton, "Making Sense of Transnational Threaths"; Moore, *Sensemaking: A Structure for an Intelligence Revolution*.
[60] Weick, Sutcliffe, and Obstfeld, "Organization Science and the Process of Sensemaking," 409.
[61] Weick, *The Social Psychology of Organizing*; Weick, *Sensemaking in Organizations*; Weick, Sutcliffe, and Obstfeld, "Organization Science and the Process of Sensemaking."
[62] Paik et al., "Cognitive Biases in a Geospatial Intelligence Analysis Task: An ACT-R Model," 2168.

alternative hypotheses.'[63] The complex, wicked, or chaotic problems intelligence analysts often face are often not possible to elucidate using traditional methods of analysis. Sensemaking in intelligence analysis can then be applied to generate an understanding of highly uncertain situations, going beyond the traditional modes of inference in both analysis and synthesis and instead focus on interpreting the results of these processes with abductive logic.

A main argument from proponents of intelligence sensemaking is that it makes interconnected and volatile complex problem easier to understand.[64] At the same time, sensemaking, like all other cognitive activity, deals with representations and models of the reality. Sensemaking can therefore be just as vulnerable to misconceptions as other cognitive processes if using the wrong models due to the inability to widen ones' horizon. Former CIA analyst Jack Davies put it this way:

> Available data on a complex issue is inherently ambiguous, open to manipulation by denial and deception, and otherwise subject to misinterpretation. The analysts' understanding of how things usually work on the subject at hand, what one academic observer calls "normal theory," does not adequately account for seemingly unprecedented or exceptional developments, overlooked key variables, foreign actors' distinctive risk-benefit calculations.[65]

Consequently, intelligence sensemaking can be just as prone to biases and pitfalls as other modes of inference. At the same time, analysts seldom limit themselves to only one mode of reasoning; a more conscious choice based upon using a mix of modes against a backdrop of sensemaking will be advantageous to increased analytic rigour. Intelligence analysts nevertheless need additional cognitive features to increase analytic rigour in estimative intelligence analysis.

Increasing analytic rigour through creative and critical thinking

In the wake of the 9/11 terror attacks, the lack of imagination was emphasized as one of the main reasons for the failure to prevent the attacks.[66] Imagination is the mind's ability to

---

[63] Fishbein and Treverton, "Making Sense of Transnational Threaths."
[64] Fishbein and Treverton.
[65] Davies, "Strategic Warning: If Surprise Is Inevitable, What Role for Analysis?," 8.
[66] National Commission on Terrorist Attacks Upon the United States, "The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States."

create mental models without prior sensory inputs, what Kant calls ‘the faculty of representing an object even without its presence in intuition.’[67] The mental models resulting from imagination can then converge with other cognitive processes. In contrast to abductive reasoning, whose purpose is sensemaking, the purpose of imagination is sense-breaking, letting the mind jump fences, widen the horizon of thoughts, and create associations beyond the limits of reasoning.[68] Others calls this creative imagination, allowing for the creation of not only mental models, but social, cultural, and even physical representations.[69] Imagination holds the power to spark the creation of new ideas of our world. Imagination is therefore just as useful in intelligence analysis as for works of fiction like *Harry Potter* or *Star Wars* because it allows us to think about unknown unknowns without having had any experiences with them. Guilford calls this divergent thinking, thinking that allows for the generation of multiple possible answers to a problem, compared to convergent thinking that seeks a single, correct answer.[70]

Although we all possess the ability to imagine, some are more creative thinkers than others. That does not mean that creative thinking cannot be learned and developed. Since Guilford's idea of divergent versus convergent thinking, the shift from an industrial economy to a knowledge economy has given rise to increased research within creative thinking skills.[71] This research is still in its infancy and there are many unanswered questions. Principal skills, however, seem to be problems generation and solving combined with divergent thinking techniques that allows us to think outside the box. Yet, a vivid imagination can by itself be counterproductive; in addition to creative thinking skills, intelligence analysts need a disposition for critical thinking to question inferences and challenge heuristics. Williamson connects imagination to counterfactual thinking, suggesting that ‘when we work out what would have happened if such-and-such had been the case, we frequently cannot do it without imagining such-and-such to be the case and letting things run.’[72] Although creative

[67] Kant, *The Critique of Pure Reason*, 276–77.
[68] Pendleton-Jullian and Brown, *Pragmatic Imagination: Single from Design Unbound*.
[69] See Krummel, “Imagination, Formation, World, and Place: An Ontology.” for a comprehensive discussion
[70] Guilford, *The Nature of Human Intelligence*, chaps. 6–7.
[71] Sawyer, “Educating for Innovation”; Greiff, Niepel, and Wüstenberg, “21st Century Skills: International Advancements and Recent Developments.”
[72] Williamson, “ARMCHAIR PHILOSOPHY, METAPHYSICAL MODALITY AND COUNTERFACTUAL THINKING.”

and critical thinking are two different concepts, they are connected and work best in tandem.[73]

There are many different definitions of critical thinking but in essence it is careful goal-directed thinking aimed at objectively evaluate all information instead of jumping to conclusions.[74] Critical thinking was introduced as an educational goal in the United States in 1910 and has now become one of the core skills wanted for in-demand jobs. This demand has led to a call from the Centre for Educational Research and Innovation of the Organization for Economic Development (OECD) in early 2018 for research on how to improve creative and critical thinking in higher education.[75] Just as for higher education, the field of intelligence analysis would benefit from utilising a reflective framework of critical thinking. Critical thinking allows one to mitigate the natural inclination for satisficing by suspending the judgement for an explanatory hypothesis until the body of information can reject the other possible explanations with a necessary degree of certainty. Moore argues that critical thinking will improve the analytic process and that analysts who adopt critical thinking stand to improve their analyses as it will 'mitigate the effects of mind-sets and biases by invoking skilful examination of evidence both for and against an issue, as well as consideration of obvious and less obvious alternative explanations.'[76] Given the ambiguous nature of available information and evidence intelligence analysts face, critical thinking ought therefore be a mix of Popper's critical rationalism and abductive reasoning in order not be dogmatic in the falsification process, but rather advocate the hypothesis that at any given time has the least amount of inconsistent evidence.

Critical thinking is not confined to evaluation of hypotheses alone but is relevant for the whole intelligence process to increase analytic rigour. The process of critical thinking consists of several components, from defining the main problem and its sub-problems, via formulating possible hypotheses and relevant evidence, to gather and evaluate this evidence

[73] Glassner and Schwarz, "What Stands and Develops between Creative and Critical Thinking? Argumentation?"
[74] https://plato.stanford.edu/entries/critical-thinking/#DefiCritThin, accessed 24 May 2020
[75] OECD [Organization for Economic Cooperation and Development] Centre for Educational Research and Innovation, 2018, Fostering and Assessing Students' Creative and Critical Thinking Skills in Higher Education, Paris: OECD. Available at http://www.oecd.org/education/ceri/Fostering-and-assessing-students-creative-and-critical-thinking-skills-in-higher-education.pdf, accessed 24 May 2020
[76] Moore, "Critical Thinking and Intelligence Analysis."

to draw conclusions.[77] A high standard critical thinking process will allow for more precise intelligence problems, better formulated hypotheses, more relevant collection and by that, arguably also better judgements. Such a critical thinking process moves beyond knowledge resulting from inference alone, and Sands makes the argument that:

> Cross-cutting critical thinking is the realization that there are forms of structured analyses and methods which may not initially drive to forecasts through a traditional analysis, but will provide means to challenge assumptions or hypotheses and cast a wider net when it comes to possible outcomes.[78]

A comprehensive process of critical thinking will therefore, in addition to experiments, observations, and inferences, also include imagination, active collection and evaluation for better identification, and analysis and judgements of arguments and hypotheses.

To be able to think critically, research conducted for the American Philosophical Association has found that people must possess certain attitudes, dispositions, and abilities.

> The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit.[79]

Accordingly, the ideal critical thinker possesses certain cognitive skills in interpretation, analysis, inference, explanation, evaluation, and self-regulation. Moreover, one must have the aptitude to exercise those skills through a variety of affective dispositions for critical thinking.[80] Since our natural way of thinking is undisciplined, cultivating critical thinking skills and the affective dispositions becomes essential. Critical thinking should consequently be at the core of a methodical and careful learning process, starting already with primary education for the skills to be used naturally on all occasions.[81] In intelligence analysis, however, additional steps might be needed, and Hendrickson suggests an adapted approach

---

[77] Hitchcock, "Critical Thinking as an Educational Ideal," 485.
[78] Sands, "Thinking Differently: Unlocking the Human Domain in Support of the 21st Century Intelligence Mission," 14.
[79] Facione, "Critical Thinking: A Statement of Expert Consuncus for Purposes of Educational Assessment and Instruction," 2.
[80] Facione, "Critical Thinking: A Statement of Expert Consuncus for Purposes of Educational Assessment and Instruction."
[81] Fastvold, "Socrates versus Peter Pan : The Difficulties of Thinking Critically."

to improve both the process and the product of intelligence analysis. His approach includes hypothesis testing, causal analysis, counterfactual reasoning, and strategy assessment, leading to best explanations, relevant causes, probable scenarios, and optimal decisions.[82]

If abductive logic combined with creative and critical thinking increase analytic rigour, will predictions also improve? Research has found that some that are better at forecasting, predicting, and foreseeing future situations than others. Tetlock and Gardener call these persons superforecasters, persons who regardless of background and training are better at predicting than subject matter experts.[83] Based on the findings from the Good Judgement Project, many of the skills and attributes presented in this chapter have been identified as essential to become a superforecaster.[84] Recent research has found support for several of these findings, especially cognitive abilities and styles.[85] Although superforecasters score higher than average on intelligence and open-mindedness, what makes them good is 'the hard work of research, the careful thought and self-criticism, the gathering and synthesizing of other perspectives, the granular judgements and relentless updating.'[86] The findings of Tetlock and Gardener resembles what Cynthia Grabo found 50 years ago to be essential intellectual attributes for intelligence warning analysts, namely an insatiable curiosity, aptitude for detailed research, imagination, a retentive memory, recognition of that which is important and not the least, the ability to entertain various hypotheses.[87]

Intelligence service can unfortunately not rely on being able to recruit or train every analyst to be a superforecaster. The complexity of contemporary intelligence problems also indicates the need for analytic teams since current problems are better answered through collaboration and coordination, not by a single subject matter expert.[88] Thus, teamwork is essential to increase process rigour since 'unaided individuals cannot do the tasks. In many areas of intelligence analysis, teamwork is not an option, it is a necessity.'[89] Collaboration

[82] Hendrickson, "Critical Thinking in Intelligence Analysis."
[83] Tetlock and Gardner, *Superforecasting: The Art and Science of Prediction*.
[84] Mellers et al., "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions."
[85] Beadle, "FFIs Prediksjonsturnering: Datagrunnlag Og Foreløpige Resultater. FFI-Rapport 21/00737."
[86] Tetlock and Gardner, *Superforecasting: The Art and Science of Prediction*, 231.
[87] Grabo, *Handbook of Warning Intelligence: Assessing the Threat to National Security*, 103–7.
[88] Fishbein and Treverton, "Making Sense of Transnational Threaths."
[89] Hastie, "Group Processes in Intelligence Analysis," 191.

can bring more perspectives, highlight information gaps, and expose divergent perspectives, resulting in richer and deeper analysis.[90] The Good Judgement Project also found significant peer effects from teams of elite superforcasters.[91] Increased collaboration may have drawbacks and necessary steps are needed to be taken to avoid groupthink, premature consensus, and polarization. R.V. Jones voiced the primacy of the single analyst, arguing that intelligence based upon committee work would not be bold enough.[92] Creating good diverse teams therefore becomes imperative for the group process. First, good teamwork skills are necessary attributes to learn in addition to good analytical skills. Exactly what skills are needed is still up for debate as more research is needed, although there is evidence of team training and shared cognition being central.[93] Second, diversity in opinion is good to a certain point, and must be balanced to the problem to be solved so that cognitive conflicts do not impede the process.[94] Third, a team-based approach can be enhanced by mixing specialists and generalists for mutual benefit. Specialists are needed for their in-dept subject matter expertise, while generalists are needed to challenge the conventional wisdom through a tentative and balanced mode of thinking, especially in cases where history at best can be used as an analogy.[95] To help both generalists and specialists, analytical methodologist should be brought in to supervise the process if not already an integral part of the analytical teams. Johnston argues that 'methodologists would act as in-house consultants for analytic teams, generate new methods specific to intelligence analysis, modify and improve existing methods of analysis, and promote the professionalization of the discipline of intelligence.'[96] Analytic rigour is consequently the result of more than the methodology behind how the individual analyst reach key judgments but just as much the methodological culture in the

---

[90] National Research Council, *Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences*, 64.
[91] Mellers et al., "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions."
[92] Jones, *Most Secret War*.
[93] Shuffler et al., "Developing, Sustaining, and Maximizing Team Effectiveness: An Integrative, Dynamic Perspective of Team Development Interventions"; McEwan et al., "The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions"; Salas, Cooke, and Rosen, "On Teams, Teamwork, and Team Performance: Discoveries and Developments."
[94] National Research Council, *Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences*, 65–66.
[95] Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*, 21; see also Kent, *Strategic Intelligence for American World Policy*, 121–22.
[96] Johnston, *Analytic Culture in the US Intelligence Community: An Ethnographic Study*, 73.

intelligence agency, working together within a shared analytic framework. For intelligence analysts to become more adept to mitigate reasons for analysis and discourse failure, a common intelligence analysis methodology can be used as a starting point for both the art and science of intelligence analysis. Such a methodology must take into consideration the cognitive aspects of intelligence analysis as well as the skills, attributes, and knowledge of modes of reasoning needed to make them better forecasters than the regular pundit. Collaborative use of Structured Analytic Techniques (SATs) can provide such a framework if used in the right manner.

Summary

Increased analytic rigour in intelligence has frequently been pointed to as the key to prevent intelligence failures, prompting a more scientific approach to intelligence analysis. However, a universal understanding of analytic rigour does not exist, although in principle it implies proper reasoning skills to ensure analytical objectivity and integrity. A rigorous intelligence analysis process is thus more than what unconsciously happens inside the head of an analyst. There are both cognitive and philosophical limitations that can impede the process of intelligence analysis. While better knowledge of these limitations cannot eliminate their impact, the effects can be reduced. Using inductive, deductive, and not the least abductive reasoning combined into sensemaking is one step forward to increase analytic rigour. However, to elucidate the range of contemporary intelligence problems intelligence analysts also need to foster the art of intelligence analysis by combining an improved reasoning process with creative and critical thinking. Consequently, intelligence analysts must possess skills and attributes for creative and critical thinking, have knowledge and experience with scientific methods, hereunder different modes of inferences, in addition to understanding how our brain works and the effect of different cognitive processes. The key to a rigorous analytic process is to combine art and science and then to encompass these social science principles into the estimative process.

Even so, individual knowledge and skills of these processes is not enough to increase analytic rigour in estimative intelligence. To mitigate discourse failure, intelligence agencies should also aim towards diverse collaborative analytic teams that work within the framework of a common analytic methodology that emphasise the use of the more controlled, conscious, and rigorous cognitive system 2. Collaborative use of Structured Analytic Techniques (SATs) can be such a step forward, and chapter 6 will present the intelligence community's effort in implementing SATs as a standard for analysts to use. At the same time, these techniques need to be put into a larger and more comprehensive framework based upon a general social science methodology to become more than just a tick in the box. Furthermore, increased analytic rigour through an improved analytical process is only one step towards mitigating discourse failure since it will not necessarily eliminate uncertainty. The next chapter will delve deeper into different issues that are central to the handling and expression of uncertainty in estimative intelligence, key elements of how to mitigate reasons for discourse failure.

Chapter 5: Assessing and expressing uncertainty: key theories and concepts

From the discussions in the previous chapters, we see that the goal of intelligence is to create knowledge and understanding and that analytic rigour is essential for reaching this goal. Yet, adhering to the principles of analytic rigour does not safeguard access to all the relevant information. Lacking or having imperfect information creates doubt and uncertainty, which again can lead to erroneous judgements. Being interdisciplinary, uncertainty as a term, concept, and construct has no universal definition but rather a history of conflation of terms. Intelligence theory is no exception, especially in its lack of a consistent terminology of uncertainty handling.[1] Intelligence analysts and policymakers have often had conflicting views as to what degree ambiguity and doubt should be conveyed in intelligence products.[2] When key intelligence judgements are being presented with the use of verbal expressions of uncertainty, the lack of a common understanding fuels discourse failure. As Heuer argues: 'They are empty shells. The reader or listener fills them with meaning through the context in which they are used and what is already in the reader's or listener's mind about that context.'[3] The ambiguity in the messages conveyed can then lead to discourse failure.

How to express uncertainty in intelligence products has been an enduring topic, but the debate resurfaced as part of the IRTPA intelligence reform.[4] The renewed focus has brought the debate forward, although lacking nuance by being too focused on increasing accuracy. To understand uncertainty, the concepts of probability and confidence are central. This chapter will more comprehensively describe central theories of uncertainty, probability, and confidence, especially how they differ between natural and social sciences. The chapter will thus show that uncertainty in general can be divided into two, an objective probabilistic tendency and a subjective degree of belief. Since people generally perceive objective and

[1] M. Isaksen and McNaught, "Uncertainty Handling in Estimative Intelligence – Challenges and Requirements from Both Analyst and Consumer Perspectives," 644.
[2] Hulnick (1986), p. 215
[3] Heuer, 1999, s. 153
[4] United States Congress, Intelligence reform and terrorism prevention act of 2004; Lord Butler, "Review of Intelligence on Weapons of Mass Destruction"; Kent, "A Crucial Estimate Relived"; Wark, "The Definitions of Some Estimative Expressions."

subjective uncertainty differently, this impacts uncertainty handling in intelligence. The main argument of this chapter is that uncertainty in estimative intelligence should be expressed as a combination of probability of the occurrence of different outcomes together with a level of analytic confidence describing the basis for the probabilistic judgement. Together they tell the relative probability of an event compared to the alternatives, in addition to saying to what degree the probability can change given new information. Although accuracy is unattainable due to the nature of estimative intelligence, improved uncertainty handling will mitigate one of the key reasons for discourse failure between intelligence and policymakers.

Probability as a measurement of uncertainty

The discourse about doubt and uncertainty can be traced back to Plato, yet without academia being able to converge on a commonly shared terminology or typology. One general definition of uncertainty is 'limited knowledge about future, past, or current events.'[5] Another one is 'any departure from the unachievable ideal of complete determinism.'[6] It is also common to talk about Knightian Uncertainty, being uncertainty that cannot be measured as opposed to Knightian Risk which can be.[7] Common for all is that uncertainty is an attribute of imperfect knowledge and understanding and the effect this will have on judgements and decision-making. The larger the uncertainty, the larger the risk. Despite Knight's notion of unmeasurable and measurable uncertainty and, as discussed in chapter 1, that there are both ontological and epistemological attributes to uncertainty, the main problem is that many sciences have, as pointed out by Kahneman and Tversky, treated 'all forms of uncertainty in terms of a single dimension of probability or degree of belief.'[8] Although there are several suggested taxonomies of uncertainty, understanding the duality of probability is consequently a necessary first step.

Probability as a concept is more novel than doubt and uncertainty, but perhaps also more difficult to understand. Bertrand Russell once stated that 'probability is the most important

---

[5] Walker, Lempert, and Kwakkel, "Deep Uncertainty," 395.
[6] Walker et al., "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support," 8.
[7] Knight, *Risk, Uncertainty and Profit*, 19–20.
[8] Kahneman and Tversky, "Variants of Uncertainty," 143.

concept in modern science, especially as nobody has the slightest notion of what it means.'[9] The birth of probability as we understand it today is often attributed to letters between Pierre de Fermat and Blaise Pascal in 1654 regarding a game of chance.[10] Simultaneously, he voiced ideas of probabilities concerned with reasoning under uncertainty such as his famous wager regarding the existence of God.[11] Leibniz, followed later by Hume, Bayes, and Bernoulli, promoted an epistemic view on probability and uncertainty.[12] However, Laplace's work *Philosophical Essays on Probability* of 1814 is considered one of the most central works of the classical theory of probability. Here he formulated what is today known as Laplace's demon, an all-knowing intelligence where '…for it, nothing would be uncertain and the future, as the past, would be present to its eyes.'[13] With this complete deterministic system as the foundation, Laplace further argued that the human mind only to a limited degree could resemble this vast intelligence and we thus had to rely on the relative nature of probability, as our uncertainty is due to lack of knowledge.[14] Laplace's theory of probability is then a measurement of our ignorance:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favourable to the event whose probability is sought.[15]

From this we get the classical definition of probability P(A)=*m/n* where P(A) is the probability of event A, *n* is all the equally possible cases and *m* all the cases favourable to outcome A. When throwing a die there are six equally possible cases and three cases favourable to getting an odd number, resulting in a probability of 0.5. Probability theory is, according to Laplace, '… at bottom only common sense reduced to calculus.'[16]

Probability theory has evolved since the time of Laplace's writing and can largely be divided into two opposite main schools of thought. Already around 1840, the classical theory of probability was questioned by what can be called revisionists; people seeing the need for a

---

[9] Kahneman and Tversky, "Variants of Uncertainty." 105.
[10] Donald Gillies, *Philosophical Theories of Probability*, 4.
[11] Arnauld and Nicole, *Logic or the Art of Thinking*.
[12] Gillies, *Philosophical Theories of Probability*, 13.
[13] Pierre Simon Marquis de Laplace, *A Philosophical Essay on Probabilities*, 4.
[14] Laplace.
[15] Laplace.
[16] Laplace, 196.

distinction between two types of probability; the objective and the subjective.[17] The early revisionists did not necessarily abandon determinism, but saw the need to differentiate between, in Cournot's words 'the double sense of probability, which at once refers to a certain measure of our knowledge, and also to a measure of the possibility of things [*possibilité des choses*] independently of the knowledge we have of them.'[18] Hacking calls this duality the Janus-face nature of probability, named after the two-faced Roman God.[19] One the one hand, probability can be seen as a feature of the objective world, involving chance, frequency, and the tendency for a specific outcome to occur. On the other hand, probability is something subjective, a measurement of a person's belief in something, how close one think one is to the true answer or how much one expects something. Over time these two views have been given different labels. Antoine Augustin Cournot was perhaps the first one to have properly distinguished between an 'objective possibility' and a 'subjective probability' in 1843.[20] In 1934 Popper, following Cournot, divided probability into an objective interpretation presenting a statement of relative frequency and a subjective interpretation, describing feelings of certainty or uncertainty.[21] Later, Hacking categorised probability into aleatory, after the Latin word for rolling a dice, and epistemic, after the Greek word for knowledge, although he later labels the two frequency-type and belief-type probabilities respectively.[22] Gillies would rather define the two objective and epistemological while Taleb calls them ontological and epistemic.[23] To confuse the matter more, there are also sub-categories, largely divided between natural and social sciences

Aleatory orientations of probability calculus

Even within the two main schools of probability, several different theories have been advocated. One of the most common theories of probability within the objective school is the frequency theory that interprets probability to be the relative frequency of outcomes in

[17] Daston, "How Probabilities Came to Be Objective and Subjective." 330.
[18] Cournot pp. 4-5, in Daston. 331.
[19] Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. 12.
[20] Antoine Augustin Cournot (1843), Exposition de la théorie des chances et des probabilités (Paris, 1843) pp. 86-87, in Daston, *Classical Probability in the Enlightenment*. 190-191.
[21] Popper (1934), section 48
[22] Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*.12; Hacking, *An Introduction to Probability and Inductive Logic*. 140.
[23] Gillies, *Philosophical Theories of Probability*. 2.

the long run. The theory originated from Bernoulli's Theorem in the 17th century and was further developed by especially Von Mises, who stated that probability is a measurement of 'the relative frequencies of certain attributes [that] become more and more stable as the number of observations is increased.'[24] The sample mean of observed results of similar experiments will get closer and closer to a fixed value but genuine empirical collectives are still essentially random.[25] When throwing a die 1000 times the probability of an odd number will be close to 0.5 while the probability of getting an even number the next time will still be 0.5 regardless of how many prior odd numbers one got. Probability is thus derived objectively, independent of the person making the estimate.[26] The frequency theory has two main problems, finding a realistic limit to the frequency and what and how much data to include in the relevant reference class collective.[27] Finding relevant collectives with long term data in intelligence can be difficult at best and misleading at worst. How can we really know that the data we are gathering in a collective in fact represent the same phenomena and have the same root causes? While wars and conflicts may share similarities, they are also unique in their own ways, being examples of situations that fall outside the scope of the frequency theory of probability.[28] An element of the reference class problem is hence the problem of finding probabilities for single events.

The other aleatory interpretation is the propensity theory. Popper argued that instead of basing probability on the frequency of occurrence, probability should be interpreted to be a property or disposition of an experiment's generating conditions, thus being relational properties like Newtonian forces.[29] These generating conditions are part of the experimental setup and henceforth propensities for the occurrence of single events, but also the propensity for the virtual statistical frequency if one were to repeat the experiment under the exact same conditions. If the conditions change, so will the outcome. The result of a coin toss is not only due to the property of the coin itself, but also to the properties of the complete coin-tossing set-up. Consequently, one can get biased results despite using a fair

---

[24] Von Mises, *Probability, Statistics, and Truth*, 12; Daston, *Classical Probability in the Enlightenment*, 234–35; Hacking, *An Introduction to Probability and Inductive Logic*, 194.
[25] Keynes, *A Treaties on Probability*, 336; Von Mises, *Probability, Statistics, and Truth*, 25–28.
[26] Gillies, *Philosophical Theories of Probability*, 89–90.
[27] Keynes, *A Tract on Monetary Reform*, 80; Reichenbach, *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*, 374.
[28] Von Mises, *Probability, Statistics and Thruth*, 11–15.
[29] Popper, "The Propensity Interpretation of Probability," 5, 37–38.

coin if the different properties of the coin-tossing system are rigged to give a different probability of heads than 0.5. The propensity theory is widely debated and alternative explanations have been promoted.[30] The propensity theory is also mainly interpreted as a way to explain aleatory, or random, uncertainty such as in quantum mechanics even though there are those that claim a relation to epistemic, or knowable, uncertainty.[31] The propensity theory is nonetheless easier to understand in a natural science like quantum physics, than as a tendency for a country's election process to yield a certain outcome or the propensity for a war breaking out between two states.

Epistemic orientations of probability calculus

One of the early epistemic probability theories is the logical theory which interprets probability to be a measurement of our rational objective degree of belief. It is an extension of inductive logical reasoning, resulting in a probability-relation between the evidence and the conclusion, what Hacking calls inference to a plausible explanation.[32] A conclusion will be relative to the evidence, making probability both conditional ($P(h|e)$), but also possible to determine *a priori*, in no need for any experiments. Since degree of belief can be difficult to measure, Keynes introduced the idea that not all probabilities needed to have a numerical value. In this qualitative approach, there is only a partial order of the set of probabilities as '…it is not even clear that we are always able to place them in an order of magnitude.'[33] However, if assigning numbers to probabilities in cases of equiprobable alternatives, the sum of different probabilities can be more than 1, and the same event can end up with different probabilities dependent upon the order of the equation.[34] The theory moreover presupposes the existence of a universal, Platonic degree of knowledge where logical relations are known by direct acquaintance.[35] If one does not have direct knowledge, one must make assumptions, which again can lead to biased judgements.

[30] Berkovitz, "The Propensity Interpretation of Probability"; Gillies, *Philosophical Theories of Probability*.
[31] Ülkümen, Fox, and Malle, "Two Dimensions of Subjective Uncertainty: Clues from Natural Language."
[32] Keynes, *A Treaties on Probability*, 5; Hacking, *An Introduction to Probability and Inductive Logic*, 16–18.
[33] Keynes, *A Treaties on Probability*, 29.
[34] Gillies, *Philosophical Theories of Probability*. 37, 42.
[35] Gillies. 33

The logical interpretation is not a widely held theory today, although there are those calling for its revival. Williamson argues that single-case probabilities of external events can be worked out by using objective Bayesianism where probability is a measurement of logical relations between the evidence and the conclusion, determined by one's background knowledge.[36] The problem is that intelligence analysts seldom have all the necessary background knowledge, especially when making judgements regarding the course of action of a policymaker from a completely different culture. What may seem logical for the policymaker may seem illogical for the analyst and vice versa. This form of ethnocentric bias has been seen in several judgements, like the Indian nuclear test and the Iraqi WMD programmes. Working out a mathematical answer using Bayes rule will hardly remedy such a bias, making the logical theory irrelevant for intelligence analysts.

The problems arising from the logical theory led to the introduction of the subjective theory where probability is defined as a personal degree of belief based upon the information available to you, thus abandoning the assumption of rationality leading to unanimity.[37] The immediate critique is that this can lead to people believing anything. However, following the Dutch book argument, the degree of belief must still cohere to the mathematical laws of probability, such as that mutually exclusive events must add up to 1.[38] Opposed to the frequency theory that treats probability as the result of long-run experiments of independent events, in the subjective theory the probability of outcome A is conditional to the occurrence of B. One is therefore also measuring the different variables instead of only the discrete outcome. Subjective probability is typically worked out mathematically by using Bayes theorem. The basics of Bayesian thinking is to specify a probability of an event based upon prior knowledge of the conditions related to the event and then to update that probability model based upon new observed data of these conditions.[39] Even though two people may have very different *a priori* beliefs of the probability of an event, they must satisfy the Dutch book requirement when they revise their probabilities in light of new evidence for or against their hypotheses. Consequently, the more, new evidence that is used for the calculations, the more the different degrees of belief tend to converge to the same

[36] Williamson, "Philosophies of Probability : Objective Bayesianism and Its Challenges." 14.
[37] Gillies, *Philosophical Theories of Probability*. 53.
[38] Ramsey, "Truth and Probability."
[39] Gill, *Bayesian Methods*. 6

probability, especially if the new evidence all pull in the same direction. The subjective theory can therefore be said to be a way of evading the problem of induction by learning from experience.[40] Moreover, it is useful to mitigate the impact of what is known as the base rate fallacy that often leads to false positives.

The subjective theory has faced criticism from non-Bayesians through a two-folded argument. First, Bayesian probability is arguably an automatic inference machine working independently of different settings, and second, it is dependent upon a subjective, unclear inference of knowledge instead of an objective knowledge.[41] Nevertheless, the subjective approach using Bayesian probability can be well adapted to intelligence analysis by following some key features: quantification of probabilistic judgements instead of using verbal terms, giving evidence a probability based on its diagnostic value, and addressing probability one evidence at the time against each hypothesis instead of at the end based on the full body of evidence.[42]

As we can see from the passages above, there is an agreement as to the mathematics of probability, especially that the whole probability space cannot exceed 100%. The main problem with the probability calculus is nevertheless that 'probability is a fundamental tool for inductive logic' while not necessarily so for abductive logic.[43] Probability is as such a measurement of a risky argument regarding a single outcome and accordingly perhaps not the best measurement of inference to the best explanation between several alternative outcomes. Is then the probability calculus suitable to convey uncertainty in estimative intelligence?

Uncertainty and probability – same same, but different?

Judgments under uncertainty are not confined to formal theories of probability but are equally important in social sciences as well as everyday life situations. Research in fields like decision theory and psychology has led to a somewhat different categorization of objective

[40] Hacking, *An Introduction to Probability and Inductive Logic*. 256.
[41] Gelman, "Objections to Bayesian Statistics."
[42] Zlotnick, "Bayes' Theorem for Intelligence Analysis."
[43] Hacking, *An Introduction to Probability and Inductive Logic*, 14, 17.

and subjective variants of uncertainty, suggesting a dual philosophical construct that challenges the single notion. Kahneman and Tversky divides uncertainty into two main categories, external (disposition) and internal (ignorance), and with two sub-categories in each.[44] The variants can then be identified based on the information used to evaluate the probability. The external, or dispositional, uncertainty attribute is either assessed distributional based on frequencies, or singular based on propensities of the case at hand. The internal, or ignorance, uncertainty attribute is either assessed through reason based on the weight of evidence and arguments, or is assessed introspectively, being a measurement of one's confidence in a particular statement or association. Three of these variants seemingly corresponds with the more formal probability theories, making one ask if there really is any difference?

Teigen, on the other hand, argues that people either makes judgments rule-dictated, using the probability calculus, or intuitively, 'from the weight of arguments, feelings of conviction, degree of knowledge, or from whatever judgemental heuristic that seems to be most appropriate for evaluating the phenomenon under consideration.'[45] Since people often have problems dealing with the probability calculus, they leave the world of rules and have to choose from a family of intuitive approaches. By applying an indetermination/determination distinction cutting across the external/internal dimension, he presents four different modes of assessing uncertainty: chance, dispositional, confidence and uncertainty by ignorance.[46]

	External	**Internal**
Emphasis on indetermination	I: Chance	IV: Ignorance, uncertainty
Emphasis on tendency	II: Disposition, propensity	III: Confidence, belief

Figure 1: Teigen's modes for assessing uncertainty

All these modes have heuristics and biases connected to them.[47] For chance probabilities the most important are representativeness and significance heuristics, resulting in a tendency to

[44] Kahneman and Tversky, "Variants of Uncertainty."
[45] Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases.," 212.
[46] Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases."
[47] Teigen.

discard chance probabilities and trying to find other reasons when low probability events do occur. For dispositional probabilities base rate neglect and violation of the complementary rule result in a tendency to treat probability as an attribute of a specific outcome instead of being considered in proportion to the set of possible outcomes. As a result, probabilities are not revised even after the introduction of new evidence, and one tends to overestimate the probability of specific outcomes. Confidence, being a measurement of one's subjective conviction, is often a second-order judgement of an already made external probability of a chosen hypothesis. Being influenced by availability and vividness heuristics, confidence judgements are often prone to suffer from the *a priori* fallacy, making it easy to go from the easily thinkable to the easily believable. Opposed to confidence, which is the result of a hypothesis testing process, uncertainty by ignorance is a result of a hypotheses generation process where one can be undecided to choose between hypotheses due to lack of diagnostic evidence. Furthermore, while the amount of information often will increase one's confidence, it can at the same time also increase one's uncertainty as one becomes aware of more possible outcomes.[48] Uncertainty by ignorance can therefore be said to be a result of Aristoteles' the more you know the more you realize you do not know.

The duality of probability between the aleatory, or objective, interpretation and the epistemic, or subjective, interpretation is not only associated with different theories of probability but from decision theory literature we also know that the differences affect how we understand probability and uncertainty.[49] The duality have consequently implications for how the intelligence community should communicate uncertainties in estimative products. Gillies argues that his propensity theory is suitable for physical sciences and subjective probability for social sciences.[50] The frequency interpretation is irrelevant, and it makes no sense to talk about analysts' batting average.[51] Since intelligence analysts usually try to make judgements about tendencies of future events, one can instead argue that estimative intelligence analysis should use the propensity approach. However, intelligence is nothing

[48] Peterson and Pitz, "Confidence, Uncertainty, and the Use of Information."
[49] Kahneman and Tversky, "Variants of Uncertainty"; Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases."
[50] Gillies, *Philosophical Theories of Probability*.
[51] Smith, "On the Accuracy of National Intelligence Estimates"; Marchio, "'How Good Is Your Batting Average?' Early IC Efforts To Assess the Accuracy of Estimates."

close to quantum mechanics, leaving the probability judgement susceptible to whatever heuristics analysts deemed appropriate at the time. Rosenberg conversely argues that it is natural for intelligence analysts to follow the subjective logical approach because one is trying to predict an indeterminate future reality where the analyst must rely on his or her intuition.[52] This argument comes quite naturally since the traditional intelligence analysis methodology has been one of induction. In a contradictory view, to remedy the heuristics and biases following inductive reasoning, there are those advocating that probability in estimative intelligence should be a result of the subjective approach using Bayes theorem.[53] Bayesian probability has, despite its many advocates, yet to get foothold within the environment of intelligence analysts. A main reason may be the lack of a common understanding of probability theory among analysts since they are recruited from many different sciences. Moreover, intelligence analysts are generally not taught a common intelligence methodology nor techniques of how to work out and represent probabilities in estimative intelligence. The lack of a common standard for handling uncertainty in intelligence has not only affected how analysts approach uncertainty and probability but has also affected how uncertainties are presented to policymakers, increasing the chances of discourse failure in the policymaker-intelligence interface.

The estimative language – ambiguous verbal expressions of probability and uncertainty

To mitigate discourse failure, intelligence producers must make sure policymakers understand the message passed on. Creating a common understanding can be a difficult task, especially when trying to describe the uncertainties connected to key judgements in estimative intelligence products. Estimative intelligence products not only need to address uncertainties of information gaps and assumptions, but also uncertainties of the range of possible outcomes. Misunderstanding or manipulating terms of uncertainty are frequent objects of attention, for instance in the British government's decision to support the United States in its 2003 invasion of Iraq.[54] As a consequence of such discourse failure, one of the

[52] Rosenberg, "The Interpretation of Probability in Intelligence Estimation and Strategic Assessment."
[53] Karvetski et al., "Structuring and Analyzing Competing Hypotheses with Bayesian Networks for Intelligence Analysis." Zlotnick, "Bayes' Theorem for Intelligence Analysis."
[54] Chilcot, "The Report of the Iraq Inquiry: Executive Summary."

key requirements stated in the IRTPA was for the intelligence community to '…properly caveat and express uncertainties or confidence in analytic judgments.'[55] The Butler Review followed suite, recommending the intelligence community to improve the language of JIC assessments.[56] The communication of uncertainty is nevertheless still an object of debate, not only in intelligence but in climate research, decision science and not the least, in psychology.

The use of verbal expressions of probability and uncertainty have been given different labels, such as probabilistic phrases, probability terms, verbal probabilities, and words of estimative probability. The study of these terms has generally followed two different approaches, one from a behavioural decision theory point of view and one from a cognitive psychology point of view.[57] Many of the conclusions are nonetheless comparable. Despite that some probability terms have been shown to correspond to different numerical levels of probability, for instance that probable is understood as approximately a 70% probability, there are still large individual differences regarding how people interpret and communicate verbal probability phrases. People receiving a verbal probability often interpret these as being elastic and more imprecise than what was intended by those communicating the phrase.[58] Sherman Kent found that there were three very different interpretations of the statement 'serious possibility' for a Soviet attack on Yugoslavia.[59] Not only was there a different understanding between the intelligence producer and the policy planners, but also within the different members of the Board of National Estimates, ranging from a 20% to an 80% probability. Some years later, a similar result was found among NATO officers.[60] Although a few phrases, like 'better than even' were interpreted in much the same ways, there were large variations to the numeric probabilities the different phrases were given. Even though one could think that people would have a better understanding of verbal probabilities today, contemporary research has reached similar conclusions. A multi-national study with samples from 24 countries and covering 17 languages found that people

[55] United States Congress, Intelligence reform and terrorism prevention act of 2004. sec 1019.
[56] Lord Butler, "Review of Intelligence on Weapons of Mass Destruction."
[57] Budescu and Wallsten, "On Two Complenetary Appraoches to the Study of Verbal Probabilities," 37.
[58] Brun and Teigen, "Verbal Probabilities: Ambiguous, Context-Dependent, or Both?"
[59] Kent, "Words of Estimative Probability."
[60] Barclay et al., *Handbook for Decisions Analysis*, 67–68.

underestimated the high probabilities and overestimated the low probabilities that were presented by the Intergovernmental Panel on Climate Change.[61]

Verbal probabilities are not only ambiguous, our understanding of the different phrases is also context dependent. Research suggests that different interpretations of verbal probabilities is also due to differences in knowledge, ideology, opinion, or other self-serving interpretations.[62] Furthermore, a severity effect impacts people to feel that a specific verbal probability for a certain event represents a higher numerical probability the more severe the outcome is perceived. A three feet rise in sea level was given a higher probability when said to produce flooding than when the effects were said to be negligible.[63] The understanding of verbal probabilities is furthermore affected by the directionality of the probabilistic message, dependent on whether the term is positively or negatively framed.[64] Positive phrases are rated as more optimistic and more accurate than negative phrases with the result that positive phrases are perceived to convey a higher probability than warranted and vice versa for negative phrases. Hence, 'a slight possibility' is perceived having a higher probability than 'unlikely' even when the two phrases are thought to correspond to the same probability. There are also other framing effects. Modal words like 'can' and 'will' are especially troublesome to use as 'will' is associated with 'certain' but also with 'at least', while 'can' is associated with 'possibly' but also outcomes of low probability.[65] The result is that, depending on frame, the same word can be associated with both high and low numeric values.

Due to the variability of how people interpret verbal probabilities it has been difficult to create a universal meaning of different terms. The use of verbal probability phrases in the intelligence community has been very inconsistent since the 1950s, indicating that analysts

[61] Budescu et al., "The Interpretation of IPCC Probabilistic Statements around the World."
[62] Brun and Teigen, "Verbal Probabilities: Ambiguous, Context-Dependent, or Both?"; Budescu, Por, and Broomell, "Effective Communication of Uncertainty in the IPCC Reports."
[63] Harris and Corner, "Communicating Environmental Risks: Clarifying the Severity Effect in Interpretations of Verbal Probability Expressions."
[64] Teigen and Brun, "Verbal Probabilities: A Question of Frame?"
[65] Teigen and Filkuková, "Can>Will: Predictions of What Can Happen Are Extreme, but Believed to Be Probable"; Løhre and Teigen, "How Fast Can You (Possibly) Do It, or How Long Will It (Certainly) Take? Communicating Uncertain Estimates of Performance Time"; Teigen, Juanchich, and Riege, "Improbable Outcomes: Infrequent or Extraordinary?"

seemingly have not given much attention to how they have communicated uncertainty to policymakers. Kesselman found that the words ‘will’ and ‘would’ were the most used estimative phrases in NIEs from the 1950s through the 2000s, pointing towards an excessive use of deterministic rather than probabilistic judgements as well as a preference for using modal words instead of verbal probability phrases.[66] Furthermore, instead of using probabilistic phrases, key judgements have used phrases like 'we assess' or 'we judge'. Consequently, intelligence estimates filled with modal words and other non-probabilistic phrases are adding to the confusion rather than functioning as a useful tool to mitigate discourse failure.

Using numbers instead of words has been suggested as a remedy. For a long time, numeric probabilities were preferred over verbal probabilities among decision theorists.[67] As a response to White House criticism, the CIA experimented with the use of both Bayes rule and numeric probabilities for a short period in the 1970's.[68] Lately, the use of numeric probabilities has been promoted by several researchers within the field of intelligence.[69] The core argument is that numbers will reduce ambiguity and increase accuracy and that the drawbacks of using numbers alone can be mitigated through training. On the other hand, other researchers have found that although people prefer to receive numbers, they also prefer to present verbal probability phrases.[70] Moreover, there are those who claim that presenting subjective probabilities in numbers gives an unjustified impression of accuracy.[71] Third, people tend to perceive an expected outcome more likely than an unexpected one

---

[66] Kesselman, “Verbal Probability Expressions in National Intelligence Estimates : A Comprehensive Analysis of Trends from the Fifties through Post 9 /11,” 62–66.

[67] Wallsten and Budescu, “A Review of Human Linguistic Probability Processing - General-Principles and Empirical-Evidence.”

[68] Marchio, “‘ If the Weatherman Can ...’: The Intelligence Community ’ s Struggle to Express Analytic Uncertainty in the 1970s.”

[69] Barnes, “Making Intelligence Analysis More Intelligent: Using Numeric Probabilities”; Mandel and Barnes, “Accuracy of Forecasts in Strategic Intelligence”; Dhami et al., “Improving Intelligence Analysis With Decision Science”; Dhami and Mandel, “Words or Numbers? Communicating Probability in Intelligence Analysis.”; Rieber, “Intelligence Analysis and Judgmental Calibration”; Chang, Chen, and Mellers, “Developing Expert Political Judgment: The Impact of Training and Practice on Judgmental Accuracy in Geopolitical Forecasting Tournaments”; Mandel and Barnes, “Geopolitical Forecasting Skill in Strategic Intelligence.”

[70] Brun and Teigen, “Verbal Probabilities: Ambiguous, Context-Dependent, or Both?”

[71] Teigen, “Risk Communication in Words and Numbers,” 244; Marchio, “‘ If the Weatherman Can ...’: The Intelligence Community ’ s Struggle to Express Analytic Uncertainty in the 1970s,” 37.

despite both being presented with the same numerical probability.[72] Lastly, as discussed in chapter 1, the accuracy debate in estimative intelligence is of little value. While secrets deal with the past and the present and can be judged accurately, mysteries and complexities are future and contingent where accuracy can only be proven retrospectively, and outcomes can change when policymakers act. A key objective of the current intelligence reform is consequently reduced ambiguity, not increased accuracy in itself.

Research has nevertheless proven that verbal probabilities can be meaningfully scaled as part of a membership function along the [0,1] probability interval rather than as a point value along the same scale.[73] An early example of a scale representing membership functions is Wallsten et al.'s scale.[74]

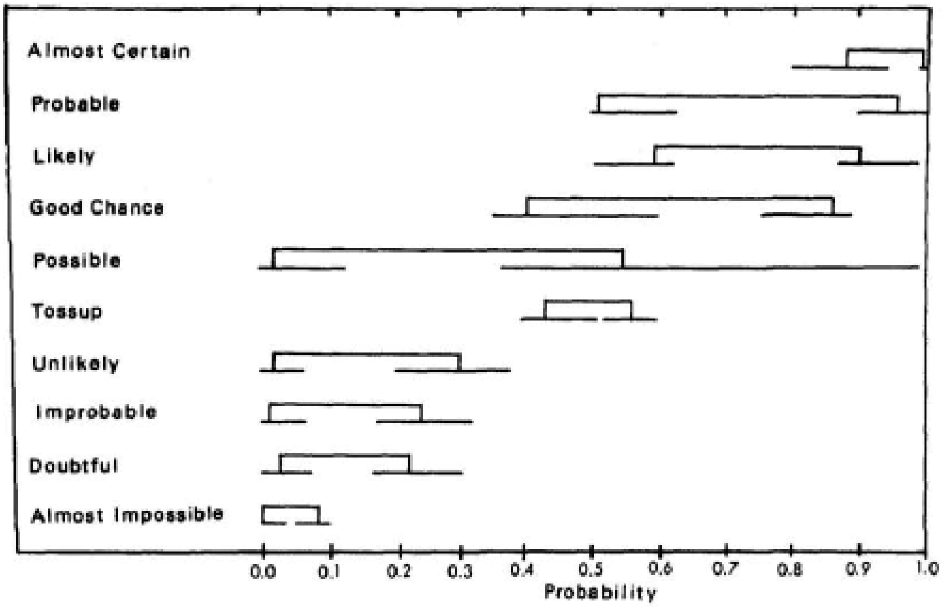


Figure 2: First, second and third quartiles over subjects of the upper and lower probability limits for each phrase in Experiment 1 of Wallsten et al.

David L. Wark came to a similar result two decades earlier when surveying both analysts and policy staff members on how different estimative expressions were understood as numeric probabilities.[75] Recent research has found that combining verbal phrases with numeric intervals resulted in increased distinction of different terms, the terms were interpreted

[72] Windschitl and Weber, “The Interpretation of ‘Likely’ Depends on the Context, but ‘70%’ Is 70% - Right? The Influence of Associative Processes on Perceived Certainty.”
[73] Wallsten and Budescu, “A Review of Human Linguistic Probability Processing - General-Principles and Empirical-Evidence.”
[74] Wallsten et al., “Measuring the Vague Meanings of Probability Terms.”
[75] Wark, “The Definitions of Some Estimative Expressions.”

more consistently and perhaps most importantly, the interpretation was more consistent with the reporting agency's guidelines.[76] This view was already proposed by Sherman Kent in the 1950s when he and Max Foster worked on a chart implementing a dual scale to be used in estimative intelligence analysis. The subsequent 'Odds Table', consisting of five verbal probabilities that represented 'The General Area of Possibility', was an early attempt of implementing a standard set of verbal probabilities in the intelligence community.[77]

100% Certainty		
The General Area of Possibility		
93%	give or take about 6%	Almost certain
75%	give or take about 12%	Probable
50%	give or take about 10%	Chances about even
30%	give or take about 10%	Probably not
7%	give or take about 5%	Almost certainly not
0% Impossibility		

Figure 3: Kent's 'Odds Table'

Unfortunately, Kent's table took a long time to get a strong foothold within the intelligence community. At the time, he met strong resistance from “poets”, the ones believing ‘…the most a writer can achieve when working in a speculative area of human affairs is communication in only the broadest general sense.’[78] The “poets” also won over the “mathematicians” after CIA's brief experiment with numeric probabilities in the 1970's.[79] The “poets’” standpoint has been strong and the explanation of the estimative language used the first post-IRTPA NIEs released to the public included only five verbal probabilities, ranging from ‘Remote’ to ‘Almost certainly’, on a scale in different shades of grey.[80] However, the NIE scale changed by the time the *Intelligence Community Assessment 2017-01D: Assessing Russian Activities and Intentions in Recent US Elections* was released to the

[76] Budescu, Por, and Broomell, “Effective Communication of Uncertainty in the IPCC Reports.”
[77] Kent, “Words of Estimative Probability.”
[78] Kent.
[79] Marchio, “‘ If the Weatherman Can ...’: The Intelligence Community ’ s Struggle to Express Analytic Uncertainty in the 1970s,” 37.
[80] National Intelligence Council, “National Intelligence Estimate July 2007 - The Terrorist Threat to the US Homeland”; National Intelligence Council, “National Intelligence Estimate January 2007 - Prospects for Iraq’s Stability: A Challenging Road Ahead.”

public. Not only does the scale now contain more verbal probabilities, but the phrases are also placed on a numeric scale showing roughly equivalent numeric probabilities.[81]

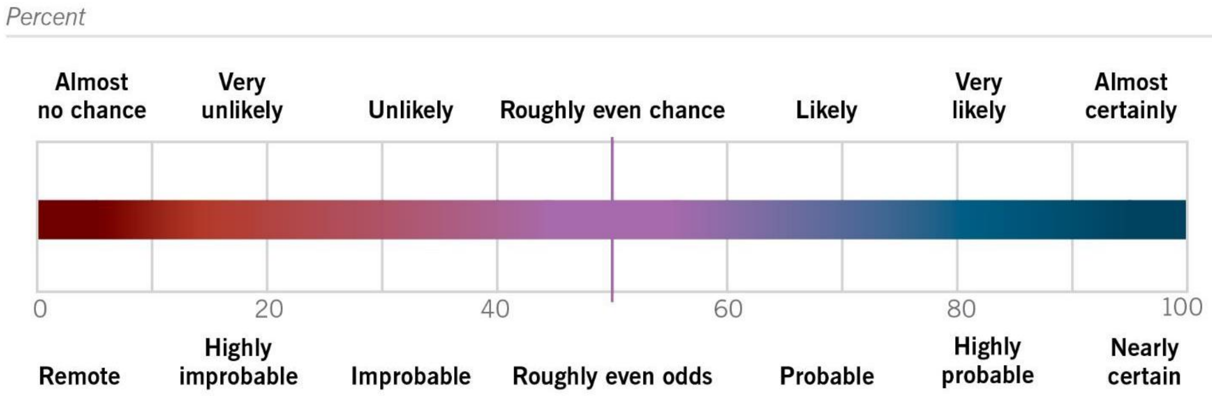


Figure 4: Likelihood scale in ICA 2017-01D

The intermittent focus on a dual scale is evident from other parts of the American intelligence community as well, especially the American military. In 1996, the American "Joint Publication 2-0: Joint Intelligence" introduced a scale to be used in estimative intelligence products.[82] This chart has later been revised several times, sometimes being a dual scale, but in its latest edition the numeric equivalents are left out.[83] Consequently, the use of a dual scale in the American intelligence community is still controversial, despite research results.

There have been some attempts outside the American intelligence community at implementing a common scale for verbal probabilities where the different phrases correspond to numeric ranges. The British Probability Yardstick was updated in March 2019 and consists of seven qualitative statements with an associated probability range.[84] The Yardstick has an intentional 5% gap between the levels to '…encourage analysts to be clear about what their assessments mean.'[85] In NATO, the first scale was introduced in NATO

[81] Office of the Director of National Intelligence, "Intelligence Community Assessment 2017-01D: Assessing Russian Activities and Intentions in Recent US Elections."
[82] Joint Chiefs of Staff, "Joint Publication 2-0 - Joint Intelligence."
[83] Joint Chiefs of Staff.
[84] Professional Head of Intelligence Assessment, "Professional Development Framework for All-Source Intellgience Assessment."
[85] Omand, *Securing the State*, 202.

doctrine as late as in June 2016, consisting of five phrases but with uneven probability ranges.[86] In Norway there have been several different scales in use, depending on the agency. The first authoritative scale to be used in the Norwegian Defence was issued as part of the first intelligence doctrine in 2013, a scale that was adapted from the American "Joint Publication 2-0" with all the problems that entailed.[87] The Norwegian police adopted the defence scale, whose use, as presented in the first chapter, led to a public debate in Norway regarding the use of verbal probabilities in threat assessments. One result of that debate was a new national standard for verbal probabilities, first issued in the Norwegian Security Police intelligence doctrine issued in 2017.[88] Although this table also included the new NATO standard with its numeric scale, a footnote made it clear that probabilistic statements were the result of a qualitative assessment and not of a mathematical calculation.[89] The attempt for a national standard was short lived as not all agencies followed the new national standard, culminating in a new, slightly different scale in the newly updated defence intelligence doctrine.[90]

The ambiguity of verbal probabilities and different scales aside, a main problem analysts face when assessing uncertainty is, as stated previously, that analysts lack guidelines for calculating probabilities. Neither analytic standards nor intelligence doctrines address this impediment, leaving it up the judgemental heuristics each analysts find appropriate. To mitigate discourse failure the intelligence community continues to need to examine methods suitable for a common framework of assessing and communicating uncertainties in estimative intelligence products. One step forward is to separate probabilities from analytic confidence.

From eliminating to assessing uncertainty by expressing analytic confidence

The traditional way of understanding intelligence has been to improve decision making by reducing uncertainty, for the most part with products that consist of two central parts, an

[86] NATO, "NATO Standard AJP-2.1(B): Allied Joint Doctrine for Intelligence Procedures."
[87] Forsvaret, "Etterretningsdoktrinen," 2013.
[88] Politiets sikkerhetstjeneste (PST), "Etterretningsdoktrine for PST."
[89] Politiets sikkerhetstjeneste (PST), 42.
[90] Forsvaret, "Etterretningsdoktrinen," 2021.

explanatory base and a forecast giving a probabilistic statement of a future outcome. Any distinction between an objective probability and a subjective confidence has historically been missing in probabilistic judgements in estimative intelligence products. The wording in the IRTPA did not help in clarifying the matter as it does not mention anything about probabilistic statements. Instead, the law adds to confusion by demanding that intelligence products need to '… properly caveat and express uncertainties *or* confidence in analytic judgments.'[91]

Friedman and Zeckhauser make the argument that instead of attempting to eliminate or reduce uncertainty, the aim of estimative intelligence products should be to assess uncertainty.[92] The reason is that attempts at eliminating uncertainty usually include a conflation of probability and confidence, resulting in reduced accuracy, clarity, and utility of these products. Assessing uncertainty, on the other hand, allows for the active management of both probability and confidence.[93] How forecasts are communicated will impact policymakers' understanding. Assessing uncertainty will facilitate a more explicit probability statement as well as an explanation of the basis for that probabilistic judgement. External (objective) and internal (subjective) uncertainties have different language markers in natural language use situations. Probability statements like 'It is likely' and 'There is a 70% probability' reflect an objective uncertainty caused by random processes or causal factors in the external world. Confidence statements like 'I am reasonably confident that…' and 'I am 70% certain' reflect a subjective uncertainty that stems from a person's internal level of knowledge or degree of conviction.[94] The difference is therefore a distinction between how likely is it versus how certain am I?

The American intelligence community attempted to implement this change in praxis of communicating uncertainty when issuing the first analytic standards. The first edition of the "Intelligence Community Directive (ICD) 203: Analytic Standards" stated that 'Analytic

[91] United States Congress, Intelligence reform and terrorism prevention act of 2004, sec. 1019. (b), (2),(A). Author's use of underlining and italics.
[92] Friedman and Zeckhauser, "Assessing Uncertainty in Intelligence."
[93] Friedman, *War and Chance: Assessing Uncertainty in International Politics*, 58–68.
[94] Teigen, "The Language of Uncertainty"; Løhre and Teigen, "There Is a 60% Probability, but I Am 70% Certain: Communicative Consequences of External and Internal Expressions of Uncertainty"; Fox and Ülkümen, "Distinguishing Two Dimensions of Uncertainty."

products should indicate both the level of confidence in analytic judgments and explain the basis for ascribing it.'[95] Soon after, page five in the National Intelligence Estimate "Prospects for Iraq's stability: A Challenging Road ahead", contained an explanation of the estimative language used.[96] Page five explained not only the words used to convey the probability for key judgements, but also a description of the three levels of analytic confidence, high, medium and low, that were to be ascribed to these judgements. The different confidence levels were described to be based upon the scope of the analytic issue and the quality of the available information, where High confidence would indicate a solid basis for a judgement while Low confidence meant a scant, questionable, or fragmented basis. Ten years later, the ICA 2017-01D added source collaboration, meaning the quantity of supporting evidence, as an additional factor for assessing analytic confidence.[97]

Confidence in the Sources Supporting Judgments. Confidence levels provide assessments of the quality and quantity of the source information that supports judgments. Consequently, we ascribe high, moderate, or low levels of confidence to assessments:

- **High confidence** generally indicates that judgments are based on high-quality information from multiple sources. High confidence in a judgment does not imply that the assessment is a fact or a certainty; such judgments might be wrong.
- **Moderate confidence** generally means that the information is credibly sourced and plausible but not of sufficient quality or corroborated sufficiently to warrant a higher level of confidence.
- **Low confidence** generally means that the information's credibility and/or plausibility is uncertain, that the information is too fragmented or poorly corroborated to make solid analytic inferences, or that reliability of the sources is questionable.

Figure 5: ICA 2017-01D Annex B Levels of confidence[98]

The concept of analytic confidence is only recently added into a few intelligence doctrines, and some, like the British standards, have no reference to analytic confidence at all. When mentioned, analytic confidence is presented without any objective guidance on how to assess it, adding ambiguity instead of clarification to the concept.[99] Then, what methods can analysts use to assess analytic confidence? The 2015 issue of ICD 203 states that confidence

[95] Office of the Director of National Intelligence, "Intelligence Community Directive 203: Analytic Standards," 2007.
[96] National Intelligence Council, "National Intelligence Estimate January 2007 - Prospects for Iraq's Stability: A Challenging Road Ahead."
[97] Office of the Director of National Intelligence, "Intelligence Community Assessment 2017-01D: Assessing Russian Activities and Intentions in Recent US Elections."
[98] Office of the Director of National Intelligence.
[99] Office of the Director of National Intelligence, "Intelligence Community Directive 203: Analytic Standards," 2015; NATO, "NATO Standard AJP-2.1(B): Allied Joint Doctrine for Intelligence Procedures."

‘may be based on the logic and evidentiary base that underpin it’ and mentions that products should make notice of causes of uncertainty, such as quality and quantity of the available information and knowledge gaps. [100] For analysts to improve their ability to assess uncertainty, standards and doctrines need to clarify what analytic confidence constitutes, the factors that impact it, and how it can be presented to policymakers. Once again, looking towards other fields may bring the debate forward.

Confidence is used for several topics within different fields like psychology, decision science, medicine, and not least natural science. Within psychology, decision science and medicine much research has looked into the issue of being under- or overconfident, usually as part of evaluating the correctness of one's knowledge or the basis for making a correct judgement where there exists a true answer.[101] Weiss therefore suggests looking to other sciences and to make use of either Tukey box plots or fuzzy pie charts to depicts zones of uncertainty.[102] In statistics, however, confidence refers to a statement of an interval of scores of which a certain percentage, typical 95%, of the mean scores fall within.[103] This can then be understood as the probability that the mean values of future experiments will lay within the stated interval. Although one can point to these issues, they cannot alone define analytic confidence in estimative intelligence. Confidence intervals belong to the frequency theory, a probability theory of little value for evaluating forecasts in estimative intelligence. Measuring under- or overconfidence in estimative intelligence is most useful if one can calibrate predictions with outcomes. A correct prediction, on the other hand, can often imply that a policymaker has not acted based on the estimate since countering a threat or exploiting an opportunity often will make the prediction void.

Research has nonetheless found that for a majority of people, the same issues that result in overconfidence in knowledge judgments are in effect when it comes to subjective

---

[100] Office of the Director of National Intelligence, “Intelligence Community Directive 203: Analytic Standards,” 2015, 3.
[101] Koriat, Lichtenstein, and Fischhoff, “Reasons for Confidence.”; Griffin and Tversky, “The Weighing of Evidence and the Determinants of Confidence”; Kuhn and Sniezek, “Confidence and Uncertainty in Judgmental Forecasting: Differential Effects of Scenario Presentation”; Sieck and Yates, “Exposition Effects on Decision Making: Choice and Confidence in Choice”; Peterson and Pitz, “Confidence, Uncertainty, and the Use of Information”; Holland, Middleton, and Uys, “Professional Confidence: A Concept Analysis.”
[102] Weiss, “Communicating Uncertainty in Intelligence and Other Professions.”
[103] Field, *Discovering Statistics Using IBM SPSS Statistics*, 54–60.

confidence in forecasts.[104] Peterson and Pitz argue that confidence is a belief of the correctness of a given prediction.[105] According to Teigen, confidence estimation is an internal, second order judgement, made after one has made up one's mind about what one thinks is the most likely outcome of an external event.[106] Consequently, when attempting to specify what factors impact on an individual's confidence in a forecast one can lean on research on over- and under-confidence. Psychological research has found that assessment of confidence relies heavily on how evidence relevant to the different alternatives is reviewed.[107] The strength or the extremeness of the evidence is often given more value than its weight or credence.[108] High strength and low weight results in overconfidence and vice versa results in underconfidence. Hence, how we treat the available evidence is obviously an important factor when ascribing a level of confidence.

Several other factors have been promoted when attempting to define analytic confidence levels in estimative intelligence. In the Intergovernmental Panel on Climate Change (IPCC) guidance notes, confidence is a measurement of evidence evaluation based upon type, amount, quality, and consistency combined with a measurement of agreement between scientists, resulting in a 3x3 confidence matrix.[109] In his Master thesis, Peterson addresses seven factors, the use of structured methods, source reliability, source agreement, level of subject matter expertise, analyst collaboration, task complexity and time pressure.[110] Former UK Defence Intelligence analyst Nick Hare makes the argument that confidence is either a measurement of the quality and quantity of the available information or a measurement of the expected cost of acquiring new information.[111] Friedman has identified three main factors; the reliability of the available evidence, range of reasonable conviction, and the judgment's responsiveness to new information.[112] All these factors were considered when the Norwegian intelligence community worked out an analytic confidence scale to be used

---

104 Fischhoff and MacGregor, "Subjective Confidence in Forecasts."
105 Peterson and Pitz, "Confidence, Uncertainty, and the Use of Information."
106 Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases."
107 Koriat, Lichtenstein, and Fischhoff, "Reasons for Confidence."
108 Griffin and Tversky, "The Weighing of Evidence and the Determinants of Confidence."
109 Mastrandrea et al., "Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties."
110 Peterson, "Appropriate Factors to Consider When Assessing Analytic Confidence in Intelligence Analysis."
111 Nick Hare, " Confidence and Probability: Summary", http://blog.alephinsights.com/search/label/confidence
112 Friedman and Zeckhauser, "Analytic Confidence and Political Decision-Making: Theoretical Principles and Experimental Evidence From National Security Professionals."

as a preliminary guideline for estimative intelligence products, later published in Norwegian intelligence doctrines.[113] The guideline ended up with these factors: information reliability, the number of independent sources, the number of possible alternative interpretations or hypotheses, the number of key assumptions and the size of the information gap. Hence, the Norwegian interpretation of analytic confidence factors is close to Friedman's.

A version of the Norwegian interpretation, with minor adaptions, was in 2020 incorporated into the NATO standard for intelligence processing, AIntP-18. Here the factors included the quality and quantity of the available information, the diversity of collection capabilities, coherence with other intelligence products of high confidence, the number of assumptions, and time and resources available for producing the intelligence product.[114] The AIntP-18 furthermore points towards Structured Analytic Techniques (SATs) as the tools to enable analysts to better assess levels of analytic confidence. The different interpretations have some common factors, especially the quality and quantity of the available information, the number of key assumptions, the judgment's responsiveness to change considering access to new information, and how difficult or costly it would be to get hold of more information. Although accuracy in intelligence estimates may be unattainable, the importance of these factors as well as SATs effect on assessing uncertainty and thereby alleviating discourse failure is unchartered territory in need of research.

Summary

The intelligence community is not alone in having difficulties in dealing with the concepts of uncertainty, probability, and confidence. Probability theory is encumbered with a central ambiguity: is probability a mathematical measurement of random phenomena or a measurement of subjective degree of belief? Psychology offers a somewhat different categorization, distinguishing between external and internal uncertainties where judgements can contain both an external probability and an internal confidence. We should not fall for the temptation to become dogmatic when dealing with the concept of uncertainty. The different theories and variants of uncertainty do not necessary compete as

[113] Politiets sikkerhetstjeneste (PST), "Etterretningsdoktrine for PST"; Forsvaret, "Etterretningsdoktrinen," 2021.
[114] NATO, "NATO Standard AIntP-18(A) Intelligence Processing," 2–8.

they sometimes can complement each other. The division therefore functions as framework to understand the duality of probability and that a probability can radically change given new information. Consequently, assessing uncertainty by communicating key judgments that convey both the relative probability for an event together with the analyst's confidence in the stated probability can help mitigate discourse failure.

The intelligence community, however, has traditionally treated uncertainty in a unilateral manner, conflating probability and confidence, while in natural language people usually understand them as different concepts. Subjective probabilistic expressions are furthermore often vague and imprecise, with the undesirable result that different people often have vastly different understanding of the same terms. First, context, directionality, and framing of expressions influence the understanding. Second, there is the question of using words or numbers, or a combination. Third, there have been, and still are, vastly different practises on the usage of these expressions, not least within the intelligence community. Lastly, even though analytic confidence has been introduced as a new term to be used as a separate scale, it has yet to be used in published intelligence products. Since assessing and communicating uncertainty is central to decision theory, the concept's ambiguity has implications for estimative intelligence products to policymakers where the conflation of probability and confidence results in assessments with equivocation, a key reason for discourse failure.

This ambiguity has affected how the difference between probability and confidence is described in the new analytic standards and how it is understood by intelligence practitioners. None of the current analytic standards or intelligence doctrines provide analysts with any methodology of assessing uncertainty, leaving probability and confidence to the analysts' intuitive approaches, which often lead to biased judgements. While the current accuracy debate has distorted research of one of the main objectives of the intelligence reform, some important issues has been raised. While a first step has been to find a better way to express probabilistic judgements, the continued conflation between probability and confidence stands as proof of why it is necessary for the intelligence community to find suitable methods for assessing uncertainty in estimative intelligence. Not only is it necessary to improve probability calculations, but the intelligence community must also define what it means with analytic confidence and finding suitable factors that contribute to the judgement of the different levels of analytic confidence in a probabilistic judgement. But research looking into techniques and methods for assessing uncertainty in estimative intelligence analysis, including both probability and confidence, is still in its infancy. Structured Analytic Techniques (SATs) are being pointed at as the solution for increased analytic rigour and a key question this thus if SATs also improves analysts' ability to assess uncertainty, and consequently what role SATs have in alleviating or mitigating potential discourse failure.

Chapter 6: Teaching SATs: different nations, different approaches

The previous chapters have established that post-9/11, the use of Structured Analytic Techniques (SATs) has been pointed at as an analytic standard to abide by to increase analytic rigour and improve uncertainty handling, thereby preventing intelligence failures. The desire for increasing analytic rigour through objectivity, neutrality, and scientific inputs to decision-making is nevertheless older, dating back to the Inquiry after the First World War. The effort has since had an intermittent focus and despite showing promise, the efforts have produced no lasting or unifying effect on analytic tradecraft. Even with today's renewed focus on developing a rigorous intelligence analysis tradecraft, a common and shared approach to the implementation of a structured approach to intelligence analysis has yet to materialize. There have been, and still are, substantial differences in how different nations and intelligence agencies have approached the teaching and implementation of SATs.

This chapter will start by presenting a brief comparison of the American and British strategic defence-level approaches to teaching SATs. While the American approach is focusing on single techniques in the analytical toolbox without procedural understanding, the British approach is in theory more methodological, loosely based on a Hegelian dialectic of thesis, antithesis, and synthesis. Diversities in teaching between civilian and military agencies has nevertheless led to different practises. Thereafter, the model taught at the Norwegian Defence Intelligence School (NORDIS) is presented, a structured methodology based upon a comprehensive and iterative framework linked to the overall intelligence process.

The current SATs-research, as described in part 1, has not taken the distinction of different approaches into consideration, producing a skewed picture of the reality. The key argument of this chapter is that contrary to a piecemeal approach of employing SATs, a more comprehensive, layered, and iterative SATs-methodology will improve the analytic process for estimative intelligence analysis. By combining a collaborative use of Structured Analytic Techniques, creativity, critical thinking, and sensemaking, the process can lay a foundation for analytic rigour through increased analytic objectivity and integrity and thereby make it easier for analysts to assess uncertainty. Rather than chasing the chimera of eliminating cognitive bias or improving analytical accuracy, this approach has a greater potential of preventing discourse failure, and thus arguably have a greater role in mitigating potential failures in the intelligence-decision-making complex.

SATs in the United States: single techniques in the analytical toolbox

An early pursuit towards a more structured and "scientific" approach to intelligence analysis is often attributed to Sherman Kent, whose efforts to improve intelligence theory and tradecraft earned him the unofficial title as 'the father of intelligence'.[1] Kent's work was nonetheless building on earlier efforts. President Wilson aimed for a 'scientific peace' through The Inquiry after the First World War, staffing it with a significant pool of social scientists drawn from US Academia and business.[2] A considerable number of scientists took part in operational research in the Second World War, contributing to major methodological innovations in technical, economic, and operational analysis.[3] The American approach to teaching and using SATs nevertheless stems from the historical work of CIA's Analytical Methodology Research Division in the 1970s, experimenting with how methods from the behavioural revolution of social and political sciences could be adapted and applied to political intelligence analysis.[4] The aim was to construct generalizations that could explain and predict political behaviour from a neutral, objective and unbiased point of view.[5] The main finding, however, was that there was a need to adapt these methods to make them useful for CIA analysts since few intelligence problems could be quantified in the same manner as in academia.[6]

---

[1] https://www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/sherman-kent-the-father-of-intelligence.html, accessed 3 June 2020
[2] Gelfand, *The Inquiry: American Preparations for Peace, 1917-1919*, 16.
[3] Marquardt-Bigman, "Behemoth Revisited: The Research and Analysis Branch of the Office of Strategic Services in the Debate of Us Policies towards Germany, 1943–46"; McCloskey, "British Operational Research in World War II."
[4] Central Intelligence Agency Office of Research and Development, "General Notice No. 91: Establishment of Analytical Methodology Research Division (AMR/ORD)"; Heuer, "Adapting Academic Methods and Models to Governmental Needs: The CIA Experience."
[5] Guy, *People, Politics and Government: A Canadian Perspective*, 57–58.
[6] Heuer, "Adapting Academic Methods and Models to Governmental Needs," 8.

A persistent political pressure for objectivity and neutrality to avoid politicisation continued to impact the CIA throughout the Cold War, resulting in DCI Gates stating in in a message to CIA analysts 1992 that 'We must recommit ourselves to the good old-fashioned scientific method — the testing of alternative hypotheses against the evidence.'[7] Jack Davis' Alternative Analysis and Douglas MacEachin's Linchpin Analysis then became the backbone of changes made to analytic training in the CIA.[8] The fundamental idea was to move from fortune-telling to a focus on facts and findings where forecasts were to be based upon a sound evidentiary base and transparent analytic logic.[9] The new analytic focus together with a need to rebuild CIA's analytical capability led to the start-up of CIA's Career Analyst Program in May 2000, which a few years later had turned into a four-month training program with 'special emphasis on creating a foundation of analytic tradecraft skills for new analysts.'[10] Outside the CIA, the issuing of new intelligence community directives by ODNI, especially "ICD 203: Analytic Standards", can be seen as an attempt to mimic CIA's effort to professionalize intelligence analysis. SATs became a quick and easy way to explain and use social science techniques, also called 'social science methodologies for dummies.'[11]

The basis for the American approach to teaching and using SATs is CIA's "Tradecraft Primer", made public in March 2009. Different SATs are divided into diagnostic techniques (Key Assumptions Check, Quality of Information Check, Indicators or Signposts of Change, Analysis of Competing Hypotheses (ACH)), contrarian techniques (Devil's Advocacy, Team A/Team B, High Impact/Low Probability Analysis, "What If?" Analysis), and imaginative techniques (Brainstorming, Outside-In Thinking, Red Team Analysis, Alternative Futures

---

[7] Gates, "A Message to Analysts: Guarding Against Politicization."
[8] Davis, "Improving CIA Analytic Performance : Strategic Warning." See also "Recommendations of the Jeremiah Report", MORI DocID 406297, https://nsarchive2.gwu.edu/NSAEBB/NSAEBB187/IN38.pdf, accessed 25 August 2020
[9] MacEachin, "The Tradecraft of Analysis: Challenge and Change in the CIA's Directorate of Intelligence"; Davis, "Defining the Analytic Mission: Facts, Findings, Forecasts, and Fortune-Telling."
[10] CIA, "A Major Agency Success Celebrated 100 Times Over: CIA's Career Analyst Program", https://www.cia.gov/news-information/featured-story-archive/2008-featured-story-archive/a-major-agency-success.html, accessed 1 July 2020.
[11] Marrin, *Improving Intelligence Analysis: Bridging the Gap Between Scholarship and Practice*, 31.

Analysis).[12] Thereafter a rough timeline for when to use which techniques throughout an analytic project is presented.

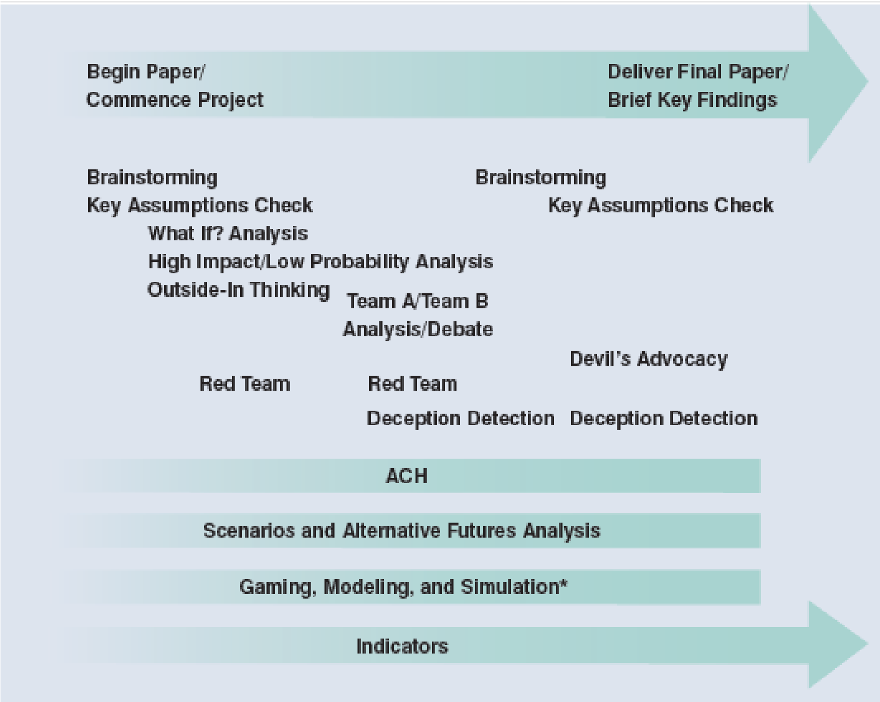


Figure 6: CIA Tradecraft Primer Timeline for SATs throughout an analytic project.[13]

Prior to the Tradecraft Primer, an early version of National Security Agency's Critical Thinking and Structured Analysis Class Syllabus was designed almost as a blueprint of Morgan D. Jones' *The Thinker's Toolkit*.[14] Later, Richards Heuer and Randolph Pherson put together a collection of more than sixty different techniques aimed at increasing analytic rigour and avoiding intuitive pitfalls in their book *Structured Analytic Techniques for Intelligence Analysis*, which was published in its third edition in 2020. In the latest version, the authors have decreased the number of categories of techniques from eight to six: getting organized, exploratory, diagnostic, reframing, foresight, and decision support techniques. Furthermore, eight core techniques are pointed out for every analyst to know how to use: Cluster Brainstorming, Key Assumptions Check, ACH, Premortem Analysis and Structured Self-

[12] Sherman Kent School Kent Center for Analytic Tradecraft, "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis"; Central Intelligence Agency, "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis."
[13] Central Intelligence Agency, "A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis," 38.
[14] Moore, "Critical Thinking and Intelligence Analysis"; Jones, *The Thinker's Toolkit: Fourteen Powerful Techniques for Problem Solving*.

Critique, What If? Analysis, Multiple Scenarios Generation, and Indicators Generation, Validation, and Evaluation.[15]

The American approach to SATs is also the one adopted by NATO. The NATO Alternative Analysis Handbook from 2012 leans heavily on the taxonomy in CIA's "Tradecraft Primer" and the SATs-course at the NATO School in Oberammergau is taught by the United States Defense Intelligence Agency (DIA), using a DIA taxonomy and syllabus.[16] This syllabus starts with addressing critical thinking and cognitive biases, and thereafter presents different SATs in the sequence of diagnostic techniques, structuring information, hypothesis generation, imaginative thinking, and contrarian analysis.[17]

The overall American approach was originally aimed towards learning how different structured techniques can aid a critical reasoning process to mitigate bias and mindsets. Today's reality, however, is that the different techniques are presented and taught as single tools in a large toolbox despite being divided into categories. The consequence is a practise that does not place the techniques in either the overall intelligence process or an adapted social science methodology. The DIA Critical Thinking and Structured Analysis course briefing states for instance that to 'impose a rigid methodology [is] NOT the objective.'[18] Similarly, a course participant on a Canadian SATs-course stated: 'The focus on the course is to perform the individual SAT in a correct manner. There is no such thing as teaching the analysis process as a chain of interconnected SATs where the output of one SAT leads to an input for the next.'[19] There is furthermore no evidence of dividing SATs into their suitability to answering different types of intelligence problems. Consequently, in America there seems to be a piecemeal approach to the application of SATs that focuses on particulars instead of

[15] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020, secs. 3.1-3.3.
[16] NATO, "Alternative Analysis Handbook."; Unclassified course material from the NATO School Oberammergau, N2-03-B-19 Intelligence Analyst Course – Critical thinking & Structured Analysis.
[17] Defense Intelligence Agency, "Critical Thinking and Structured Analysis (CTSA) Syllabus". See also NATO School N2-03-B-19 Intelligence Analyst Course – Critical thinking & Structured Analysis, "20191028_0930_U_CTSA Full Brief with Instructor Guidance – Complete."
[18] NATO School N2-03-B-19 Intelligence Analyst Course – Critical thinking & Structured Analysis, "20191028_0930_U_CTSA Full Brief with Instructor Guidance – Complete."
[19] Personal communication with an Intelligence Officer who attended a Canadian SAT-course in 2015, 20 November 2020.

comprehensiveness, hoping that the analysts already understand social science methodology and how to conduct intelligence analysis.

SATs in the United Kingdom: aiming at a "Hegelian Triad"

The British approach to intelligence assessments has historically been one with a passion for gifted amateurism, based on an idea that good analysts could be recruited among clever people from good universities, people who would be good at seeking information and writing reports.[20] For British civil servants working as analysts in the Assessment Branch in support of the Joint Intelligence Committee, the focus was on apprenticeship and where the use of a specific intelligence methodology was at best optional.[21] The Butler review addressed the British government's overconfidence in being able to recruit adept persons to serve as analysts by calling for a professionalization of the intelligence community with room for a career within the Assessment Branch.[22] The function of Professional Head of Intelligence Analysis (PHIA) was subsequently established as part of the Cabinet office in 2005. While PHIA had little impact in the beginning, its community-wide role today is to advise on analytical training, recruitment, and career structures as well as on development of analytic methodology and training, based on skills levels described in the Professional Development Framework for All-source Intelligence Assessments.[23]

Prior to the conclusions of the Butler review, however, Defence Intelligence (DI) had already a few years earlier taken the lead to improve the British approach to intelligence analysis and assessments. In the wake of 9/11, key persons working with teaching and developing intelligence analysis methodology had started a push to mirror the American approach to achieve greater academic rigour in DI assessments. They received support from both the CIA Sherman Kent School and the United States Defense Intelligence Agency, making it possible

---

[20] Davies, *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 2: Evolution of the UK Intelligence Community*, 219; Strong, *Intelligence at the Top: The Recollections of an Intellgience Officer*, 334–36.
[21] Davies, "Jointery versus Tradecraft: The Brunel Analytical Simulation and Alternative Approaches to Intelligence Analysis and Analytical Professionalization in Postgraduate Academic Teaching," 202.
[22] Lord Butler, "Review of Intelligence on Weapons of Mass Destruction."
[23] Professional Head of Intelligence Assessment, "Professional Development Framework for All-Source Intellgience Assessment."

to run the first courses on Structured Analytic Techniques in the United Kingdom in 2004-05.[24] In 2006, the Futures and Analytical Methodology unit at DI issued "Quick Wins for Busy Analysts", a pamphlet describing how to use different Structured Analytic Techniques to answer different types of intelligence problem types. The core idea was to address the problem of analysts' sole focus on current intelligence by creating a coherent, consistent, and scientific framework to make analysts to also focus on the longer term.[25] Today, DI is still the leading agency in the United Kingdom in developing and issuing all source analytic standards, analysis handbooks, and tradecraft guiding notes as well as the training of analytical methods. PHIA, since the publishing of the Chilcot Inquiry Report, has sought to take a more central role in spreading analytical methods, with the foundation of the UK Intelligence Assessment Academy in August 2022 demonstrating their new central role.[26]

The British approach towards SATs, as described in "Quick Wins for Busy Analysts", seemingly aims at pushing analysts towards a Hegelian Triad of thesis, antithesis and synthesis, where the hypotheses and future scenarios are being tested and challenged through diagnostic techniques aimed at improving communication of uncertainty.[27] In practise, the British approach is focused on selecting the best analytical approach and techniques based upon how the intelligence problem is classified.[28] The problems are either closed or open, and either present or future. The different techniques are then divided into data organisation (Mind Maps, Chronologies, Timelines, Matrices, SWOT), hypothesis generation and testing (Environmental Scanning, Structured Brainstorming, and ACH), and scenario generation and evaluation (Cone of Plausibility, Backcasting, and Key Assumptions Check). While data organization techniques are vital for all intelligence problems, questions about the present are subject to hypothesis generation and testing techniques, and questions about the future are subject to scenario generation and evaluation techniques.

[24] Gustafson, K., 'History, Evolution, Future of SATs', lecture at NORDIS 2016.
[25] Personal communication with Nick Hare, former DI senior analysts and co-author of "Quick Wins for Busy Analysts".
[26] "Joint Intelligence Organisation", The Government of the United Kingdom, 2 April 2022, https://www.gov.uk/government/groups/joint-intelligence-organisation#:~:text=The%20UK's%20Intelligence%20Assessment%20Academy,Assessment%20as%20an%20academic%20discipline.
[27] Gustafson, K., "History, Evolution, Future of SATs", lecture at NORDIS 2016. See also Defence Intelligence, "Quick Wins for Busy Analysts."
[28] Professional Head of Defence Intelligence Analysis, "Quick Wins for Busy Analysts," 2–4. Also interview K

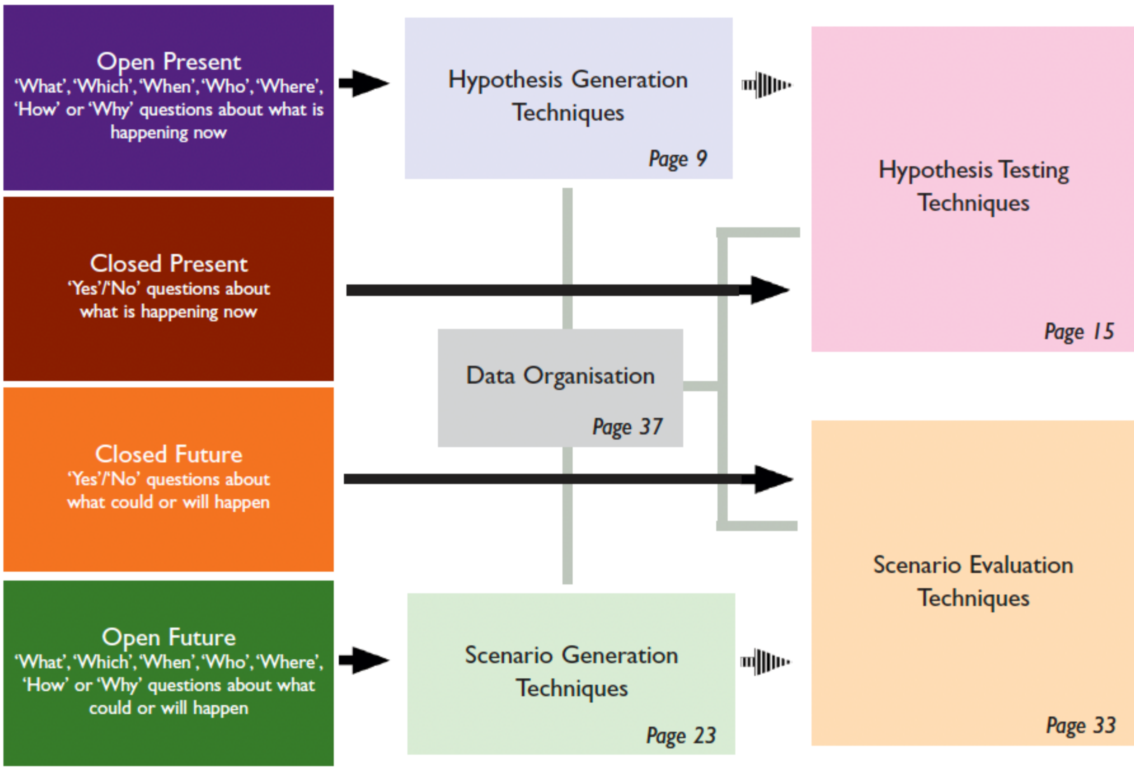


Figure 7: UK DI SATs approach[29]

Unlike the American approach, the British approach focuses on environmental scanning using the STEMPLES mnemonics at the start of the project, and indicator identification and Key Assumptions check in the generation and testing phases.[30] Moreover, even though it is not evident from the published material, the 6-day long Defence Intelligence Analysis Module is based upon having teams of analysts approaching intelligence problems in a layered system of structured techniques where the output of one is used as input in the next.[31] Hence, the use of SATs is not only based upon social science methodology but also placed within the overall context of the assessment process. PHIA's equivalent, the 5-day Government Intelligence Analysis Techniques course, leans on the other hand more towards the American approach of teaching single techniques for independent problems, and by that adapting a piecemeal approach without procedural understanding.[32]

[29] Defence Intelligence, "Quick Wins for Busy Analysts," 3.
[30] Defence Intelligence, 3, 6.
[31] Interviews J & K. in 2022 DIAM was replaced with the Intelligence Analyst Foundation Course providing some moderate improvements.
[32] Course material from PHIA; personal communication with an analyst participating on a 2015 PHIA course.

SATs in the Anglosphere - impact on previous research

Both the American and British approaches to SATs can broadly be seen as a project pathway, from using divergent creative thinking aimed at hypothesis generation to using convergent critical thinking aimed at hypothesis testing. However, even though "Quick Wins" and *Structured Analytic Techniques* have become the ultimate go-to sources for teaching and using SATs, both the categorization and the suggested taxonomy of which techniques to use when and why, can result in a tendency for analysts to just pick a technique to prove for their supervisors that SATs has been applied to a specific type of question.[33] This problem is enhanced by a piecemeal approach to SATs as single tools in a toolbox instead of a structured methodology based upon principles from social sciences.[34] Even though the approach of the British Defence Intelligence is more comprehensive and layered, the taxonomy in "Quick Wins" can limit analysts to believe they have to choose between two distinct problem sets that have no connection. ACH is, for instance, not connected to evaluate and monitor scenarios, but rather limited to hypotheses of the present situation. The main pitfall of both the American and British approaches is, however, that the techniques can be perceived as limited to a single analytic project that is finished as soon as the key assessments have been presented to the policymakers. Accordingly, neither approach sets the use of Structured Analytic Techniques into the overall iterative intelligence process where there is also a need for continuous feedback and dialogue with the other sub-processes, especially direction and collection. Consequently, the result is that the use of Structured Analytic Techniques can by some be perceived as a mechanical and counterproductive checklist procedure.[35] When the use of tools is not situated within a structure, they are less likely to help the analyst derive a comprehensive understanding of the problem, thereby also less conducive to alternative explanations and clear expressions of uncertainty in assessments. This lack of procedural understanding has, as described in part 1,

[33] Coulthart, "Improving the Analysis of Foreign Affairs: Evaluating Structured Analytic Techniques," 175.
[34] Moore, "Critical Thinking and Intelligence Analysis."; Unclassified course material from the NATO School Oberammergau, N2-03-B-19 Intelligence Analyst Course – Critical thinking & Structured Analysis; Course material from PHIA.
[35] Coulthart, "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency," 942. See also Defence Intelligence Analytic Standards Checklist.

impacted most of the research on SATs, using correct methods but for the wrong questions and settings.

The structured approach of a small NATO nation

Norway has chosen a different approach; one where different SATs have been put into a more comprehensive system that allows for better interaction with the overall iterative intelligence process (cycle). Although the usefulness of the intelligence cycle has been widely debated, using it as a backdrop can be beneficial since it is regularly referred to in most other intelligence courses. The problems of the intelligence cycle are nevertheless important to keep in mind and one must especially give recognition to the fact that intelligence analysis is a socio-cognitive process where, in Johnston's words, 'a collection of methods is used to reduce a complex issue into a set of simpler issues within a secret domain'.[36] He furthermore argues that:

> A carefully prepared taxonomy can become a structure for heightening awareness of analytic biases, sorting available data, identifying information gaps, and stimulating new approaches to the understanding of unfolding events, ultimately increasing the sophistication of analytic judgments.[37]

The Norwegian perspective of teaching and using SATs builds upon several approaches, from within the wider intelligence community as well as ideas from different academic fields. Early SATs-courses were adapted from an early Dutch approach, building on Morgan D. Jones' *The Thinker's Toolkit: 14 Powerful Techniques for Problem Solving,* together with classes in network analysis and the Royal Dutch Shell/Global Business Network scenario analysis.[38] The Norwegian military commitments in Afghanistan resulted in a large increase in the number of analysts needed, making it necessary to develop a more comprehensive training programme for the analysts at the IntBn. The result was a 6-month programme involving everything from information management and data structuring, via classes in both conventional and COIN tactics, combined with how to utilise SATs and critical thinking in different types of military operations. Apart from the course material inherited from the

[36] Johnston, "Developing a Taxonomy of Intelligence Analysis Variables," 65.
[37] Johnston, 61.
[38] Course material from the Netherlands Defence Institute for Security and Intelligence course at Setermoen, Norway, 18 August 2007

Netherlands Defence Institute for Security and Intelligence, the curriculum was based upon both NATO and American doctrines, like US Army *FM 3-24 Counterinsurgency* and USMC *MCWP 3-12 MAGTF Intelligence Production and Analysis*, together with Richard Heuer's *Psychology of Intelligence Analysis*, and specific COIN literature.

SATs training outside the IntBn started when the Norwegian Defence Intelligence School (NORDIS) in 2010 introduced a two-week SATs-course open to all in the Norwegian intelligence community. In this course, the initial ideas from the analyst education in the IntBn was synchronized with new ideas from other authors, like Robert M. Clark, Nate Silver, Richards Heuer, and Randolph Pherson. The NORDIS SATs-course became mandatory for attendance in subsequent advanced intelligence courses as well as part of the curriculum for the NORDIS' bachelor's programme in intelligence and language studies from 2013.[39] Using SATs therefore became a natural part of the more advanced intelligence courses, such as intelligence support to joint operations, where the use of SATs was incorporated in all phases, from operational planning to current operations and indication and warning. Furthermore, in 2016, the Norwegian Security Police (PST) developed their own intelligence analysis education, based upon recruiting key personnel from the Intelligence Battalion. The main course constituted three 3-day long modules where two of the modules focused on SATs like Key Assumptions Check, Devil's Advocacy, Scenario Generation, and ACH, as a foundation for their approach to improve both collection and threat assessments.[40] By 2017, Structured Analytic Techniques had become the principal advocated methodological approach for multiple source analysts in the Norwegian intelligence community.

The NORDIS SATs-methodology is a framework that combines Robert M. Clark's overall iterative prediction methodology as described in *Intelligence Analysis: A Target-Centric Approach* with a taxonomy of relevant Structured Analytic Techniques against a backdrop of creative and critical thinking and sensemaking.[41] The purpose is to enhance analytic reasoning through a shared and collaborative methodological construct that increases analysts' awareness of the scope of possibilities, leaves a clear audit trail, and where the

[39] Norwegian Defence Intelligence School, "Study Programme Bachelor of Arts in Intelligence Studies."
[40] *Politiets Sikkerhetstjeneste (PST), 'Kurs i etterretningsanalyse'*, 11 January – 4 March 2016.
[41] Borg, "Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology."

shared subjective experience and logical reasoning can improve judgements and products through continuous review and testing.

The NORDIS SATs-methodology is based upon a few, key steps. Although the steps of the process are presented in a flowchart, the output will be best when conducted in an iterative manner to better accommodate for new information. The process is initiated by defining the intelligence problem followed by estimating a past and present state model of the issue based upon existing patterns, key assumptions, and an analysis of driving forces. The subsequent step is to model a comprehensive set of alternative future state models of the problem based upon changes in the driving forces. The last step is to test and monitor the alternative future models, and also nominate indicators for collection. New information will bring new knowledge and understanding, creating a need to update the current and future models, making the process iterative. The aim is to not only create knowledge of the current situations but also to understand how the situation can develop given certain circumstances and changes in the key driving forces.[42] The methodology thereby combines horizon scanning with indication and warning. Although the NORDIS model suggests the use of a limited set of techniques in the different stages, the use of other, alternative techniques is encouraged if these are better suited for the problem at hand. Hence, the methodology functions as a framework that can be adjusted to the needs for the different agencies, analytical teams, or individual analyst.

[42] Borg, 11.

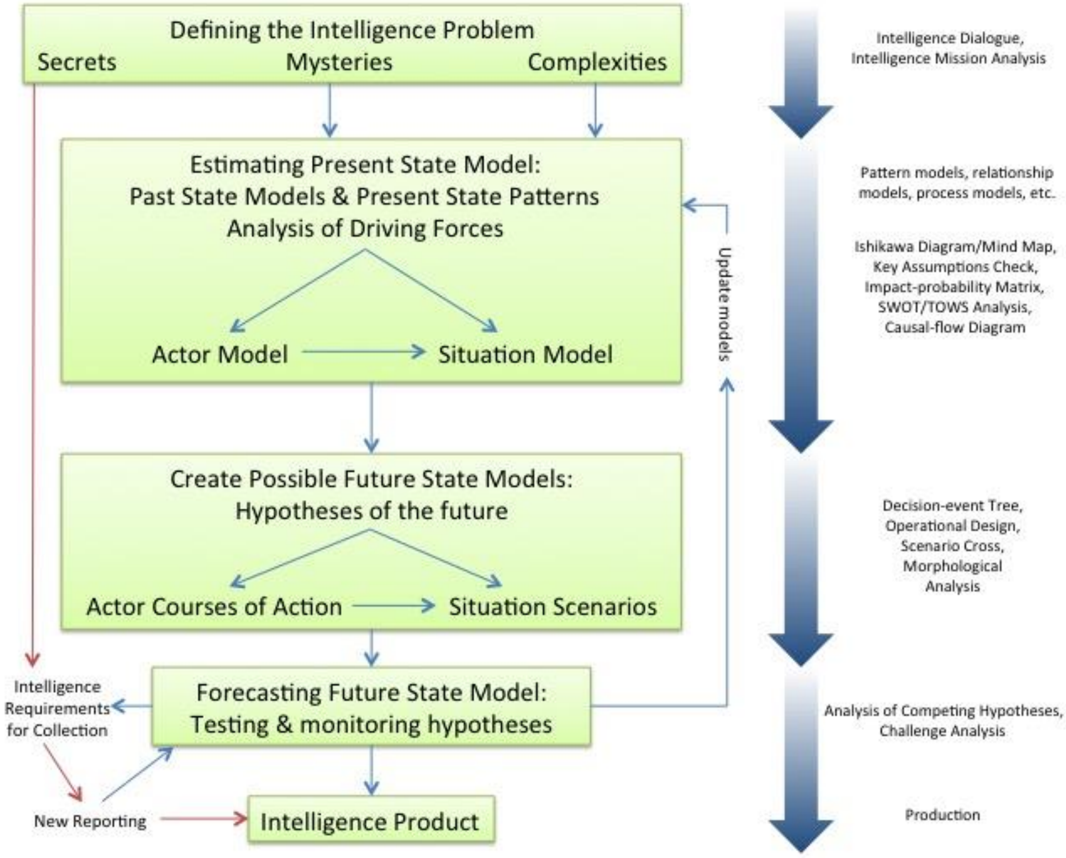


Figure 8: The NORDIS SATs-methodology[43]

A good process starts with well-defined questions or problem sets, in the NORDIS model initiated by an intelligence mission analysis based on a dialogue with the policymakers. The purpose is to get an improved comprehension of what is needed with regards to type of product, purpose, timeframe, and not the least, what they really want elucidated.[44] How the intelligence problem is categorized in the intelligence mission analysis will largely direct the subsequent analytical approach. Unlike the British open/closed, present/future taxonomy, the NORDIS model is using the secrets, mysteries, and complexities delineation as described in chapter 3. Unlike other SATs approaches, the NORDIS SATs-methodology does not have a separate focus on secrets, but rather emphasizes how to answer mysteries and complexities. These are intelligence problems where the analysts become more the centre of gravity due to the increased importance of assumptions, assessments, and use of different techniques

[43] Borg, 11.

[44] Clark, *Intelligence Analysis: A Target-Centric Approach*; Vandepeer, *Applied Thinking for Intelligence Analysis: A Guide for Practitioners*.

and methods in the process.[45] In the NORDIS model, a mystery is typically an actor's alternative courses of actions while complexities deal with long-term scenarios based upon the development of key driving forces, including how competing actors will attempt to reach their conflicting desired end states.[46] Complexities will therefore inevitably contain one or more mysteries and both problem sets depend upon uncovering secrets. Breaking down and restating the intelligence problem consequently becomes a vital starting point to focus the subsequent methodological approach.

With the intelligence problem defined, the next step is to construct a past and a present state model of the problem, often called basic intelligence. These models are central, functioning as a current body of knowledge from which future models can be created and toward which new information can be judged. Like the British SATs-model, an essential element of this step is data structuring and visualization, often through sub-models like pattern analysis, link charts, human terrain mapping, process models, matrices, and other relationship models. Unlike the British approach, however, pattern and trend models are only treated as the tip of the iceberg, a necessary starting point to understand the driving forces at play. These underlying factors have acted on the issue in a way that created the present situation and are usually also factors that influence the development into the future.[47] Driving force analysis can be dated back to Sun Tzu's five factors that govern the art of war and is today widely used within several other fields than the military, often through mnemonics such as DIME, PMESII, STEMPLES, or PESTLE.[48] Structured brainstorming techniques in combination with visualization techniques like Ishikawa diagrams and mind maps help identify and connect the driving forces and the sub-factors.[49] The driving force analysis provides inputs for other techniques where strength-weakness-opportunities-threat (SWOT) analysis is used for actor evaluation while creating causal-flow diagrams are more

[45] Krizan, "Intelligence Essentials for Everyone."
[46] Fishbein and Treverton, "Making Sense of Transnational Threaths."
[47] Clark, *Intelligence Analysis: A Target-Centric Approach*, 216–24.
[48] DIME (diplomacy, information, military, economic) is based upon Grand Strategy, see Liddell Hart, *Strategy*, 322; Defence Intelligence, "Quick Wins for Busy Analysts." uses STEMPLES (social, technological, environmental, military, political, legal, economic, security), United States Joint Forces doctrines use PMESII (political, military, economic, social, information, infrastructure), for more acronyms see https://en.wikipedia.org/wiki/PEST_analysis.
[49] Borg, "Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology," 13.

suitable for modelling regional situations where many factors and actors are at play.[50] The objective of this stage is to first use a divergent mode of thinking to identify all the relevant factors to create a better overview and thereafter to employ a convergent mode of thinking to focus the analysis on the key factors and issues.

The third step of the NORDIS SATs-methodology is to use the results from the present state analysis for the subsequent modelling of alternative future state models. Identifying the most important, or key, driving forces is imperative in this step. If the driving forces do not change, the current pattern and trends will continue into an extrapolated future model, what is called a base-line scenario when using the Cone of Plausibility technique.[51] However, complex situations are unstable and can rapidly change, invalidating current pattern and trends, situational logic, historical analogies, or applied theory. Intelligence analysts must therefore identify a wide range of possible future situations based upon incorporation of both changing and possible new driving forces.[52] Whereas stable high impact factors will be part of any future situation, the high impact-high uncertainty factors are key to model alternative futures. In the NORDIS SATs-course weighted ranking and cross-impact matrices are techniques used to rank the driving forces to find the driving forces that are assessed to have both the highest impact but also have the largest potential for change.[53]

With a comprehensive analysis of the driving forces as a basis one can start creating future models, where the choice of techniques is usually dependent upon type of intelligence problem. TOWS-analysis combined with a decision-event tree and operational designs are useful techniques to an actor's alternative courses of action. Where the original SWOT-analysis typically focus on maximising opportunities and minimising weaknesses in the present environment, the TOWS-analysis produce an interaction analysis where an actor's internal strengths and weaknesses are seen in conjuncture with the opportunities and threats in the present as well as a possible future operational environment. Hence, the

[50] Borg, 13–14.
[51] Clark, *Intelligence Analysis: A Target-Centric Approach*, 295–98; Professional Head of Defence Intelligence Analysis, "Quick Wins for Busy Analysts," 27.
[52] Clark, *Intelligence Analysis: A Target-Centric Approach*, 221–22.
[53] Borg, "Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology," 13.

TOWS-analysis aims at identifying alternative actions and strategies an actor can adopt in the short, middle, and long term to reach the desired end state.[54] By visualizing the different strategies and actions in operational designs, analysts can model an actor's alternative courses of action based upon a combination of the different strategies, tactics, and actions from the TOWS-analysis, conditions needed to be achieved towards fulfilment of the desired end state.[55] The combination of a TOWS-analysis and a COA-generation technique will therefore also produce relevant indicators for collection.

When analysing complexities, like how a regional situation like the Middle East will look like in the future, one needs to model a series of alternative scenarios based upon how the most important driving forces may interact at a given time in the future. The NORDIS SATs-course teaches the Royal Dutch Shell/Global Business Network (GBN) scenario planning technique, often called 'the gold standard of corporate scenario generation'.[56] This technique is the same as the Alternative Futures Analysis and the Quadrant Crunching described in the CIA "Tradecraft Primer" and in *Structured Analytic Techniques* respectively. The technique produces four alternative scenarios based upon identifying the two driving forces with both the highest impact and the highest degree of uncertainty regarding development and thereafter construct a 2x2 matrix, or scenario cross, with well-defined extremes for each driving force in the given timeframe of the intelligence problem.[57]

The original GBN scenario planning technique has limitations since it is 'almost impossible to fully characterize the uncertainties of the future with just two dimensions.'[58] Slow-changing or less important factors that have a bearing on the problem are not necessarily explicitly delt with. In the NORDIS model, analysts mitigate this by creating an updated causal-flow diagram within each quadrant to depict more comprehensively how all the different driving forces exert influence on each other, contingent on the more radical changes of the two key

---

[54] Weihrich, "The TOWS Matrix - A Tool for Situational Analysis."
[55] For more on operational design, see NATO Allied Joint Doctrine for Planning of Operations (AJP 5(A)v2 with UK National Elements), May 2019, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/837082/dcdc_doctrine_nato_planning_of_ops_ajp_5.pdf, accessed 8 September 2020
[56] Bishop, Hines, and Collins, "The Current State of Scenario Development: An Overview of Techniques," 5.
[57] For an outline of the steps, see for instance MacKay and McKiernan, *Scenario Thinking: A Historical Evolution of Strategic Foresight*, 60–61.
[58] Bishop, Hines, and Collins, "The Current State of Scenario Development: An Overview of Techniques," 20.

drivers.[59] This mitigation can also be applied to the Cone of Plausibility technique described in "Quick Wins". Alternatively, one can either create several GBN-models by combining various pairs of key drivers or apply Morphological Analysis that allows for a combination of more factors.[60]

While the abovementioned scenario techniques are useful to identify a set of alternative hypotheses to the known unknown, there is still a chance of being surprised when scenarios unfold outside what was assessed to be the limit of possibilities. These so-called black swan events come as a surprise because they generally exist as unknown unknowns.[61] To remedy this situation, other creative hypotheses-generation techniques must be utilised, like what-if analysis, and high impact/low probability scenarios, to exhaust the scope of opportunities. Allowing for some unstructured, creative, and convergent thinking in the hypotheses generation part of the process can ensure that analysts find all relevant possible future courses of action and scenarios, including wild cards.[62]

A major hurdle for intelligence analysts is that they deal with elucidating problems when others want to keep the necessary information secret and hidden.[63] When lacking data and information, analysts must make assumptions to fill in the gaps. A useful definition of assumption for intelligence analysis is 'judgments which is not known to be 100% true but *whose truthfulness is not under current inquiry*.'[64] However, if these assumptions are not questioned or challenged, they tend to be treated as facts, which again will lead to intelligence failures if wrong. The key assumptions, those that can make assessments invalid if wrong, must be checked and validated throughout the estimative process.[65] For that reason, key assumptions check cannot be placed in a specific place in the SATs-model, as is done in "Quick Wins", but must be applied to validate both past, present, and future models.

---

[59] Borg, "Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology," 14.
[60] Schwartz, "Scenarios for the Insurgency in Iraq"; Johansen, "Scenario Modelling with Morphological Analysis."
[61] Taleb, *The Black Swan: The Impact of the Highly Improbable*.
[62] Kant, *The Critique of Pure Reason*, 276–77; Pendleton-Jullian and Brown, *Pragmatic Imagination: Single from Design Unbound*.
[63] Gill and Phythian, *Intelligence in an Insecure World*, 105.
[64] Gustafson, "Assumption... the Necessary Evil."
[65] Clark, *Intelligence Analysis: A Target-Centric Approach*, 57,103; Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020, 20.

Some key assumptions will be judged to be solid, being based upon proven theory or credible reporting. Other key assumptions can be assessed to be key uncertainties and should therefore be prioritised in the intelligence requirement management (IRM) process.

The fourth step of the NORDIS SATs-model is monitoring and testing of the hypotheses. While alternative courses of action and scenarios can widen policymakers' scope of what may happen, what they really want to know is what is most likely to happen now.[66] Being provided with sufficient indication and warning will help policymakers to take action to prevent or lessen the blow. Analysts' predicament is that it is impossible to provide policymakers with a detailed, absolute, and accurate prediction of the future.[67] The best analysts can do is to monitor the situation to track indicators that can point towards which course of action and/or scenario is about to come true and to report their conclusions with a degree of uncertainty. However, if only looking for indicators that will confirm a given scenario, the process will inevitably lead to failure. Applying a back-casting technique to identify significant indicators for each course of action and scenario and then combining this with ACH will improve the monitoring and testing phase and will be of immense value when drafting key judgements. Additionally, by assessing the diagnosticity of future indicators, analysts will stimulate a more focused collection effort, increasing the likelihood of more precise warning ahead of time.

ACH in the NORDIS SATs-model is adapted from the original version to mitigate some of the main pitfalls. First, the alternative future models from step 3, both courses of action and scenarios, replace ACH-step one, identify hypotheses. If using ACH as a tick-box technique, the development of hypotheses can easily be the result of an intuitive system 1 approach instead of a structured system 2 approach, thereby drastically increasing the probability of not having exhausted the field of possible hypotheses. Second, when creating several ACH matrices to cover different problem types, the conclusion from an ACH-matrix for a mystery can be used as evidence in an ACH-matrix dealing with a complexity, thereby limiting the effect of an unbalanced set of evidence due to a difference of resolution. The third adaption

---

[66] Gardiner, "Squaring the Circle: Dealing with Intelligence-Policy Breakdowns"; Dahl, *Intelligence and Surprise Attack*, 21–26.

[67] Ben-Israel, "Philosophy and Methodology of Intelligence: The Logic of Estimate Process."

is done to prevent diminishing returns, a problem that shows up when tracking a situation for a long time where a few new, inconsistent indicators will be unable to change a conclusion despite clearly pointing towards a different hypothesis being more likely. Dividing the ACH matrix into several sub-matrices, one for each reporting period, makes it much easier to spot a changing situation as the new indicators will stand out as signposts. Although seemingly trivial, in the NORDIS SATs-model, the word evidence is exchanged with indicator to better reflect ACH's role in indication and warning, to be in better agreement with NATO AAP-6 Glossary of Terms and Definitions, and lastly due to translation issues where the Norwegian word for evidence mainly belongs to the judicial system.

The final step is a product that answers the intelligence problem, which in estimative intelligence most often is about the future. Describing the present is usually not enough, analysts also need to forecast how the future may look like since the future is plural and not limited to a single outcome.[68] Simultaneously, there is often a need for more tactical warnings to get the attention from policymakers. The combination of several techniques makes this possible, where tactical indicators can be derived from strategic long-term scenarios. At the same time, SATs are not exact science, and the use of these techniques should never be dogmatic. There is always a need to challenge assumptions, judgements, and conclusions, not only at the end of the process but continuously from the mission analysis to the ACH results. Key assumptions check, premortem analysis, structured self-critique, devil’s advocacy, and what-if analysis are all techniques that can help analysts to challenge their thinking.[69] These techniques allow for creative and critical thinking as well as sensemaking to be an integral part of the use of SATs, vital aspects of dealing with intelligence problems that are pestered with uncertainty, ambiguity, and socio-cultural misconceptions. New information will also give new insights. Due to the iterative nature of intelligence analysis, the NORDIS SATs-methodology depict the need to loop back one or several times to update models, key assumptions, and hypotheses. Just as repeated randomized testing is vital in other fields of science, indication and warning intelligence cannot assume that the world of tomorrow will look the same as the one yesterday. Unless

[68] Peter Schwartz, cited in Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020, 252.
[69] Heuer and Pherson, chap. 8.

the models are continuously updated and refined with new information, the models will be outdated and the analysts' conclusions will not reflect the reality, thus severely increasing the risk of intelligence failure.

While the NORDIS SATs-methodology is based upon a more comprehensive and iterative approach than many other approaches, the real advantage comes with how it is taught to be applied. Some intelligence analysis courses, such as the DIA run NATO course and PHIA's SATs-course, relies on teaching the use of single techniques to the individual analysts. This approach assumes analysts intuitively will understand how to apply SATs in the analytical process, something which is often not the case. One impression is that the teaching of intelligence analysis suffers from a significant deficiency: 'The real-world way of thinking about SATs is not how they are thought of or taught by those responsible for either training or education.'[70] Other courses, such as the NORDIS and DI ones, focus on collaborative and layered use of SATs in small teams.[71] Decomposition, externalization, and visualization are key principles of a structured approach to intelligence analysis, principles that are enhanced when the analysis can be reviewed, discussed, and critiqued not only at the end, but throughout the process.[72] In the NORDIS SATs-course, the participants are divided into teams of five to six analysts, mixing generalists and specialists, military and civilians, and with the instructors acting as methodologist. To be enhance the team effort, the participants are first given lectures on how to individually use a structured technique before coming together for a team-based application of the same technique in a pre-scripted simulation scenario. At the end of the course, the teams are provided with new information in two iterations, making it necessary to update the previous work, implementing the iterative nature of the intelligence analysis process.

By combining an iterative analysis model with the use of a variety of Structured Analytic Techniques, the NORDIS SATs-methodology provides an improved taxonomy of which

[70] Stephen Marrin, Personal communication, 15 November 2020
[71] Interview with senior member of the Defence Intelligence professionalisation team. See also Borg, "Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology," 16.
[72] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020, 20–21.

techniques to use when.[73] Additionally, the NORDIS model focuses on an externalization and visualization of the analytic process, which can bring new or alternative inputs as well as transparency to the process, including a more explicit differentiation of information, assumptions, and assessments. By synthesizing these concepts into a comprehensive and iterative methodology, the NORDIS approach brings focus to courses of action, scenarios, and relevant indicators, key elements of both warning and forecasting in estimative intelligence products. Most importantly, however, the NORDIS SATs-methodology combines the inductive nature of pattern analysis with a deductive nature of creating models and a mainly abductive nature of hypotheses testing, thereby also stimulating creative and critical thinking. In sum, the NORDIS SATs-methodology adhere to the principles of increased analytic rigour and improved uncertainty handling, thereby reducing the risk of discourse failure.

The NORDIS SATs-methodology is significantly different than the approaches of American intelligence agencies and PHIA in the United Kingdom but has its equivalent in other small European nations. The SATs-model taught by the Netherlands Joint ISTAR Command is almost a blueprint of the NORDIS model, from problem analysis to producing an intelligence estimate.[74] In Sweden, SATs are taught as tools in support of early warning, helping analysts identifying changes in current trends.[75] Consequently, opposed to a mainly civilian ad-hoc approach there seems to be a military North European integrated social science approach to the use of SATs in estimative intelligence analysis.

Summary

The current research on Structured Analytic Techniques (SATs) lacks nuance by assuming that SATs are taught in a uniform manner. This chapter has revealed that there are distinct methodological differences among different nations in how SATs are being taught. The typical American approach is one of teaching single techniques as tools in a toolbox that

[73] Borg, "Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology."
[74] Interviews H & I, The SATs-methodology is also described in "*Instructiekaart inlichtingen analyse*", IK 30-1, SAP10002050512, Ministerie van Defensie
[75] Henrik Häggström, Swedish Defence University, personal communication 13 December 2021.

lacks procedural understanding of how to best implement different SATs in the overall analytical process. The British approach to SATs is largely based upon a Hegelian triad of thesis, antithesis, and synthesis where techniques are differentiated by problem types. Unlike DI, however, PHIA does not teach a layered use of the techniques where the output of one is used as input the next.

Both the single techniques approach and the division of different pathways for different questions can be problematic as analysts do not necessarily see the linkage between the different techniques, especially when hypotheses and scenarios are treated differently. Moreover, the British and American approaches seem more focused on using SATs as techniques for single projects, not as an integrated part of the overall intelligence process. One can thus get the perception that some of the people responsible for teaching SATs today, and the academics researching the effect, have not understood that the people who came up with all these SATs used them within an analytical framework, not as one-offs. Consequently, when the techniques are being used in a piecemeal fashion instead of as part of a system, the effect is arguably also piecemeal. The ameliorative effect of applying SATs becomes a tincture dissolved in a much larger volume of unstructured analysis, with all the well-documented pitfalls that this entails.

The NORDIS SATs-methodology ventures further than most other approaches. By integrating different elements of logic into sensemaking and combined with creative and critical thinking in collaborative groups of intelligence analysts, the NORDIS SATs-methodology synthesizes several concepts into a more comprehensive and iterative framework. This approach to teaching SATs can consequently be seen as an integrated social science approach to intelligence analysis methodology as opposed to the common piecemeal approach. The iterative nature of the NORDIS model furthermore combines long term analysis with current intelligence, enhancing the process of indication and warning in several timeframes. In that manner it sets the foundation for a more structured as well as more creative and critical approach by describing the use of different analytic techniques, the relationship between them, and the problem at hand. Moreover, the methodological foundation is based upon a mix of inductive, deductive, and not the least abductive logic, thereby being set to produce new knowledge and understanding. Accordingly, the NORDIS

SATs-model accomplishes two main objectives. First, it provides intelligence analysts with a framework and a common taxonomy of structured techniques beyond what has been presented by others. Second, it supports collaboration, ensures traceability, and exposes analysts to divergent and conflicting perspectives, thereby creating better shared mental models. Most importantly, however, using a comprehensive structured methodology will arguably increase analytic rigour in the estimative process as well as improve analysts' ability to assess uncertainty, key mitigations to intelligence failure. The next part will then investigate the validity of these claims.

Part 3: Data analysis & discussion

The Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 gave, as discussed in part 2, an increased community wide focus on improving the tradecraft of intelligence analysis to prevent surprises and discourse failure. A result is that Structured Analytic Techniques (SATs) have become the intelligence community's go-to approach for this tradecraft improvement. Succeeding this latest of intelligence reforms came a call for validation of the new so-called alternative analysis methods. Tetlock, for instance, argued that:

> Players high up in the political system—who really do want the best-possible forecasts—could decide that it is worth investing a nontrivial share of their intelligence agencies' budgets into a series of long-term forecasting tournaments designed to distinguish the more from the less promising forecasting approaches across policy problems.[1]

But research on the effect of SATs has been limited and apart from Coulthart's research, it has had a narrow focus, investigating the effect on bias mitigation or judgement accuracy of a single technique: Analysis of Competing Hypotheses (ACH). Consequently, the current research has investigated topics that are arguably somewhat subsidiary to what the Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 called for: increased analytic rigour, as in analytic objectivity and analytic integrity, and improved communication of uncertainty, as in statements of both probability and analytic confidence. This research therefore aims at investigating the effects using a comprehensive, layered, and iterative SATs-methodology has on the wanted outputs of the reform. This approach has led to one main research question and three sub questions:

Main RQ: What are the experienced effects of using a comprehensive Structured Analytic Techniques (SATs) methodology in estimative intelligence?
Sub RQ 1: What is the effect of SATs on the ability to assess uncertainty?
Sub RQ 2: What is the effect of SATs on analytic rigour?
Sub RQ 3: What is the main impact on analysts' usage of SATs?

[1] Tetlock, "Reading Tarot on K Street."

To answer these questions, two surveys have been issued, one among participants of the Norwegian Defence Intelligence School (NORDIS) SATs-course and one among intelligence analysts in three Norwegian Intelligence agencies: The Norwegian Intelligence Service (NIS), The Norwegian Army Intelligence Battalion (IntBn), and the Norwegian Police Security Service (*Politiets sikkerhetstjeneste*, PST). The survey results were thereafter compared to qualitative data retrieved from a few semi-structured interviews with experienced intelligence analysts.

Chapter 7: Survey 1: the effects of the NORDIS SATs-course

Participants

Survey 1 was conducted as a pre- and post-course questionnaire issued to participants on the NORDIS SATs-course between August 2017 and January 2019. The survey consisted of 120 respondents after removal of 9 cases that had answered only either the pre- or post-course part of the survey. The gender mix among the respondents was 77% male and 23% female and with an age mix of 62% between 20 and 30 years of age and 30% between the age of 31 and 40. 84% of the respondents had a military rank while 16% were civilians. The mix between respondents from the Armed Forces and the Norwegian Intelligence Service, however, was 55% and 45% respectively. The level of education varied from high school to PhD, where 31% had a high school diploma, 42% had an undergraduate degree and 27% had a postgraduate degree. Years of service was divided between one-third for both 2-4 years and 10+ years while the last one-third was divided equally between 1 year or less and 5-9 years. At the same time, 61% reported to have intelligence experience of one year or less, 24% had 2-4 years, 9% had 5-9 years and 6% had 10 years or more.

The demographic statistics for the respondents show that the average course participant was a young military male with little intelligence experience holding a bachelor's degree. At the same time there was some variance to this. The Norwegian Armed Forces, which the Norwegian Intelligence Service is a part of, is still predominately a male organisation, despite opening for females in 1977 and introducing obligatory military service for both genders in 2015. The 70-30 mix of males-females is as such well within the general gender mix for the armed forces. The other demographic outputs can largely be explained with the fact that there has generally always been a mix of non-commissioned officers (NCOs), officers, and civilian analysts attending the NORDIS SATs-course. Most course participants attend the course early in their intelligence career, but several officers and NCOs have had several other postings before turning to intelligence. Furthermore, in general, NCOs have a high school background, officers have a bachelor's degree from a military academy or a master's degree from staff college, while civilian analytic positions usually require a master's degree. Lastly, the SATs-course has become mandatory in most all intelligence job descriptions, explaining why there are also some older and more experienced (10+ years) people attending.

Exploring, screening, and cleaning the data

Missing data

The missing value analysis (MVA) in SPSS showed that the maximum missing data was 2.5% (one variable) and showing no obvious predictable pattern. The result of Little's MCAR test on variables with more than 1% missing data was also nonsignificant, $p = .20$, indicating that the missing data was missing completely at random and that any method of dealing with missing data was acceptable.[1] In order not to unnecessarily limit the sample size, the exclude cases pairwise option was the main choice when appropriate.

Reliability

The reliability analysis carried out on all 33 scale-items in Survey 1, showed an acceptable reliability of Cronbach's alpha = .75. Nonetheless, investigating the other outputs from the reliability test revealed some issues that needed consideration.[2] First, there were several small negative values in the inter-item correlation matrix despite no need to reverse scores. Second, while there were no negative values in the list of corrected-item total correlation values, there were as many as 19 items having a corrected item-total correlation value of <.3, suggesting a removal. Third, the inter-item correlation showed a mean = .09 with values ranging from -.26 to .60, indicating a low internal consistency. At the same time, the Cronbach's alpha if deleted showed that the removal of any single variable would not increase the alpha. With as many as 33 items the likelihood of a reliable alpha is high, but this does not prove that the questionnaire is unidimensional.[3] Instead, the high number of low inter-item correlations rather suggested that the questionnaire consisted of several sub-scales that needed to be tested for reliability by themselves.[4] A subsequent factor analysis

---

[1] Tabachnick and Fidell, *Using Multivariate Statistics: Pearson New International Edition*, 100.
[2] Pallant, *SPSS Survival Manual*, 104.
[3] Cortina, "What Is Coefficient Alpha? An Examination of Theory and Applications."
[4] Field, *Discovering Statistics Using IBM SPSS Statistics*, 709.

using SPSS showed that the questionnaire did indeed consist of several separate subscales that measured different, although related constructs.

Assessing normality

Running the explore function in SPSS showed that both the Kolmogorov-Smirnov and the Shapiro-Wilk tests of normality were significant, $p < .001$, indicating that none of the items have a normal distribution. Furthermore, 22 items had a $M > 4$ ($SD < 1$), resulting in a large negative skewness, up to -2.49 ($SE = .22$), and a high positive kurtosis, up to 9.85 ($SE = .44$). An examination of the bar graphs, box plots and Q-Q plots for these 22 items revealed that neither of them had a normal distribution, all also having one or more outliers. Outliers can typically be dealt with by deleting cases or items or by changing the scores of the outliers to reduce the impact.[5] On the other hand, the largest difference between mean and 5% trimmed mean was .13, indicating that the outliers would not interfere with any analytic tests and their values were thus not changed.[6] The remaining 11 items had a much lower degree of skewness as well as kurtosis. Nevertheless, only eight of these items showed to have a more normal distribution in bar graphs, box plots and Q-Q plots. Since the data was on the threshold of normal distribution, it was appropriate to employ both the parametric and the non-parametric tests, but to report results from the non-parametric test unless when there were differences between the two.

Exploratory Factor Analysis

Factorability

Since the results of the reliability test indicated that survey 1 could consist of several separate subscales, exploratory factor analysis (EFA) was carried out using SPSS.
Initially, the factorability of the data was assessed. A first issue to consider was the adequacy of the sample size, where a general rule-of-thumb has been at least 300 cases.[7] However, 100 cases can be enough if there are well defined factors, enough high loading marker

---

[5] Tabachnick and Fidell, *Using Multivariate Statistics: Pearson New International Edition*, 111.
[6] Pallant, *SPSS Survival Manual*, 63.
[7] Field, *Discovering Statistics Using IBM SPSS Statistics*, 684.

variables (>.80) and communalities in the range of .5.[8] An opposing view is to focus on the cases to variables ratio where a five to one ratio is assessed as adequate.[9] A second issue is to screen the correlation matrix for coefficients greater than .3.[10] Third, other issues to consider is a Kayser-Meyer-Olkin (KMO) value above .5, Bartlett's test of sphericity being significant, no diagonals in the anti-image correlation matrix less than .5, and a determinant above 0.[11]

Factorability was assessed using all 33 scale variables in the survey. The sample size of 120 cases in survey 1 was on the low side, resulting in a cases-to-variables ratio of 3.6. The correlations matrix revealed 66 correlations with a coefficient > .30, suggesting that factor analysis still could be valuable, even though the determinant implied singularity (value = 3.022E-6). The Kayser-Meyer-Olkin (KMO) value was .66, a mediocre value according to Hutcheson and Sofroniou.[12] Bartlett's Test of Sphericity reached statistical significance ($p <$ .001), although that has less relevance given the number of respondents.[13] More importantly, while communalities were above .5 for all but 1 variable (.49), the anti-image correlation matrix revealed the presence of 4 diagonals with a value under .5. Consequently, some results supported factorability while others did not.

Principal component analysis output

The initial factor analysis in SPSS using PCA with oblique rotation (direct oblimin) was ambiguous, revealing 11 components while the scree plot had points of inflection at both 4 and 8 components. The components also contained an unsystematic mix of pre- and post-course items, suggesting that the number and content of components was not optimal, thereby underlining the problems of factorability mentioned above. Removing items that scored low on communalities or anti-image diagonals or loaded low and/or equal on more than one variable, resulted in a principal component analysis conducted on 17 items using

---

[8] MacCallum et al., "Sample Size in Factor Analysis."
[9] Pallant, *SPSS Survival Manual*, 184.
[10] Tabachnick and Fidell, *Using Multivariate Statistics: Pearson New International Edition*, 667.
[11] Field, *Discovering Statistics Using IBM SPSS Statistics*, 686–87.
[12] Hutcheson and Sofroniou, *The Multivariate Social Scientist*, 225.
[13] Field, *Discovering Statistics Using IBM SPSS Statistics*, 685–86.

oblique rotation (direct oblimin).[14] This gave an adequate case to item ratio of 7 to 1. Factorability was furthermore supported with a determinant of .004, a KMO value = .77 (a middling value) and Bartlett's Test of Sphericity reached statistical significance ($p < .001$). Moreover, there were no anti-image diagonals or communalities < .5. The analysis revealed 5 components with an eigenvalue greater than 1, explaining a total of 62.7% of the variance and with 49 % nonredundant residuals with absolute values greater than .05. Both the pattern matrix and the structure matrix showed a clear and simple structure of 2 to 7 items on each component except for one item loading equally on two components. Removing this item had a negative effect on the factor analysis, resulting in a lower KMO, diagonals and communalities < .5, and higher residuals, suggesting that the item should be kept. The structure matrix revealed that all items loaded greater than .56 and with at least one high loading of .8 on all but 1 component. Furthermore, running Horn's parallel analysis revealed no over-dimentionalisation of factors.[15]

The cluster of items on the components suggested that component 1 represented the post-course experienced effect using SATs have on ability to assess uncertainty, component 2 represented the pre-course ability to assess uncertainty, component 3 represented the assumption level's effect on analytic confidence, component 4 represented the post-course experienced effect using SATs have on analytic rigour, and component 5 represented pre-course SATs' expected effect on ability to assess uncertainty.

Component analysis

Component reliability and test of normality

The reliability of the components was tested, resulting in a Cronbach's alpha = .83 for component 1, .65 for component 2, and .70 for component 5. Using Cronbach's alpha to test for reliability on a two-item scale usually produces underestimated reliability and therefore the Spearman-Brown coefficient is preferable.[16] Here, however, the reliability test for the two 2-items components was .33 for component 3 and .71 for component 4 on both tests.

[14] Appendix 5
[15] Horn, "A Rationale and Test for the Number of Factors in Factor Analysis."
[16] Eisinga, Grotenhuis, and Pelzer, "The Reliability of a Two-Item Scale: Pearson, Cronbach, or Spearman-Brown?"

While component 3 was discarded in the following analysis due to clear unreliability, component 2 was kept due to being close to an alpha = .7.

The means of the four remaining components were calculated using the transform function in SPSS and are presented in table 1.

	N	Mean (Median)	SD
Pre-course ability to assess uncertainty (component 2)	120	2.98 (3.00)	.75
Pre-course SATs' expected effect on ability to assess uncertainty (component 5)	120	4.59 (4.67)	.51
Post-course SATs' experienced effect using on ability to assess uncertainty (component 1)	120	4.25 (4.29)	.51
SATs' experienced effect on analytic rigour (component 4)	118	4.48 (4.50)	.57

Table 1: Descriptive statistics survey 1 components

Exploring the components produced to a large degree the same result as with the original 33 items, a large negative skewness (up to -2.9 (*SE* = .22)) and a large positive kurtosis (up to 14.7 (*SE* = .44)), on all but component 2, pre-course ability to assess uncertainty (skewness = -.42 (*SE* = .22), kurtosis = -.22 (*SE* = .44)). The lack of normality was supported by the histograms, box plots and Q-Q plots, although component 2 was bordering normality. Although the four other components had one or two outliers, the difference between the mean and the 5% trimmed mean was 0,1 at maximum and values were not changed as the outliers would not affect the analysis.[17]

Component correlations

For an initial examination of the relationship between the components, a bivariate correlation analysis was conducted. There was no significant relationship between pre course ability to assess uncertainty and any of the other components, p > .05. For the other components there were correlations, although a non-conclusive correlation for one of the components. Bias corrected and accelerated bootstrap 95% Cis are reported in square brackets. Pre-course SATs' expected effect on assessing uncertainty was significantly correlated with post-course SATs' experienced effect on ability to assess uncertainty, *r* = .51

[17] Pallant, *SPSS Survival Manual*, 64–65.

[.18, .71], $p < .001$, with a nonparametric correlation $r_s = .32$ [.12, .49], $p < .001$. There was also a significant correlation between post-course SATs' experienced effect on ability to assess uncertainty and the post-course SATs' experienced effect on analytic rigour, $r = .44$ [.15, .67], $p < .001$, $r_s = .28$ [.11, .45], $p < .001$. While Pearson's correlation showed a significant correlation between pre-course expected SATs' effect on ability to assess uncertainty and post-course SATs' experienced effect on analytic rigour, $r = .40$ [.02, .65] ($p < .001$), there was no significant correlation using Spearman's correlation. Using Cohen's interpretation of the strength of the correlations, the values suggest a medium to large positive relationship, where especially high levels of expected effect were associated with high levels of experienced effect on assessing uncertainty.[18]

The effect of using SATs on analysts' ability to assess uncertainty

The output of Explore in SPSS revealed what seemed like a large difference in means between the components. To investigate this difference for the three components measuring uncertainty, the non-parametric Friedman test was run. The medians are presented in figure 9. There was a significant effect of using SATs on analysts' ability to assess uncertainty, $X^2(2) = 167.26$, $p < .001$.[19]

[18] Cohen J., *Statistical Power Analysis for the Behavioural Science*, 79–81.
[19] Appendix 6, a

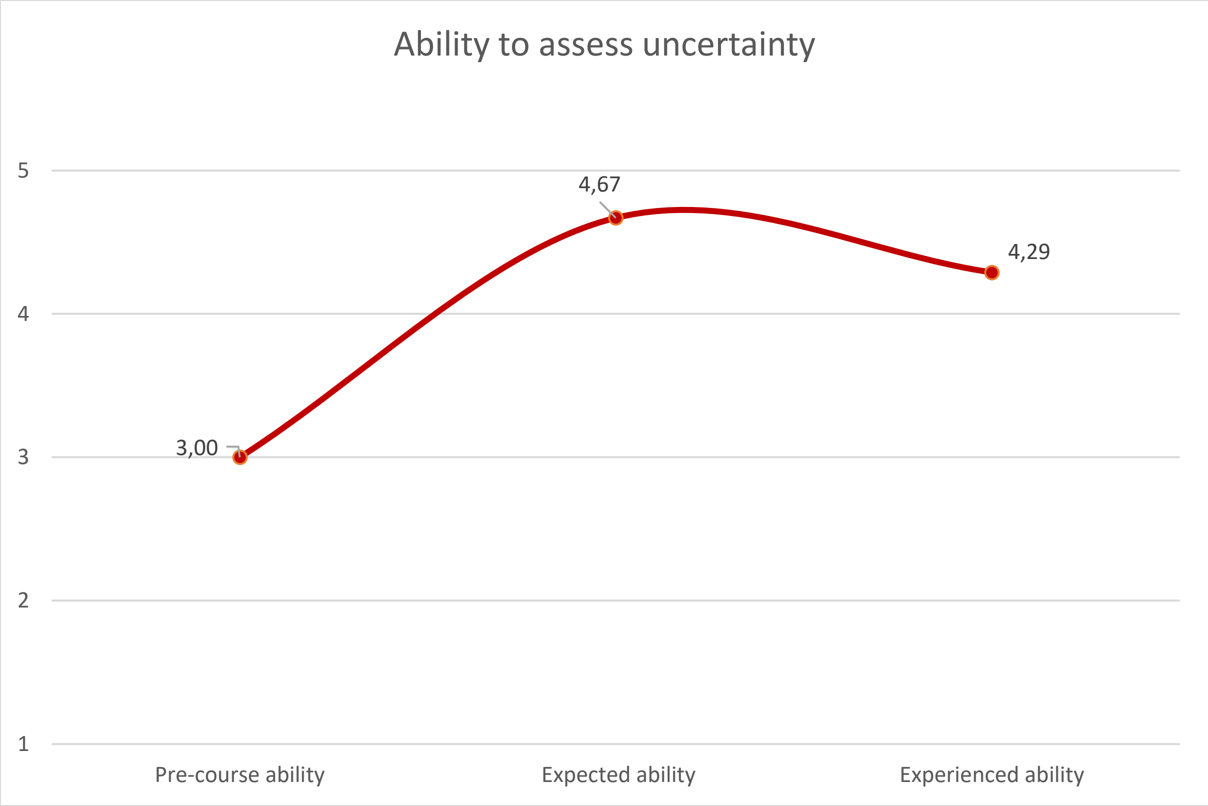


Figure 9: Descriptive statistics Analysts' ability to assess uncertainty

The effect size can be calculated in several ways. A common method is conducting Wilcoxon signed-rank tests and thereafter calculate effect by using *z*-scores divided by the square root of observations in total.[20] Another way is a simple difference formula using rank sums to obtain a matched-pairs rank-biserial correlation.[21] Both these approaches measure effect size only between two points of time. As an alternative, calculating the coefficient of concordance (*W*) can produce an overall effect size when using the Friedman test.[22] By using this calculation, the effect of using SATs on assessing uncertainty $W = .70$. a very large effect using an adapted version of Cohen's guidelines.[23]

To disclose the effect between time 1 to time 3 and time 2 to time 3, the pairwise comparison of the Friedman test output in SPSS was investigated. The pairwise comparison using the Dunn-Bonferroni post hoc test showed a significant difference with a large effect

[20] Pallant, *SPSS Survival Manual*, 236; Field, *Discovering Statistics Using IBM SPSS Statistics*, 257.
[21] Kerby, "The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation."
[22] Gray and Kinnear, *IBM SPSS Stat. 19 Made Simple*, 332.
[23] Rosenthal, "Qualitative Descriptors of Strength of Association and Effect Size." Effect size using *W* can be interpreted the same way as for *r,* see Gray and Kinnear, *IBM SPSS Stat. 19 Made Simple*, 333.

between post-course experienced effect on assessing uncertainty (time 3) and the pre-course ability (time 1), $z = -8.04$, $p < .001$, $r = .52$. At the same time, there was also a significant difference with a small effect between the post-course experienced effect on assessing uncertainty (time 3) and the pre-course expected effect on assessing uncertainty (time 2), $z = 4.45$, $p < .001$, $r = .29$.

Probability and analytic confidence

Investigating the two main aspects of assessing uncertainty, probability, and analytic confidence, a similar, but nevertheless also slightly different picture emerged, as depicted in figure 10. The Friedman test showed a significant difference with a large effect between pre-course easiness of ascribing probability (time 1) (*Mdn* = 3.0), pre-course SATs makes it easier to ascribe probability (time 2) (*Mdn* = 5.0), and post-course SATs helped in ascribing probability (time 3) (*Mdn* = 4.5), $X^2(2) = 141.20$, $p < .001$, $W = .59$.[24] The pairwise comparison using the Dunn-Bonferroni post hoc test showed a significant difference with a large effect between time 1 and time 2, $z = -9.84$, $p < .001$, $r = .64$, and between time 1 and time 3, $z = -8.26$, $p < .001$, $r = .53$, but a non-significant difference between time 2 and time 3, $z = 1.58$, $p = .34$, $r = .10$.

There was also a significant difference with a large effect between pre-course easiness of stating analytic confidence (time 1) (*Mdn* = 3.0), pre-course SATs will help state analytic confidence (time 2) (*Mdn* = 5.0), and post-course SATs made it easier to ascribe a level of analytic confidence (time 3) (*Mdn* = 4.0), $X^2 (2) = 120.39$, $p < .001$, $W = .52$.[25] The pairwise comparison using the Dunn-Bonferroni post hoc test showed a significant difference with a large effect between time 1 and time 2, $z = - 9.85$, $p < .001$, $r = .63$, a significant difference with a medium effect between time 1 and time 3, $z = - 6.40$, $p < .001$, $r = .42$, as well as a significant difference with a small effect between time 2 and time 3, $z = 3.45$, $p = .002$, $r = .23$.

[24] Appendix 6, b
[25] Appendix 6, c

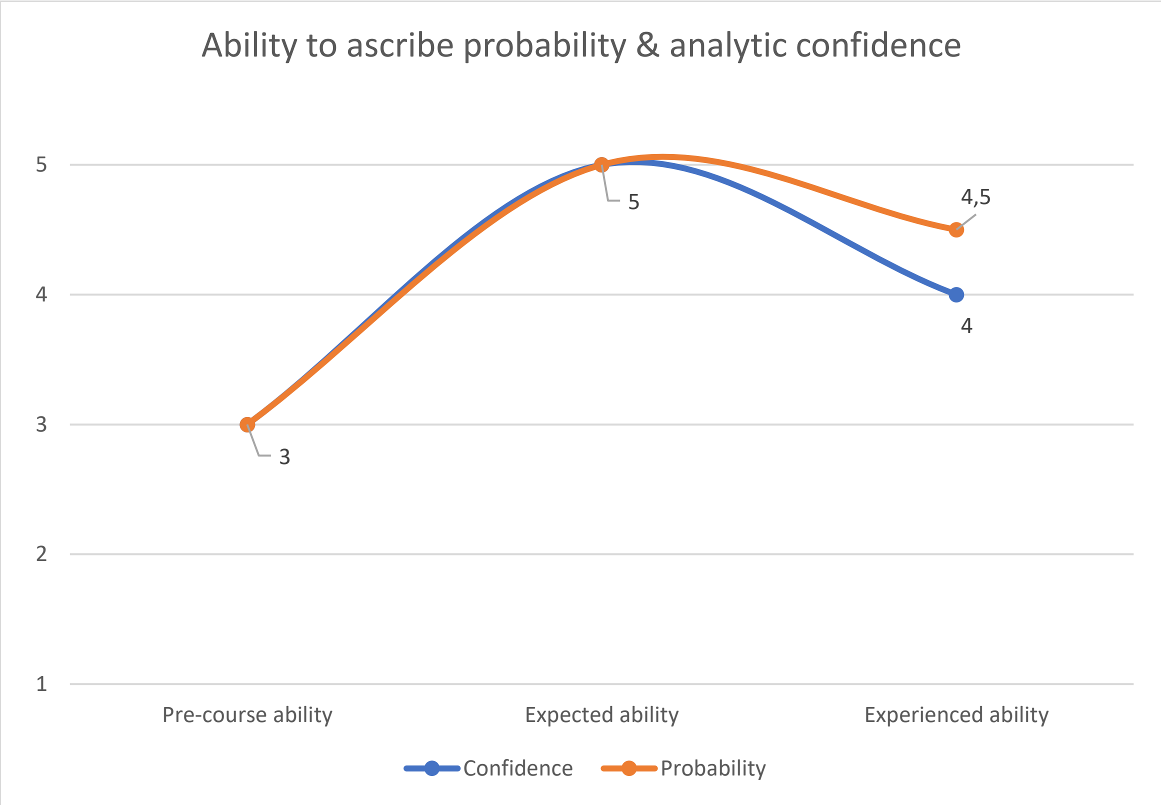


Figure 10: Descriptive statistics Probability vs Analytic confidence

A more detailed investigation of items measuring probability and analytic confidence shed more light on differences for the demographic variable unit. An independent-samples Mann-Whitney U tests showed that there was a significant difference in ranks with a small effect on pre-course easiness of ascribing probability between personnel from the Armed Forces (AF) (Mdn = 3.00/*mean rank* = 66.83, *n* = 66) and from the Intelligence Service (NIS) (Mdn = 3.00/*mean rank* = 52.76, *n* = 54), *U* = 1364.00, *z* = -2.37, *p* = .022, *r* = .21.[26] Post-course, however, there was no difference in ranks of post-course statement Q1, SATs made ascribing probability easier, *U* = 1611.5, *z* = -1.01, *p* = .314, *r* = .09.

For the analytic confidence dimension, an independent-samples Mann-Whitney U tests showed that there were significant differences in ranks between personnel from the Armed Forces (AF) and from the Intelligence Service (NIS) on three items.[27] First, there was a significant difference with a small effect in ranks of pre-course high analytic confidence in assessments (pre-course statement Q4) between AF (Mdn = 3.50/*mean rank* = 65.45, *n* = 66)

[26] Appendix 6, d
[27] Appendix 6, e

and NIS (Mdn = 3.00/*mean rank* = 53.22, *n* = 53), *U* = 1389.50, *z* = -2.04, *p* = .041, *r* = .19. Second, post-course there was a significant difference with a small effect in ranks of SATs made it easier to ascribe a level of analytic confidence (post-course statement Q8) between AF (Mdn = 4.00/*mean rank* = 64.66, *n* = 66) and NIS (Mdn = 4.00/*mean rank* = 52.17, *n* = 53), *U* = 1334.00, *z* = -2.16, *p* = .030, *r* = .20. Third, post-course there was also a significant difference with a small effect in ranks of structured methodology is an important factor for rating analytic confidence (post-course statement Q14) between AF (Mdn = 5.00/*mean rank* = 65.91, *n* = 66) and NIS (Mdn = 5.00/*mean rank* = 53.89, *n* = 54), *U* = 1425.00, *z* = -2.37, *p* = .018, *r* = .22.

SATs' effect on analytic rigour

The analytic rigour component from the factor analysis consisted of two items, the degree of SATs improving analytic objectivity and the degree of SATs improving analytic integrity. The medians, means and standard deviations are presented in table 2.

	N	**Median (Mean)**	**SD**
SATs' experienced effect on analytic rigour (comp. 4)	118	4.50 (4.48)	.62
SATs improved analytic objectivity (post-course Q6)	118	4.50 (4.39)	.73
SATs improved analytic integrity (post-course Q5)	118	5.00 (4.57)	.67

Table 2: Descriptive statistics the experienced effect of using SATs on analytic rigour

Running a series of non-parametric tests on component 4 revealed that there was evidence of a significant difference in ranks for only one demographic variable. A Kruskal-Wallis test showed a significant difference in SATs' experienced effect on analytic rigour across the categories of years of service *H*(3) = 9.78, *p* = .021.[28] The Dunn’s pairwise tests with Bonferroni adjusted *p*-values for the four groups revealed that there was a significant difference with a medium effect between people with 5-9 years of service (*Mdn* = 4.00/*mean rank* = 43.89) and people with 1 year or less of service (*Mdn* = 5.00/*mean rank* = 73.87), *p* = .026, *r* = .45. There was no evidence of a difference between the other pairs. Investigating the two items, analytic objectivity and analytic integrity separately, however,

[28] Appendix 7

showed that there were no demographic differences on levels for SATs improved analytic objectivity for any of the demographic variables. The only difference between demographic variables was in distribution of SATs improved analytic integrity across years of service, $H(3) = 9.88$, $p = .020$. The Dunn's pairwise tests with Bonferroni adjusted *p*-values for the four groups revealed that there was a significant difference with a medium effect between people with 5-9 years of service (*Mdn* = 4.00/*mean rank* = 43.05) and people with 1 year or less of service (*Mdn* = 5.00/*mean rank* = 71.00), $p = .017$, $r = .47$. There was no evidence of a difference between the other pairs.

Discussion

This research set out to investigate analysts' experienced effect of using a comprehensive Structured Analytic Techniques (SATs)methodology for estimative intelligence analysis. For survey 1 the research was limited to investigating the effects a two-week course learning to use the NORDIS SATs-methodology has on analytic rigour and on analysts' ability to assess uncertainty. One can conclude there was a correlation with medium to large effect where high levels of pre-course expected effect were associated with high levels of post-course experienced effect of the dimension of ability to assess uncertainty. Given that the data was bordering normality, there is also a tendency towards conclusiveness, especially post-course, with regards to the correlation between SATs' effect on ability to assess uncertainty and SATs' effect on analytic rigour.

For the uncertainty dimension, the post-course experienced effect on assessing uncertainty was significantly higher, with a large effect, than how the participants rated their pre-course ability. Simultaneously, the experienced effect was significantly lower, although with a small effect, compared to the participants' pre-course expectations to the effect the NORDIS SATs-methodology has on analysts' ability to assess uncertainty. A more detailed investigation revealed that while the pre-course expectations of SATs' effect on ascribing probability were met, there was a small significant negative difference between the pre-course expectations and the post-course experienced effect on ascribing analytic confidence. This indicates that properly distinguishing between information, assumptions, and assessments in the estimative process makes it easier to decide where key judgements belong on a standardized probability scale like the ones described in intelligence doctrines. It also indicates that there are more factors impacting analytic confidence than just employing SATs, such as the credibility and quality of the information or any of the other factors previously discussed in chapter 5.

The data analysis furthermore revealed that there were a few significant differences of the experienced effect of using SATs on the ability to assess uncertainty between respondents of different units, providing nuances to the general results. Personnel from the Armed Forces (AF) expressed higher ability to assess probability and analytic confidence pre-course than personnel from the Norwegian Intelligence Service (NIS). Post-course there was no difference in the dimension of probability while the AF respondents found the NORDIS SATs-methodology to be more effectful on analytic confidence. Together this could indicate that there is also a cultural difference that impact intelligence analysis methodology and the effect on assessing uncertainty, an important point that was investigated further in survey 2.

For the analytic rigour dimension the analysis showed that using a comprehensive SATs-methodology improved analysts' analytic rigour by improving both analytical objectivity and analytical integrity. There was nevertheless a difference to this general result as there was a significant difference in SATs' effect on analytic rigour between inexperienced course participants and experienced ones. The significant difference lay not in SATs' effect on analytic objectivity, but the effect on analytic integrity where the more experienced (5-9 years) reported lower effect. It is, however, difficult to say whether the results point towards a Dunning-Kruger effect where the novice show overconfidence in the effects of SATs or if the results point more towards a skilled-unaware pattern as found in Burson et al.[29]

Based on the analysis of the survey 1 data, one can conclude that the use of a comprehensive Structured Analytic Techniques (SATs)-methodology for estimative intelligence analysis fulfils two of the key requirements of the post-9/11 intelligence reform. Participants of the NORDIS SATs-course, centred around a comprehensive, layered, and iterative use of several key SATs, experienced a large positive effect both on analytic rigour

[29] Burson, Larrick, and Klayman, "Skilled or Unskilled, but Still Unaware of It: How Perceptions of Difficulty Drive Miscalibration in Relative Comparisons."

and on their ability to assess uncertainty. Furthermore, there was a significant correlation between the SATs-methodology experienced effect on analytic rigour and the experienced effect on assessing uncertainty, indicating that these two issues are more interlinked than what has been previously presented in intelligence theory literature. Consequently, the mix of a creative, critical, and sensemaking logic-process as the NORDIS SATs-methodology enables analysts to approximate a pragmatic 'best truth' in their estimates. Two key questions nevertheless emerged from the general findings. First, do the effects of being taught and trained in the use of a comprehensive SATs-methodology last when the analysts go back to their workplaces, dealing with their everyday intelligence problems? Second, are there any inter-agency cultural differences that impact the effect and usage of SATs? The follow-up survey 2 was conducted to research these matters.

Chapter 8: Survey 2: Long-term effects of SATs training and use

Participants

To investigate long-term effects of having received SATs-training, survey 2 was conducted in January and February 2019 by sending out a questionnaire to intelligence analysts in three different intelligence agencies in Norway, the Norwegian Intelligence Service (NIS), the Norwegian Army Intelligence Battalion (IntBn), and the Norwegian Police Security Service (PST). A total of 105 respondents participated in survey 2, and just as for survey 1, reporting a response rate would be in breach with Norwegian security regulations. Additionally, due to the need for absolute anonymity for the respondents, it was not possible to find out how many of the respondents from survey 1 participated in survey 2. As part of exploring, screening, and cleaning the data, nine cases were excluded due to missing data, leaving 96 cases for the data analysis.

Output of descriptive statistics in SPSS showed that 53% of the respondents came from the Norwegian Intelligence Service (NIS), 29% from the Intelligence Battalion (IntBn), and 18% from the Police Security Service (PST), with an overall mix between civilians and military personnel of 41% and 59% respectively. However, while all respondents from the IntBn had a military rank and all respondents from the PST were civilians, respondents from the NIS were divided by 57% military personnel and 43% civilians. The gender mix was 70% male and 30% female, a slight difference from survey 1. The age mix was also different, with 48% between 20 and 30 years of age, 30% between 31 and 40 and as many as 22% of more than 40 years of age, the latter a large difference from the 8% in survey 1. For other demographic variables such as education, years of service, and experience, there were also differences to survey 1. The level of education still varied from high school to PhD, but in survey 2 only 8% had a high school diploma, 43% had an undergraduate degree and as many as 49% had a postgraduate degree. Years of service was divided between 15% having one year or less, 31% with 2-4 years, 25% with 5-9 years, and 29% with 10+ years. In survey 2, years of experience was more in line with years of service where 21% reported to have intelligence experience of one year or less, 33% had 2-4 years, 28% had 5-9 years and 18% had 10 years or more. While in survey 1, where all respondents were subject to the NORDIS SATs-course, in survey 2 13% of the respondents reported to have had other SATs-training and 10% reported to not have attended any formal SATs-course. Time from attending SATs training varied where half of the respondents had attended a SATs-course more than 3 years ago.

Exploring, screening, and cleaning the data

Missing data

In survey 2, the dataset initially contained 105 cases. The result of the initial missing values analysis showed that the amount of missing data could be an issue and nine cases were excluded due to large amount of missing data coded 99 (61-94% missing), all cases from NIS where the respondents had answered they never used SATs and thereafter chose not to answer any more questions. Even after excluding these cases, there were 12 variables with missing data of more than 5% where the maximum of missing data was 13,5% (one variable), resulting in a valid N (listwise) of 77 for scale data and 93 for categorial data. However, coding ‘Not relevant’ as missing (66) explained the percentage of missing data and the maximum of actually missing data (99) was 1%. The difference was evident in the results of Little's MCAR test, being significant $p = .003$ with 'Not relevant' set as missing and $p = 1.00$ when only including missing data coded 99. Furthermore, running MVA for only the variables with more than 1% missing (both 99 and 66) revealed a Little's MCAR test $p = .15$ and with no obvious pattern of missing data. In conclusion, the MVA indicated that the data was missing at random and that both the exclude cases pairwise and the exclude listwise option could be applied, where which one to use was dependent on the degree of missing data for the items being analysed.

Reliability

The reliability analysis of the 22 scale items in survey 2 showed a Cronbach's alpha = .90, a more than appropriate result, indicating a large degree of internal consistency. Notwithstanding the alpha, investigating the other outputs from the reliability test revealed some issues. First, there were a few negative values in the inter-item correlation matrix despite no need to reverse scores. Second, while there were no negative values in the list of corrected-item total correlation values, three of the variables had a corrected item-total correlation value of <.3, suggesting a removal. Third, the summary of inter-item correlations showed a *M* = .29, range -.17 to .90, indicating a low to medium internal consistency.[1] Lastly, while the Cronbach's alpha if deleted showed that the removal of any single variable would not increase the alpha significantly, the item-total statistics suggested removal of three items.

Survey 1 analysis showed that a high number of variables can result in a high alpha without the questionnaire being unidimensional but rather consisting of several subscales.[2] A subsequent factor analysis revealed a two-factor structure where the three items with low reliability were left out and without representing a scale of their own. A removal of the three abovementioned items resulted in an alpha of .93 and an inter-item correlations *M* = .43, range .22 to .91. Despite that, keeping the three items made it possible to shed some light on the actual usage of verbal probabilities and analytic confidence and they were thus kept as part of the data.

Assessing normality

The data was explored using descriptive statistics and graphs to check for any violations of statistical assumptions. Both the Kolmogorov-Smirnov and the Shapiro-Wilk tests of normality turned out p <.001 on all 22 items, indicating that none of the items had a normal distribution. On the other hand, only 6 variables had a large negative skewness (above –1.5), and a high positive kurtosis (above 2.3). The other 16 variables had both a more normal degree of skewness as well as kurtosis. By examining the bar graphs, box plots and Q-Q plots for all items, it was evident that as many as 13 items had a non-normal distribution, half also having one or more outliers. The outliers were assessed not to interfere with any analytic tests as the largest difference between mean and 5% trimmed mean was .16, their values were therefore not changed.[3] The remaining 9 variables showed to have a more normal distribution in bar graphs, box plots and Q-Q plots. All in all, the data in survey 2 showed a higher degree of normality than the data for survey 1.

---

[1] Pallant, 104.
[2] Field, *Discovering Statistics Using IBM SPSS Statistics*, 709.
[3] Pallant, *SPSS Survival Manual*, 63.

Due to the mix of normal and non-normal data, both parametric and non-parametric tests were used, dependent upon the variables in question. Just as for survey 1, for the non-normal data it was therefore appropriate to employ, and report results from the non-parametric tests unless when there were differences between the two.

Exploratory Factor Analysis

Factorability

Based upon the results of the reliability test, which suggested that the questionnaire consisted of two or more separate subscales, the 22 scale items in survey 2 were subjected to a principal components analysis (PCA) using direct oblimin as rotation method.[4] As for survey 1, the aim was to reduce the items to a smaller set of related components as there were theoretical grounds to suggest that the components would correlate.[5] First, the factorability of the data was evaluated. A sample of 96 cases was on the low side, especially since the cases to items ratio was less than 5. Nevertheless, the correlation matrix revealed the presence of a large number of coefficients > .3, the Kaiser-Meyer-Olkin (KMO) value was .84, (a meritorious value[6]), the Bartlett's Test of Sphericity was significant ($p < .001$) and there were no communalities < .5. On the other hand, the determinant was 0, indicating singularity, and there was one anti-image correlation as low as .24, questioning the factorability.

Principal component analysis output

The initial factor analysis in SPSS using PCA with oblique rotation (direct oblimin) produced an ambiguous result of 5 components while the scree plot suggested either a two- or four-component solution. Factor analysis was rerun several times, removing items that scored low on anti-image diagonals, communalities, and low and/or equal on more than one variable. In the end, a primary component analysis was conducted on 17 items using oblique rotation (direct oblimin).[7] Factorability was nevertheless still non-conclusive. On the one

[4] Appendix 8
[5] Field, *Discovering Statistics Using IBM SPSS Statistics*, 681.
[6] Hutcheson and Sofroniou, *The Multivariate Social Scientist*, 225.
[7] Appendix 8

hand, there was an adequate case to item ratio and Bartlett's Test of Sphericity reached statistical significance ($p < .001$). Moreover, the Kaiser-Meyer-Olkin (KMO) value = .87 (a meritorious value) and all anti-image diagonals were ≤ .8. On the other hand, the determinant was still 0, there were 56 % nonredundant residuals with absolute values > .05 and there was one communality < .5 (.43). The latter two was due to letting SPSS extract components based upon a fixed number of two components that was suggested by the scree plot instead of using the eigenvalue limit of 1. With only two components there was no need to run Horn's parallel analysis to assess over-dimensionalisation of factors.

The analysis revealed 2 components, explaining a total of 61.5% of the variance. Both the pattern matrix and the structure matrix showed a clear and simple structure with 10 items clearly loading on component 1 and 7 items clearly loading on component 2. Moreover, all items loaded between .60 and .87 and there were no items loading equally on two components. The pattern matrix suggested that component 1 constituted the respondents' opinion of the effect using SATs have on assessing uncertainty while component 2 constituted the degree of SATs usage among the respondents.

Component analysis

Component reliability and test of normality

The reliability of the components was tested, resulting in a Cronbach's alpha = .93 for component 1 and .88 for component 2. Although item 18, SATs improving analytic integrity, and item 19, SATs improving analytic objectivity, did not come up as a separate component in the factor analysis for survey 2, they were nevertheless combined into a component 3, SATs' effect on analytic rigour, to compare the data from survey 2 with the data from survey 1. The reliability of component 3 was tested, resulting in a Cronbach's alpha of .74, the same value as for the Spearman-Brown coefficient.

The means of the three components were calculated using the transform function in SPSS, outputs in table 3. Exploring the components revealed a lack of normality for 2 out of 3 components. The normality tests for component 2, SATs usage, seemingly turned out non-conclusive with the Kolmogorov-Smirnov test p =.026 while the Shapiro-Wilk test turned out *p* = .094. However, with an N = 96, the Shapiro-Wilk test is much more powerful than the Kolmogorov-Smirnov test.[8] Moreover, there was a Skewness of .03 (*SE* = .25) and a Kurtosis of -.60 (*SE* = .49), indicating a close to normal distribution. Normality was also supported by the histogram, box plot and Q-Q plot. For component 1, SATs' effect on assessing uncertainty, and component 3, SATs' effect on analytic rigour, however, both the Kolmogorov-Smirnov and the Shapiro-Wilk tests of normality turned out $p \leq .001$. Furthermore, the components had a negative Skewness and a positive Kurtosis of -1.51 (*SE* = .25)/3.86 (*SE* = .50) and –1.97 (*SE* = .26) /4.40 (*SE* = .51) respectively. The lack of normality for components 1 and 3 was supported by the histograms, box plots and Q-Q plots. Although the non-normal components had one or two outliers, the difference between the mean and the 5% trimmed mean was 0,1 at maximum and values were not changed as the outliers would not affect the analysis.[9] Given the results, component 2 was analysed using parametric tests while components 1 and 3 were analysed using non-parametric tests.

	N	**Mean (Median)**	**SD**
SATs' effect on assessing uncertainty (component 1)	90	3.98 (4.00)	.77
SATs usage (component 2)	96	2.61 (2.71)	.91
SATs' effect on analytic rigour (component 3)	89	4.53 (5.00)	.78

Table 3: Descriptive statistics survey 2 components

Component correlations

To examine the relationship between the components, a bivariate correlation analysis was run in SPSS. There were significant relationships between all the components and bias corrected and accelerated bootstrap 95% Cis are reported in square brackets. First, there was a significant correlation between SATs usage and SATs' effect on assessing uncertainty, $r = .48$ [.32, .63], $r_s = .48$ [.29, .64], $p < .001$. Second, there was also a significant correlation between SATs usage and SATs' effect on analytic rigour, $r = 39$ [.20, .54], $r_s = .33$ [.10, .52], $p \leq .002$. Third, there was a significant correlation between SATs' effect on assessing

[8] Mohd Razali and Bee Wah, "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests."
[9] Pallant, *SPSS Survival Manual*, 64–65.

uncertainty and SATs' effect on analytic rigour, $r = 60$ [.37, .76], $r_s = .38$ [.17, .56], $p < .001$. Using Cohen's interpretation of the strength of the correlations, the values suggest a medium to large positive relationship between the components.[10]

At the same time, a positive relationship does not necessarily imply causation, but one can use R^2 as a measurement for the amount of variance is shared by the components.[11] For survey 2, using Pearson's r, 23% of the variability of SATs usage shared the variability of SATs' effect on analytic uncertainty, 15% of the variability of SATs usage shared the variability of SATS' effect on analytic rigour, and 36% of the variability of SATs' effect on assessing uncertainty shared the variability of SATs' effect on analytic rigour. This still left a large amount of variability that was accounted for by other variables.

The effect of using SATs on analysts' ability to assess uncertainty

From the analysis of the survey 1 data, we saw there were significant differences in ranks for both the probability and confidence dimensions of assessing uncertainty, especially between personnel from the Norwegian Armed Forces and from the Norwegian Intelligence Service. For the survey 2 data, an independent-samples Kruskal-Wallis test showed that there was an almost significant difference in SATs' effect on assessing uncertainty across the three different units, the Norwegian Intelligence Service (NIS, $n = 45$), the Intelligence Battalion (IntBn, $n = 28$), and the Police Security Service (PST, $n = 17$), $H(2) = 5.92$, $p = .052$.[12] Making a decision not to reject the null hypothesis in this case could very well be an example of a dogmatic and dichotomous reject/not reject decision that could lead to a type II error.[13] Unfortunately, SPSS does not produce a pairwise comparison for a non-significant Kruskal-Wallis test.

A one-way between groups ANOVA analysis was conducted for a parametric comparison.[14] Given the non-normality of the data, the result would only give an indication despite

---

[10] Cohen J., *Statistical Power Analysis for the Behavioural Science*, 79–81.
[11] Field, *Discovering Statistics Using IBM SPSS Statistics*, 276.
[12] Appendix 9 A
[13] Kirk, "Practical Significance: A Concept Whose Time Has Come."
[14] Appendix 9 A

Levene's test of homogeneity of variance being non-significant (*p* = .189). There was a significant difference in SATs' effect on assessing uncertainty between the three units, *F*(2, 90) = 3.27, p = .043, $\omega^2 = .05/\eta^2 = .07$. Omega squared ($\omega^2$) is often preferred as a measurement of effect for between groups ANOVA as it is generally less biased than ETA squared (η^2).[15] For an unbalanced ANOVA as this one, other biases may apply, making appropriate to also report the ETA squared when the two give different results. For this analysis there was a small to medium effect using Kirk's interpretation (.01 = small effect, .06 = medium effect, .14 = large effect).[16] Post-hoc tests were carried out to investigate comparisons between the groups, where in addition to the Bonferroni's test applied to test for type I error, Hochberg's GT2 and the Games-Howell procedure were chosen due to unequal sample sizes.[17] All post-hoc tests revealed similar results and the Bonferroni test indicated that the mean score for the NIS group (*M* = 3.80, *SD* = .90) was significantly different from the IntBn group (*M* = 4.26, *SD* = .92), *p* = .038. The PST group (*M* = 4.02, *SD* = .68) did not differ significantly from the two other groups.

There were also significant differences between units regarding the use of standardised verbal probabilities and the degree of stating analytic confidence in intelligence estimates. A Kruskal-Wallis test showed that there was a significant difference in stating standardised verbal probabilities in intelligence estimates across the three different units, NIS (n = 51), IntBn (n = 28), and PST (n = 17), *H*(2) = 14.71, *p* = .001.[18] Dunn's pairwise comparisons with Bonferroni adjusted *p*-values showed that there was a significant difference with a medium effect between NIS (*Mdn* = 5.00/Mean rank = 41.44) and IntBn (*Mdn* = 5.00/Mean rank = 58.50), *z* = 3.68, *p* = .001, *r* = .41. There was no significant difference between the other pairs.

As for explicitly stating analytic confidence in intelligence estimates, a Kruskal-Wallis test showed that there was a significant difference across the three different units, NIS (n = 51),

---

[15] Albers and Lakens, "When Power Analyses Based on Pilot Data Are Biased: Inaccurate Effect Size Estimators and Follow-up Bias."
[16] Kirk, "Practical Significance: A Concept Whose Time Has Come."
[17] Field, *Discovering Statistics Using IBM SPSS Statistics*, 458–60.
[18] Appendix 9 B

IntBn (n = 28), and PST (n = 17), *H*(2) = 24.65, *p* < 001.[19] Dunn's pairwise comparisons with Bonferroni adjusted *p*-values showed that there was a significant difference with a large effect between NIS (*Mdn* = 2.00/Mean rank = 39.15) and PST (*Mdn* = 4.00/Mean rank = 76.68), *z* = -4.96, *p* < .001, *r* = .57. There was also a significant difference with a large effect between IntBn (*Mdn* = 2.00/Mean rank = 48.43) and PST, *z* = -3.40, *p* = .002, *r* = .51. There was no significant difference between NIS and IntBn.

SATs' effect on analytic rigour

Analysis of the component SATs' effect on analytic rigour revealed that 6.3 % of the survey 2 respondents reported that SATs were not relevant for analytic rigour. Investigating the component against different variables by running a series of non-parametric tests revealed only one significant result. While there were no significant differences for any of the demographic variables, a Mann-Whitney U test revealed that there was a significant difference in ranks with a small effect on SATs' effect on analytic rigour between those reporting the reason for not using SATs was that other methods are better (*n* = 22, *Mdn* = 4.75/*mean rank* = 36.57) and those not reporting this (*n* = 67, *Mdn* = 5.00/ *mean rank* = 47.77), *U* = 922.5, *z* = 2.04, *p* = .041, *r* = .22.[20] The underlying significant difference in ranks was for SATs improving analytic integrity across those reporting that other methods are better being the reason for not using SATs (*Mdn* = 5.00/mean rank = 36.77) and those not reporting this (*Mdn* = 5.00/mean rank = 47.70), *U* = 918.0, *z* = 2.13, *p* = .033, *r* = .23. There were no significant differences for any other variables, neither for SATs improving analytic integrity nor for SATs improving analytic objectivity.

Analysts' degree of SATs usage

The fact that the component SATs usage had a normal distribution, but a larger standard deviation than the two other components, could indicate that there were several other variables explaining the variance. A series of one-way between groups ANOVA analysis in SPSS was run on the component SATs usage to investigate this further, finding several significant differences across different demographic variables.

[19] Appendix 9 C
[20] Appendix 10

First, a one-way between groups ANOVA analysis found there was a statistically significant difference with a very large effect in SATs usage between personnel from the Intelligence Service (NIS) (*n* = 51), the Intelligence Battalion (IntBn) (*n* = 28), and the Police Security Service (PST) (*n* = 17), *F* (2, 96) = 38.66, $p < .001$, $\omega^2 = .44$.[21] The Hochberg's GT2 test was used for the post-hoc comparisons due to unequal sample sizes.[22] The test showed that the mean score (see table 4) for IntBn (*M* = 3.48, *SD* = .69) was significantly different from NIS (*M* = 2.08, *SD* = .73), $p < .001$ as well as from PST (*M* = 2.79, *SD* = .47), *p* = .004. There was also a significant difference between NIS and PST, *p* = .001. There was a nuance to this result, found by investigating service background regardless of unit. An independent samples t-test revealed that there was also a significant difference in scores of SATs usage with medium effect between military personnel (*M* = 2.82, *SD* = .96) and civilians (*M* = 2.30, *SD* = .75), *t*(94) = 2.81, *p* = .006, *d* = .69.

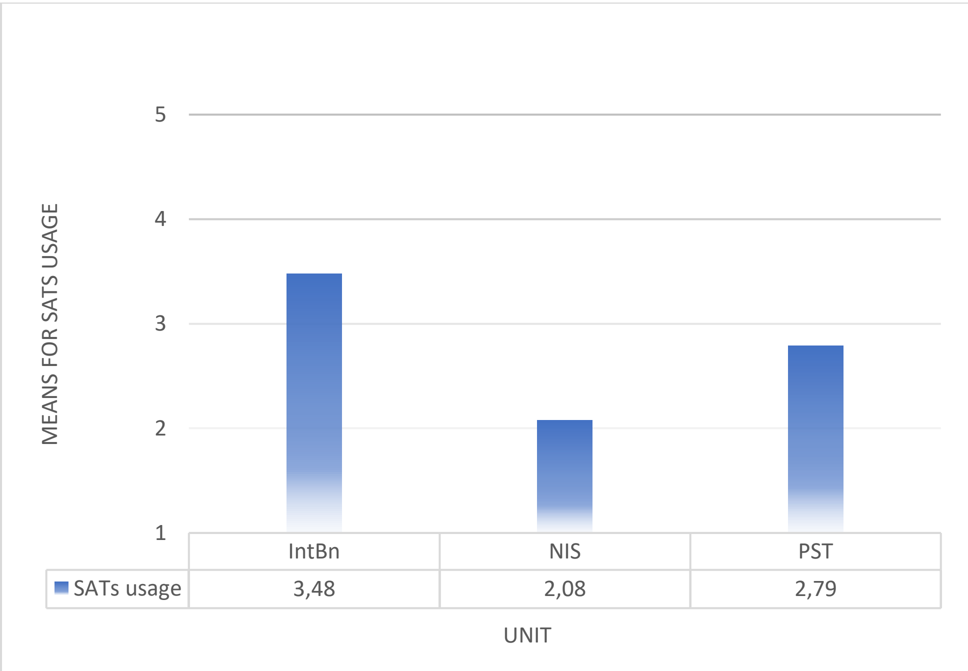


Table 4: Degree of SATs usage across units

Second, by investigating reasons for not using SATs, an independent samples t-test found that there was a significant difference in SATs usage scores, although with a small effect,

[21] Appendix 11 A
[22] Field, *Discovering Statistics Using IBM SPSS Statistics*, 459.

between those reporting there was no culture for using SATs (*M* = 2.30, *SD* = .77) and those not reporting that culture was an issue (*M* = 2.80, *SD* = .95), *t*(94) = -2.71, *p* = .008, *d* = .39.[23] A follow up chi-square test indicated a significant association between unit and culture being a reason for not using SATs, X^2 (2, *N* = 96) = 13.01, *p* = 001. The effect size using Cramer's V (ϕ_c) = .37 indicated a large effect.[24] While half of the NIS and PST respondents reported culture being an issue for not using SATs, the equivalent number for the IntBn was only 1 in 10. There was also an indication for a significant association with a medium effect between unit and SATs not being relevant, X^2 (2, *N* = 96) = 9.36, *p* = .009, ϕ_c = .31. Again, half of the NIS respondents reported SATs not being relevant where the equivalent numbers were 1 in 10 for the PST and as much as one-third for the IntBn. There were no associations between unit and other methods being better, between unit and SATs being time demanding, or unit and SATs being difficult.

Survey 1 versus survey 2

The effect of using SATs on analysts' ability to assess uncertainty

A comparison between survey 1 data and the data from survey 2 was conducted to investigate if there were any significant differences in opinion between the SATs-course participants and people working as intelligence analysts after having received some sort of SATs-training. Due to minor differences in the items making up the two components, survey 1 post-course SATs' effect on assessing uncertainty and survey 2 SATs' effect on assessing uncertainty, this comparison does not have complete validity, but does produce interesting indications. The means, medians and standard deviations are presented in table 5.

	N	**Mean (Median)**	**SD**
SATs' effect on assessing uncertainty (survey 1)	120	4.25 (4.29)	.56
SATs' effect on assessing uncertainty (survey 2)	90	3.98 (4.00)	.77

Table 5: Comparison of assessing uncertainty from surveys 1 and 2

[23] Appendix 11 B
[24] Pallant, *SPSS Survival Manual*, 222.

Although both the means and medians of the two components would translate to 'agree' on the surveys' Likert scales, a Mann-Whitney U test revealed that there was a significant difference, although with a small effect, in SATs' effect on assessing uncertainty between the two populations, *U* = 4073.0, *z* = -3.05, *p* = .002, *r* = .21.[25] Furthermore, there were between 7.3% and 13.5% 'Not relevant' answers on the items that made up the survey 2 component. The analysis nevertheless indicated a lasting opinion of SATs' impact on the ability to assess uncertainty as there was a significant difference with a large effect between the survey 1 pre-course ability and the survey 2 reported ability, *U* = 9005.0, *z* = 8.29, *p* < .001, *r* = .57.

Survey 2 also measured how the respondents separately perceived SATs' impact on ascribing probability and analytic confidence. Since the population of survey 2 was different than the population of survey 1, an independent samples Mann-Whitney U-test was run to investigate any differences between the two surveys.[26] The means, medians and standard deviations are presented in table 6.

	Survey	**N**	**Mean (Median)**	**SD**
SATs make it easier to ascribe probability	1	120	4.38 (4.50)	.77
	2	89	4.11 (4.00)	.98
SATs make it easier to ascribe analytic confidence	1	117	4.15 (4.00)	.80
	2	83	3.98 (4.00)	.94

Table 6: SATs' impact on ascribing probability and analytic confidence surveys 1 and 2

For the SATs' impact on ascribing probability there was a significant difference with a small effect between survey 1 and survey 2, *U* = 4533.5, *p* = .040, *r* = .14. For the SATs' impact on analytic confidence there was no significant difference.

SATs' effect on analytic rigour

It was possible to compare the SATs' effect on analytic rigour between respondents from survey 1 and respondents from survey 2 since the components consisted of equally worded

[25] Appendix 12
[26] Appendix 12

items, SATs improving analytic objectivity and SATs improving analytic integrity. The means, medians and standard deviations are presented in table 7.

	N	Mean (Median)	SD
SATs' effect on analytic rigour (survey 1)	118	4.48 (4.50)	.62
SATs' effect on analytic rigour (survey 2)	89	4.53 (5.00)	.78

Table 7: Survey 1 vs survey 2 SATs' effect on analytic rigour

A Mann-Whitney U test revealed that there was a significant difference with a small effect for the SATs' effect on analytic rigour between the two populations, $U = 6067.5$, $z = -2.06$, $p = .039$, $r = .14$.[27] However, a look at the small difference in means could indicate that this result was a type I error. An independent samples t-test showed that there was no significant difference between the two groups, $t(163.9) = -.49$, $p = .623$, $r = .04$, a negligible effect.

Discussion

Survey 2 set out to investigate the long-term effect the NORDIS SATs-course had on analytic rigour and on analysts' ability to assess uncertainty, both key aspects of the post-9/11 intelligence reform. In addition, survey 2 also investigated to what degree analysts used SATs for their assignments. Any inter-agency cultural differences were investigated, where analysts from an intelligence unit (PST) not having attended the NORDIS SATs-course were asked to participate for a comparison. The key results of the two surveys were also compared.

Just as for survey 1, the survey 2 results revealed there was a large significant correlation between the effect SATs have on analytic rigour and the effect on assessing uncertainty. There were also correlations with medium effect between the use of SATs and SATs effect on analytic rigour and ability to assess uncertainty. The survey 2 results had a larger effect size and was more conclusive than the survey 1 results, probably due to the survey 2 population being a more representative group of Norwegian intelligence analysts than the

[27] Appendix 13

survey 1 population. One can thus conclude that analysts that often use SATs experience high levels of analytic rigour and ability to assess uncertainty. Simultaneously, there was a large amount of variability that was accounted for by other variables.

The opinion of the effect using SATs have on analytic rigour was high for survey 2 and did not significantly differ between the two surveys. Although it was not possible to investigate an exact effect of receiving SATs-training, the mean opinion of survey 2 was on the same high level as for survey 1, indicating a lasting effect even years after having taken part in SATs-training. The one exception for survey 2 was a significant difference on SATs' impact on analytic integrity between those reporting that other methods are better than SATs and those that did not report this. Hence, while SATs are effectful for objectivity, this indicates that perception of integrity depends on familiarity of methods used or that some intelligence problems are solved by techniques people do not regard as SATs. Despite this exception, it is safe to conclude that SATs have more than face value when it comes to fulfil one of the main requirements of the post-9/11 intelligence reform.

For the uncertainty dimension, the overall result was that analysts experience that using SATs improves their ability to assess uncertainty. Based upon a comparison with the survey 1 results, there was a small decrease from the immediate post-course opinions on SATs' effect on assessing uncertainty, due to a decrease in the opinion of the SATs' effect on ascribing probability. While SATs' effect on analysts' ability to assess analytic confidence was the same as the post-SATs-course effect, more analytic experience seemingly gives analysts a further reality check on SATs' impact on ascribing probability. A natural explanation is that none of the SATs taught today explicitly deal with how to assign probabilities to possible outcomes in a rule dictated and objective manner. Instead, probability is still an intuitive, subjective judgement of the analysts involved.[28] Outside the safe boundaries of a simulation exercise, other factors than SATs alone appear to have an impact.

There was, however, a small significant difference between personnel from the Norwegian Intelligence Service (NIS) and personnel from the Norwegian Intelligence Battalion (IntBn),

[28] See chapter 3, especially Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases."

where personnel from the NIS were alone in reporting that SATs were not relevant for assessing uncertainty. The opinion of personnel from the Police Security Service (PST) lay in between the personnel from the two other units and did not significantly differ from any of them. There was also a significant difference in use of standard verbal probabilities and explicitly stating analytic confidence in intelligence estimates. The Norwegian Intelligence Service (NIS) stands out as the unit with the least use of both standard phrases of verbal probabilities and expressions of analytic confidence in intelligence estimates. For the Intelligence Battalion (IntBn) the use of verbal probabilities was high and the use of expressions for analytic confidence low and vice versa for the Police Security Service (PST).

Consequently, just as for survey 1, survey 2 revealed that analysts are of the opinion that there is more to assessing uncertainty than just a SATs-approach to the estimative process. These results furthermore indicate that there exists a significant difference in analytic culture between the units where especially personnel from the NIS are of a lower opinion of SATs' effect and relevance for uncertainty handling compared to personnel from the IntBn and the PST. A reasonable explanation for this is that the use or non-use of both verbal probabilities and expressions of analytic confidence follow the culture among the units' customers. While the PST was tasked by the Justice Department to revise its use of verbal probabilities and with that included a standard for statements of analytic confidence in its revised intelligence doctrine, there has been no such push for either in the NIS or the IntBn. For the IntBn, however, the use of a standardized set of probability phrases is far more mandatory as it is explicitly described in military doctrines, making it an analytic standard to abide by.

Investigating to what degree analysts use SATs, a key finding was a very large significant difference in SATs usage between the three different units (NIS, IntBn, PST), where SATs were used the most in the IntBn and the least in the NIS. Military personnel also used SATs significantly more than civilians, regardless of unit. Agency culture was reported to be the main reason for not using SATs, followed by SATs not being relevant, where for both reasons there were significant differences between the three units. A reasonable explanation to this would be that while personnel from the NIS and the PST in general only have a two-week course in the use of SATs, personnel from the IntBn undergo a 6-months long education program in intelligence analysis prior to being certified as intelligence analysts. The finding furthermore supports one of the results from Coulhart's research, revealing a difference in analytic culture towards using SATs between civilian subject-matter experts and military generalists.[29]

The main output of the survey 2 data was that there is almost a unison of opinion among Norwegian intelligence analysts that the use of SATs improves their ability to assess uncertainty and increases analytic rigour in the analytic process, especially by increasing analytic objectivity. The effect seemingly also lasts for years after having received SATs-training, that being the NORDIS course or other training. Any demographic differences were of a small effect only. At the same time, the actual use of SATs is in general at a much lower level. One could therefore think that the average analyst would suffer from cognitive dissonance, not using SATs even though one is of the opinion that their use would enhance the quality of the analytical process. However, in addition to SATs training, a prerequisite for SATs usage is an agency culture endorsing the use. Another key finding from survey 2 is that there exist differences in analytic cultural between different Norwegian intelligence agencies. The differences in agency culture have a large impact not only on the use of SATs but also on how uncertainty is expressed in intelligence assessments. Qualitative data analysis based on interviews conducted among some experienced intelligence analysts were analysed to gain further insight to the matter of the impact of analytic culture on the degree of use and the effect of using SATs in the different agencies.

[29] Coulthart, "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency," 941–42.

Chapter 9: Interviews and survey comments

Participants

To shed more light on the key results from the two surveys, a qualitative analysis of in-depth interviews with some selected experienced analysts in the Norwegian intelligence community was conducted. Moreover, free-text comments from survey 2 were analysed. Seven Norwegian analysts were interviewed, and they made up a more homogeneous group than the survey populations. Due to the nature and sensitivity of their work, a full disclosure of their background is not possible. All were male aged from 20-50 and with a range of 2 to 25 years of service. 5 out of 7 had a military rank and all but one had a postgraduate degree. 40 respondents in survey 2 left comments of value as part of answering the questionnaire. A little more than one-third of these were female, with age groups ranging from 26-30 to 46-50. They were equally divided between civilians and military ranks as well as equally divided between the different categories of years of intelligence experience. Moreover, two-thirds had a postgraduate degree. Half of those leaving comments had also attended a SATs-course prior to the start-up of the data collection for survey 1.

Interview analysis and discussion

The interviews with the analysts were thematised into the same main categories as the components resulting from the Exploratory Factor Analysis of the survey 2 data: degree of SATs usage, SATs' experienced effect on assessing uncertainty and analysts' experience of SATs' effect on analytic rigour. All interviewees used SATs regularly, although the types of techniques varied depending on the type of analytical tasks and intelligence problems to answer. In comparison, 14 of the survey respondents leaving comments stated that they never or seldom used SATs, either due to not having attended any SATs-training, a lack of a culture for using SATs, or time constraints. All 14 were from the NIS, 10 of them civilians.

The key finding from the qualitative data is that there is a unison agreement that SATs increase analytic rigour and make it easier to assess uncertainty. More specifically, using SATs increases analysts' confidence in their judgements, which makes it easier to assess uncertainty, resulting in a perception of increased analytic integrity. The respondents were also of the opinion that using SATs help them be more aware of heuristics and biases that can impede the analytical process, resulting in increased analytic objectivity by eliminating satisficing and single-outcome analysis. The techniques that were presented as the most valuable were driving force analysis with a subsequent causal-flow diagram, Scenario Generation, Key Assumptions Check, and Analysis of Competing Hypotheses (ACH). When asked what alternative methods or techniques they would make use of instead of SATs, the general comment was that they would fall back on whatever methods they had been taught at university.

The combination of scenario generation, Key Assumptions Check, and ACH was highlighted as distinctly valuable. Instead of falling into the trap of single-outcome forecasts, by using these techniques analysts must consider alternatives, identify and check assumptions that otherwise would have been overlooked, and objectively evaluate both the available information and to what degree new information would alter a current conclusion. The increased objectivity and integrity of this process first and foremost results in a feeling of increased analytic confidence. The interviews thus revealed that the increased analytic confidence of analysts could be understood more as reduced analytic uncertainty, that using SATs make them more certain which key judgements to report and the evidentiary basis for those judgements. There are three key take-aways to this finding. First, using SATs is first and upmost a critical thinking approach mixing elements from both art and science, aimed at widening analysts' horizon to prevent surprises and discourse failure. Debiasing and judgement accuracy may come as second-order effect, if at all. Second, the concept of analytic confidence is still in its infancy and needs further clarification and research. Third, there is no established reliable or valid method for probability calculus in the current SATs, making probability statements subject to, in Teigen's words, 'whatever judgemental heuristic that seems to be most appropriate for evaluating the phenomenon under consideration.'[1]

[1] Teigen, "Variants of Subjective Probabilities: Concepts, Norms, and Biases.," 212.

Despite that using SATs was expressed to have a positive impact on the quality and rigour of the estimative intelligence process, there are different factors impeding the actual use. There was a unison opinion that the use of SATs depends on a combination of training and experience with SATs and an organisational culture that commends the use of SATs. The most important factor reflecting on the use of SATs is agency culture. A general comment given by the analysts working in NIS was that there was no culture for the application of SATs and that the management did not provide the time needed to work more structured. An opposing view from respondents working in the IntBn was that the use of SATs is imperative to ensure analytic rigour as well as a common, shared understanding of the situation and how it may develop. The experience in the IntBn was that using SATs gave the analysists an opportunity to work on many topics across several focus areas without necessarily having to be a subject matter expert. A senior analyst put it this way: 'Using SATs is the reason we are punching way above our weight.'[2] For analysts in the PST, the focus on using SATs was based upon designing a structured approach of data organization and scenario development, where the subsequent use of ACH was tailored to the unit's process of ranking threat actors to prioritize collection.[3] Hence, there seems to exist, using Tetlock's terminology, a foxes versus hedgehogs culture between the IntBn and the NIS which is not beneficiary for the latter. As for the PST, the use of SATs seems to make subject matter experts, or hedgehogs, better at their specific working process, where process efficiency seemingly has outweighed the goals of improved analytic rigour and uncertainty handling.

A conclusion is that the use of SATs needs to be endorsed by the senior management to be part of a unit's analytical culture. Several interviewees commented that the use of SATs among analysts increased immediately after returning from SATs training only to decrease as management only showed interest in evaluation of finished estimates, not any interest in overseeing a critical evaluation of the analytic process leading to the estimates' key judgements. One interviewee stated: 'Analysts using SATs produced more precise and logically coherent products, but there was no management interest in quality assurance of the process leading to these products.'[4] Analysts who have working experience in an

[2] Interview C
[3] Interview G
[4] Interview G

organisation with a strong culture for using SATs are nevertheless more likely to use SATs when working in organisations where the use of SATs is not necessarily promoted by senior management. One interviewee stated: 'Indoctrination works, those who come from a culture where they are used to apply SATs in everyday use have this in their analytical DNA, the others fall down to their level of training.'[5] A strong endorsement of using SATs from the mid- and top-level management is therefore needed for moving beyond SATs being the analytic methodology of the mass instead of the few devoted ones.

A second factor is that the use of SATs is very context dependent. What analysts associate with using SATs largely depends on the type of intelligence problems one is tasked to handle, especially in a large unit like the NIS. Analysts that mainly work with secrets, such as surveillance and monitoring of enemy units and equipment, generally do not seem to connect their form of data and information structuring, such as using matrices, timelines, network diagrams, etc, as a specific use of SATs. On the contrary, techniques analysts think of as SATs are techniques like Key Assumptions Check, Scenario Generation, and ACH, techniques that to a larger degree are used in small teams primarily working with mysteries or complexities. This is especially evident in situations in which developments emerge that impact key driving forces, which again necessitate analysts to rethink possible future outcomes. When major developments like the fall of ISIS or the killing of the Iranian General Soleimani in Iraq occur, there is a growing understanding that the NORDIS SATs-methodology is useful to apply to ensure an objective analysis and not least that all assumptions are explicitly dealt with.

Moreover, the interviews revealed that the use of the NORDIS SATs-methodology and adapted versions of it, is considered somewhat more relevant in support of military operations, especially in the planning phase, but also during a tour, as these situations are perceived to have a clear start and end for the analysts involved. A reason for this may be that several SATs bear resemblance to techniques presented in military doctrines, where SWOT/TOWS can be used in to evaluate the operational environment and event matrix is

[5] Interview F

comparable to ACH.[6] Simultaneously, many analysts have a harder time applying SATs when monitoring a slow-moving target like, the Russian Armed Forces in the High North, every day, year after year, since there may not be significant changes in the patterns of activity. A possible reason can be shortcomings in the indication and warning system, where extrapolation of current trends instead of monitoring and testing alternative scenarios is seemingly the basis for the strategic estimative process.[7]

A third factor is that the NORDIS SATs-methodology can appear as too elaborate and "one size fits nobody". All interviewees reported that there is a need to adapt the methodology to the situation at hand, especially due to time constraints. Both the PST and the IntBn have designed their own structured methodological approach to better serve their typical tasks. The NIS, on the other hand, has yet to find a most effective approach. On several occasions, teams of NIS analysts have tried to apply the complete NORDIS SATs-methodology in one-day workshops, only to realise that too much time is needed at the start of the process just to get all participants on the same sheet of paper. For some teams, this has justified going back to the old ways, primarily working by oneself with whatever methodology, or lack thereof, one personally prefers. Other teams, however, have had key team members that instead have adapted the NORDIS SATs-methodology to make it work for their team and their specific tasks. The success of these teams depends on having one or several SATs - experts to guide the team in the process, in addition to a senior manager that supports the process and ensures that all necessary subject matter experts (SMEs) are available when needed. The key here is that there will always be a need for in-dept knowledge of different subjects which SATs alone cannot replace. One interviewee stated: 'SATs do not replace SMEs but enhance their contribution; a common methodology is what binds the SMEs together and ensures high quality products.'[8] Just as academia alone could not ensure high

---

[6] Joint Chiefs of Staff, "Joint Publication 2-01.3: Joint Intelligence Preparation of the Operational Environment," II–81.

[7] The Norwegian Intelligence Service, "Focus 2016: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2017: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2018: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2019: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2020: The Norwegian Intelligence Service's Assessment of Current Security Challenges."

[8] Interview F

quality products for the Inquiry in 1919, SATs cannot improve the analytic process without having the insights of true experts as part of the process. Babetski conveyed the idea this way:

> The trick is in using structured techniques and approaches – or applied system 2 thinking - in a way that eliminates biased intuitive forecasts and predictions without also discouraging, delaying or even eliminating the intuitive insights that true expertise provides.[9]

An important element to keep in mind is that people working alone or in homogeneous teams are more prone to suffer from heuristics and biases, while working in small heterogeneous teams, utilising a structured methodology, improves the ability to keep cognitive limitations in check.[10] There are as such methodological but also organisational aspects that need to be addressed for an effective use of SATs.

The fourth factor is that intelligence estimates do not necessarily present the methodology behind the assessments, impacting both the use of SATs and to what degree senior management understands that rigorous high-quality products stem from a structured methodology. Those receiving a finished estimate from the NIS are usually not presented with alternative outcomes with a relative probability attached, but rather only the most likely conclusion and not always with a probability statement.[11] The recipients have therefore little or no insight in the process behind the product. Analysts in military units, like the IntBn, are on the other hand used to presenting two or more alternatives in most intelligence estimates, such as intelligence support to planning or intelligence summaries, as this is dictated in NATO and national military doctrines. Other members of a military operational staff can usually also participate in parts of the process, thereby witnessing the methodology in use. For the PST, use of SATs was explicitly mentioned in the 2020 threat

[9] Babetski, "Intelligence in Public Literature: Thinking, Fast and Slow."
[10] Heuer, "Small Group Processes for Intelligence Analysis"; Convertino et al., "The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis."
[11] The Norwegian Intelligence Service, "Focus 2016: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2017: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2018: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2019: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; The Norwegian Intelligence Service, "Focus 2020: The Norwegian Intelligence Service's Assessment of Current Security Challenges."

assessment, although without presenting any alternative scenarios.[12] A conclusion is that the use of SATs seems to increase the more the different techniques enhance and complement an already established overall analytical process.

The lack of transparency in the products can furthermore impact how uncertainty is expressed in intelligence assessments. The interview data indicated large differences between the three units. First, reporting both probability and confidence has started to become the norm in the PST.[13] Second, the IntBn only reports probability after having received feedback from commanders that expressions of analytic confidence in assessments only confused them and increased their uncertainty.[14] Third, in the NIS, there is no use of expressions of analytic confidence and the use of standard probabilistic phrases is decreasing.[15] A reason can be, as one NIS analyst stated: 'The managers are uncomfortable expressing uncertainty in the intelligence assessments.'[16] From the interviews it can seem that SATs mostly affect uncertainty when it comes to assessing analytic confidence and to a lesser degree when assessing probability.

The key take-away from the interviews is that using SATs in the estimative intelligence process increase analytic rigour and thereby reduce analytic uncertainty. At the same time, analysts' actual use of SATs, and thus also gaining the wanted effects from using them, depends on having a common framework for using the different techniques that suits the tasks given, and an agency culture advocating the use of SATs. An appropriate use of SATs is consequently a management issue, something not appraised by all Norwegian intelligence agencies. Only by recurrent use of SATs as part of a structured estimative process will the techniques become a natural part of an analyst's job. The interview data clearly point towards the benefits outweighing the costs.

[12] Politiets sikkerhetstjeneste (PST), "National Threat Assessment 2020."
[13] Interview G
[14] Interview A
[15] Interview E
[16] Interview F

Chapter 10: Overall discussion of the effects using a comprehensive Structured Analytic Techniques methodology has on analytic rigour and ability to assess uncertainty

The intelligence reform instigated by the Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004 led to a renewed focus on improving intelligence analysis tradecraft to prevent discourse failure. Structured Analytical Techniques (SATs) became the new tradecraft standard, aimed at increasing analytic rigour and improving the communication of uncertainty in estimative intelligence products. Along with the focus on SATs came a call for validating these so-called “social science for dummies” techniques. The problem is that only ACH has been repeatedly tested, and that the findings have been generalized to SATs as a whole. The testing of ACH has furthermore been underwhelming for several reasons.[1] One is that most of the previous research has had a limited focus, investigating the effects on bias mitigation and judgement accuracy, two aspects only auxiliary to the main goals of the intelligence reform. The main pitfall is nevertheless that ACH has not been tested in environments or situations similar to how ACH is used in real-life.

This research, on the other hand, set out to investigate the experienced effects using a comprehensive, layered, and iterative Structured Analytical Techniques (SATs) methodology have on analytic rigour and the ability to assess uncertainty, key elements of the intelligence reform. The NORDIS SATs-course teaches the use of several different SATs in a situation simulating a real-life intelligence process. Both the immediate effects of having conducted a two-week SATs-course as well as the more long-term effects were investigated. There are three main findings from this research. First, the NORDIS SATs-course had a large positive effect on analysts’ ability to assess uncertainty. Moreover, there is only a small difference between the two surveys, indicating a lasting effect of SATs-training. Second, the participants experienced that the SATs-course increased analytic rigour, although more so for inexperienced analyst than for the experienced ones. Again, there were indications of a lasting effect as there were no significant difference between the two surveys. Third, agency culture impact both the usage of SATs and how uncertainty is expressed in estimative

[1] See Chapter 1 for discussion

intelligence products. The overall main effect of using SATs in estimative intelligence can be said to be reduced analytic uncertainty through increased analytic rigour. The wanted effect towards discourse failure may nonetheless be absent if the intelligence products do not reflect all aspects of using a structured methodology.

The first main finding was that participants on the NORDIS SATs-course experienced a large positive effect on their ability to assess uncertainty. The course participants had a disproportionate pre-course belief in the effect of using the SATs-methodology, resulting in a small decrease for the post-course experienced effect. Analysts seemingly undergo a further reality check over time, which resulted in a small negative difference between the two surveys. The survey 2 results were nonetheless significantly higher than the reported pre-course ability, indicating a lasting effect of receiving SATs-training. The qualitative data supported the survey results, although arguably more with regards to increased confidence in key judgments than to increased ability to state probability.

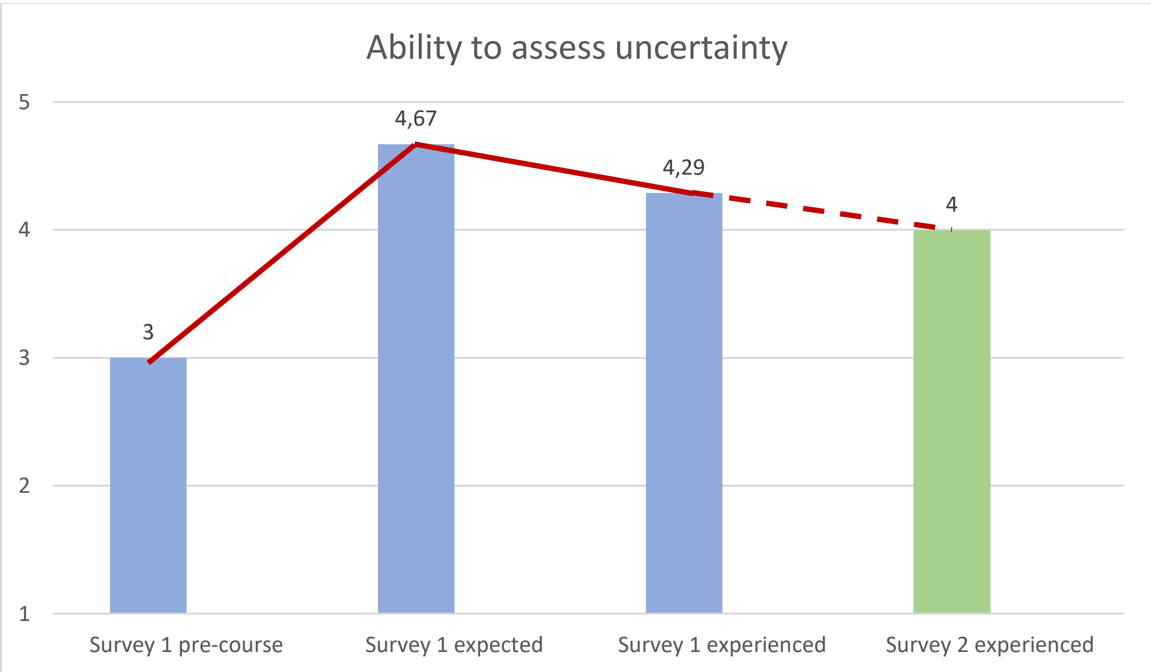


Figure 11: Analysts' opinion of ability to assess uncertainty before and after SATs training. 5= agree, 1 = disagree

A reasonable explanation for this reality check is that there is more to assessing uncertainty than structuring the analysis. Many other factors may impact, such as the quality and amount of available information, the number of key assumptions, or difficulties differentiating between hypotheses. While these factors can be easier to evaluate with a structured approach to the analysis since the process forces one to explicitly lay out the evidentiary basis for one's judgements, the nature of mysteries and complexities makes it impossible to account for all uncertainties. Moreover, uncertainty handling in intelligence analysis is still resting on subjective judgments, not objective and unbiased calculations as in several, especially natural, sciences. Since people are prone to leave the rules of probability calculus, they become susceptible to judgemental heuristics instead. A key reason is, as argued by others, limitations of current techniques as to reliable uncertainty handling.[2] ACH, for instance, does not in its original form present any objective way of assessing probability or analytic confidence. The original technique will give an indication of the relative likelihood for the hypotheses, but this is purely a subjective approach and far from a scientific probabilistic measurement. The result is often, as addressed by Chang et el., that subjective judgments become falsely dressed in a 'cloak of objectivity.'[3] Consequently, the intelligence community need to continue to research how to adopt viable concepts from other sciences to improve information evaluation and uncertainty handling.

The second main finding was that the SATs-course participants experienced a high effect on analytic rigour, although there was a nonconclusive result between novice and more experienced analysts. While survey 1 participants with some experience reported a lower effect than the novice, there was no such difference among the participants in the follow-up survey, indicating that SATs-course participants with some experience had an inflated confidence in own competence. Furthermore, there was in both surveys a significant correlation between analytic rigour and uncertainty where high levels of SATs' effect on analytic rigour were associated with high levels of SATs' impact on ability to assess uncertainty. A comparison of the data from surveys 1 and 2 also revealed that SATs' effect on analytic rigour endured over time as there was no significant difference in opinion between the two populations. The survey findings were supported by qualitative data, underlining that training and repeated use of SATs have a lasting effect due to SATs becoming a part of the analytical DNA.

[2] Chang and Tetlock, "Rethinking the Training of Intelligence Analysts." See also chapter 1
[3] Chang et al., "Restructuring Structured Analytic Techniques in Intelligence," 344.

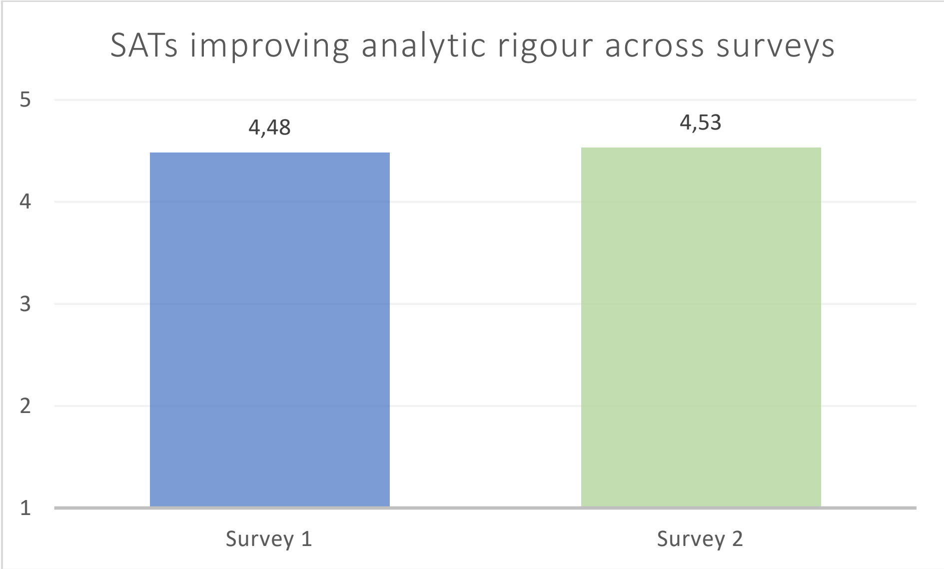


Figure 12: Analysts' opinion on SATs' effect on analytic rigour. 5 = agree, 1 = disagree

The finding of this research ties well in with the finding from Coulthart's study, that SATs were perceived as 'fairly effective' in improving analytic rigour.[4] However, survey 2 revealed that a significant number of analysts were of the opinion that methods other than SATs were more important for analytic integrity. An explanation can be that SATs are by many treated in a uniform manner instead of being tailored to the problem at hand. Interview data supported this, that the NORDIS SATs-model may function as "support wheels" for novices and generalists, improving their ability to consider all aspects and information belonging to an intelligence problem. The more experience one gains, the more one understands that there is a need to adapt and design one's own methodology to achieve the wanted effect. Simultaneously, there were no significant differences for any variables for SATs' effect on objectivity, indicating that using SATs widen analysts' perspectives, a key element to prevent single-outcome forecasts and discourse failure. In sum, lending Tetlock's argument, it is not about <u>what</u> you think but <u>how</u> you think, making cognitive style a key criterion also for increased analytic rigour.[5] Consequently, the finding on SATs' effect on analytic rigour gives prominence to the fact that SATs alone cannot replace experts and knowledge, but rather enhance them, thereby supporting the argument that intelligence analysis do not belong

---

[4] Coulthart, "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency," 942.
[5] Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*, chap. 3.

into a fixed category of art or science, but is rather a combination of key elements from both.

The third main finding is that there exists a distinct difference in analytic culture between the main intelligence agencies in Norway. This cultural difference is manifested not only in the degree of SATs usage, but also to how uncertainties are expressed in estimative products. SATs were used the most in the Army's Intelligence Battalion (IntBn) and the least in the Norwegian Intelligence Service (NIS), where agency culture and management indifference was reported to be the main reason for not using SATs. There were also differences between IntBn, NIS and the Police Security Service (PST) when assessing and expressing probability and analytic confidence in their products. Again, The Norwegian Intelligence Service stands out as the unit with the least use of both standard probabilistic phrases, or words of estimative probability (WEP), and levels of analytic confidence, where management discomfort for usage was pointed to as the main explanation.

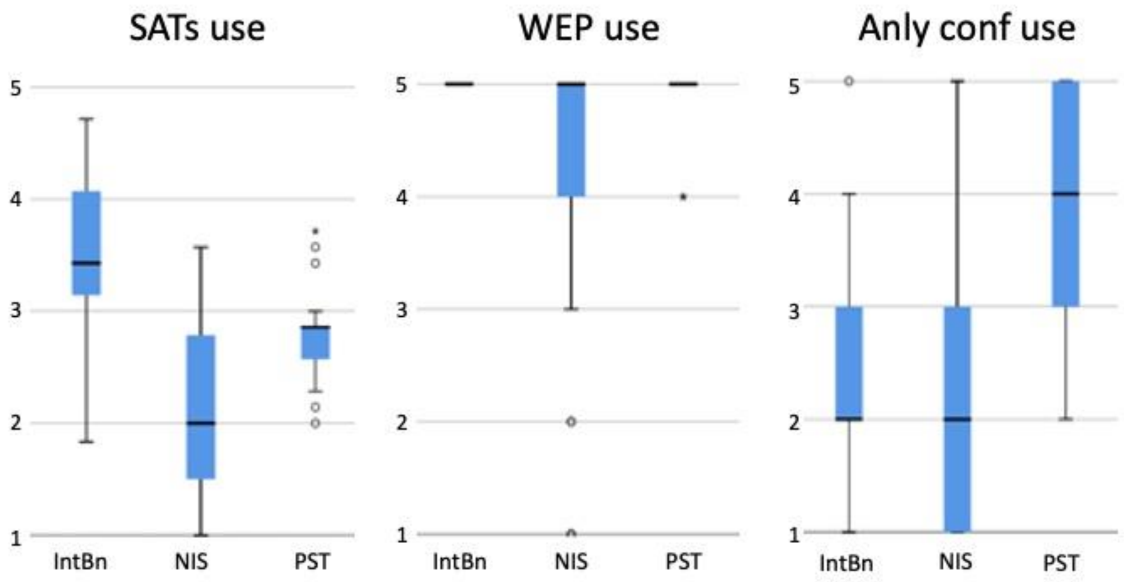


Figure 13: Use of SATs, WEP, and analytic confidence across units. 5 = always, 1= never

The most obvious explanation to the finding is that not only SATs-education but also extensive training is a prerequisite for use of SATs, standard probabilistic phrases, and levels of analytic confidence. Those who have the longest training also report the most use. Another explanation may be that the products are tailored to the different customers where many recipients may find both standard probability phrases and expressions of analytic confident alien and confusing. A third reason may be that there seems to be a continued civilian tendency of presenting single-outcome analysis in estimative products, where the use of SATs may not seem necessary.[6] The core idea of utilising a structured intelligence analysis methodology that endorses both creativity and critical thinking is not to present only the most likely outcome, but rather to also present the other key alternatives, with their respective assumptions, uncertainties, and likelihoods, relative to the most likely one. This form of formalized multiple estimates is in consonance with Abbott Smith's description of what constitutes a good NIE, made years ago:

> Sophisticated estimating ought always to be something more than bald prediction. The course of events is seldom inevitable or foreordained, even though hindsight often makes it look that way. A good paper on a complicated subject should describe the trends and forces at work, identify the contingent factors or variables which might affect developments, and present a few alternative possibilities for the future, usually with some judgment as to the relative likelihood of one or another outcome.[7]

The findings indicate it can be difficult to achieve intercommunity relevance for SATs due to two main reasons. First, agency culture, and especially leadership endorsement, is a major factor not only for SATs usage but for uncertainty handling as well, where management must care about and set standards for both process and product rigour. Second, SATs should not be seen as a universal methodology but rather a framework to increase objectivity and integrity together with creative and critical thinking in a rigorous and traceable manner, where different problems need different techniques.

The findings in this research are based upon data collected from analysts in a small NATO nation are consequently not universal in nature. Similar research on structured approaches in other intelligence communities is therefore most welcomed. Since SATs are not taught and used in the in same manner across the different intelligence communities, new research may give different results regarding the effects of using SATs. A starting point would therefore be to conduct similar research in relation to SATs-courses in other nations, such as

[6] As examples, see: The Norwegian Intelligence Service, "Focus 2022: The Norwegian Intelligence Service's Assessment of Current Security Challenges"; Politiets sikkerhetstjeneste (PST), "National Threat Assessment 2021."
[7] Smith, "On the Accuracy of National Intelligence Estimates."

the United Kingdom, the Netherlands, and the United States, preferably with a similar follow-up study after having used SATs as intelligence analysts for a certain amount of time.

There are also other issues left unanswered and for which future research may bring new insight. A first issue is, going back to Tetlock, to investigate the effect of more structured techniques, especially forecasting techniques. A particularly interesting angle would be to look at alternatives to the Global Business Network (GBN) Scenario Cross and the Cone of Plausibility techniques. Morphological Analysis is a more complex methodology, but also more suitable for identifying and structuring all possible aspects and solutions for non-reducible, complex problem spaces. The driving forces are broken down to a comprehensive set of mutually exclusive states, and when combined, they constitute a multidimensional matrix containing all theoretical scenarios.[8] Thereafter a cross consistency assessment is run to delete irrelevant and inconsistent combinations, reducing the solution space by 90% or more. The main issue with morphological analysis is that, being rooted in the *La Prospective* school of scenario techniques, it often relies on quantitative and probabilistic data and models where the need for computer support becomes intangible due to a multitude of parameters.[9] The largest advantage is that morphological analysis is extremely scalable and can just as well be used with qualitative data and information and is therefore just as functional when analysing and modelling mysteries, like an actor's alternative courses of action, as it is for analysing and creating alternative models of complexities.[10] Investigating how Morphological Analysis or other scenario techniques compare to the GBN technique and Cone of Plausibility as to practicability, comprehensiveness, and the ability to create mutually exclusive and collectively exhausting alternatives would be very valuable for the intelligence community.

A constructive next step would be to research the effect of combining different scenario techniques with hypotheses testing techniques, such as the research done by Granåsen and Karasalo, combining Morphological Analysis with ACH.[11] There are shortcomings to how

[8] Johansen, "Scenario Modelling with Morphological Analysis."
[9] Bradfield et al., "The Origins and Evolution of Scenario Techniques in Long Range Business Planning," 807–8.
[10] For some examples, see Álvarez and Ritchey, "Applications of General Morphological Analysis: From Engineering Design to Policy Analysis."
[11] Granåsen and Karasalo, "Methodology and Tool to Facilitate Structured Analysis of Multiple Hypotheses."

much of the current literature explain the use of ACH, especially the problem of giving evidence proper diagnostic weight. Today the indicators are assumed to have equal relevance towards all hypotheses, something which is rarely the case. A mitigating solution would be to improve evidence diagnosticity. An example of such an approach is to implement the use of subjective logic to improve the ACH technique (ACH-SL).[12] By implementing ideas taken from Bayesian probability, fuzzy logic and formalized reasoning, the analyst can assign both probability and analytic confidence to each indicator, allowing for an improved calculation of the diagnosticity and overall conclusion. A different methodology is TRACE, an overall methodology that can in general be compared to the NORDIS SATs-methodology.[13] Future research should therefore also investigate the utility of other hypotheses testing techniques and methodologies in addition to adapted forms of ACH.

The second issue that has not been investigated in detail in this research is a more detailed uncertainty handling in intelligence analysis. Probability and analytic confidence are not limited to SATs alone but central aspects of all judgements and assessments. The intelligence community will benefit from an improved understanding of uncertainty handling, for which the breadth of probability theory and communication of uncertainty warrant a research project of its own. Although this thesis has found that analysts found SATs to improve their ability to assess uncertainty, there are today no techniques in the commonly referred SATs-toolbox that properly address the issue of probability.[14] Most standard probability theory methods are of little utility for estimative intelligence, except perhaps for Bayesian methods. Recurring attempts to incorporate Bayesian probability in intelligence analysis has so far been of no or limited success and standardized uncertainty scales are still largely resting on subjective inputs, with all that entails. Since estimative intelligence mainly deals with epistemic uncertainty, several authors point towards looking to other sciences, especially climate change, decision science, and risk management.[15] This research is ongoing, where

[12] Pope and Jøsang, "Analysis of Competing Hypotheses Using Subjective Logic."
[13] Stromer-Galley et al., "Flexible versus Structured Support for Reasoning: Enhancing Analytical Reasoning through a Flexible Analytic Technique."
[14] Heuer and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 2020; Defence Intelligence, "Quick Wins for Busy Analysts."
[15] Budescu et al., "The Interpretation of IPCC Probabilistic Statements around the World"; Walker, Haasnoot, and Kwakkel, "Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty";

different alternatives to ACH have been promoted, such as a factorized Bayes's theorem (FBT) elicitation method, or techniques related to Dynamic Adaptive Planning (DAP) and Robust Decision Making (RDM).[16] The beforementioned ACH-SL is another alternative technique, but which needs a proper software to be of utility for the intelligence community. One of the most recent results is that instead of using single techniques, a combination of methods increased judgements accuracy.[17] The latter can be of interest for future research, especially since this research found that a combined use of SATs gave analysts a perception of improved ability to assess uncertainty. Future research should therefore look further into how SATs can be improved to better handle uncertainty in complex intelligence problems. At the same time, the accuracy debate in intelligence analysis may be another distorted debate masked by a pretence of helping to improve decision making. A reason is that there has been very little research done into what kind of intelligence estimates policy- and decisionmakers prefer to be presented with.

The third issue not addressed in depth in this dissertation runs parallel with the art-science and accuracy debates, namely whether intelligence should be focusing on current or long-term analysis.[18] Intelligence is more than knowledge, it is also about understanding, and one should perhaps say that the aim of intelligence is to elucidate; to bring insight and foresight into a policymaking process. The current vs long-term debate includes thus an element of preventing discourse failure. To be able to elucidate, an intelligence product must often function as a wake-up call in a timely fashion to attract the right responsiveness from policymakers. How to gain responsiveness lies in the centre of the debate regarding where to put the emphasis on intelligence analysis, on the long-term or the short-term. Robert Gates expressed already 20 years ago: 'One of the CIA's greatest concerns over the years has been the unwillingness or inability of most policymakers to spend much time on longer-range issues - looking ahead several steps - or in helping to guide or direct the agency´s long-

---

Dhami et al., "Improving Intelligence Analysis With Decision Science"; Karvetski, Mandel, and Irwin, "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation."

[16] Karvetski, Mandel, and Irwin, "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation"; M. Isaksen and McNaught, "Uncertainty Handling in Estimative Intelligence – Challenges and Requirements from Both Analyst and Consumer Perspectives."

[17] Karvetski, Mandel, and Irwin, "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation."

[18] Dahl, *Intelligence and Surprise Attack*; Dahl, "Missing the Wake-up Call: Why Intelligence Failures Rarely Inspire Improved Performance."

term efforts.'[19] Omand warns against using all resources on situation updates, taking no time to see the long lines.[20] He goes on to argue that an important prerequisite for being able to identify fractures in patterns is that you have time to lift your mind from the current details. Former CIA deputy director of intelligence Bruce Clark described it as 'Intelligence research is putting money in the bank; current intelligence is making a withdrawal'[21] A sole emphasis on current intelligence means, as Robert Clark pointed out, that it is possible to go bankrupt.[22] Consequently, one cannot do one without the other. As for SATs, some techniques are most suitable for current intelligence and other techniques are most suitable for long term analysis. Almost all contemporary writing on SATs disregards this spectrum and treats all SATs the in the same manner as alternative analysis. Future research could shed more light on this division.

There is consequently a need to further develop the taxonomy of SATs by increasing the range of SATs. Input from several different fields, for example statistics, data mining, and artificial intelligence can improve the identification and communication of trends and patterns. It will also be beneficiary to look to future studies to expand the field of scenario developments with more techniques. Moreover, structured methodologies like the NORDIS SATs-methodology, the Dutch approach to SATs, or the model in "Quick Wins for Busy Analyst" should be approached as a framework for guidance, not a cookbook or checklist. Lastly, advances are needed to improve uncertainty handling. It is consequently necessary to continuously update analytic standards to further improve analytic tradecraft, paying more attention to key principles as well as depicting suitable approaches within a common framework.

---

[19] Gates, "The CIA and American Foreign Policy."
[20] Omand, *Securing the State*, 225.
[21] Clark, *Intelligence Analysis: A Target-Centric Approach*, 101.
[22] Clark, 101.

Conclusion

The pursuit to find ways to mitigate intelligence failures has been an enduring one, for which numerous official committees and review boards have suggested mitigations, although usually limited to separate parts of the intelligence process. The post-9/11 and Iraq WMD intelligence and policy failures led not only to further debates about several aspects of intelligence theory, but also a more encompassing intelligence community reform than previously implemented, both in content and geography. A major driver for the impact of the reform was that many of the recommendations from governmental review committees were turned into policy, and for the United States also enshrined in law, the Intelligence Reform and Terrorism Prevention Act (IRTPA) of 2004. On the subject of estimative intelligence, prominence was given to implementation of analytic tradecraft standards aimed at increasing analytic rigour and improving communication of uncertainties to policymakers. The core idea behind these reforms was arguably to prevent discourse failure, a new and more comprehensive theory of intelligence failure arguing that failures are the result of incomprehension of the identification and communication of possible significant threats.

Following the IRTPA, the newly established office of the Director of National Intelligence (ODNI) issued new analytic standards, aimed at increasing analytic rigour through Structured analytic Techniques (SATs), together with new and altered standardised expressions of uncertainty by implementing separate tables of verbal probabilities and levels of analytic confidence. For the United States, the strive to improve both intelligence analysis methodology and expressions of uncertainty was not new but has had an intermittent focus, especially in the CIA, ever since Sherman Kent's days. This time, however, the reform spread to Europe as well with early efforts in Defence Intelligence in the United Kingdom. Soon after, other countries, such as Norway and the Netherlands, followed suit, mostly because of the international effort to fight Islamic terrorism. 10-15 years after 9/11, SATs had become the *de facto* "science" of intelligence analysis in the western intelligence community, and new standardized scales of both verbal probabilities and analytic confidence had been implemented in intelligence doctrines.

Together with the reform came a call for validation of the new so-called alternative analysis methods. Alternative analysis, better known as Structured Analytic Techniques (SATs), has by many been claimed to be debiasing techniques that will challenge established mindsets and produce more accurate intelligence, thereby mitigating intelligence failures. The intelligence reform did, however, originally focus more on alternative analysis to ensure increased objectivity and integrity, hereunder clear communication of key assumptions, uncertainties, and possible alternative outcomes. Coupled with an American obsession for metrics, the academic focus has nonetheless been on measuring to what degree single techniques mitigate certain biases or increase accuracy, to which the research has yielded mixed results. The research correctly points at some limitations as to how several of the SATs are described today, but the value of the critique is lessened since only Analysis of Competing Hypotheses (ACH) has been repeatedly tested, but unfortunately not in the same manner or in real-life like situations, making one question the reliability and validity of the research.

The previous research and academic literature have in general treated SATs as being applied in a far more mechanistic and orthodox manner than they are in practice. Previous research has moreover taken the North American approach as a universal application of SATs and has thus not sufficiently considered of how different SATs function together. By doing so, academics forget that intelligence analysis is conducted as part of a larger intelligence process. SATs are seldom used as stand-alone techniques, but rather as part of a process, in which the previous research for instance has not differentiated between SATs for different intelligence problem types. Furthermore, there has been no recognition that different nations and agencies may have different approaches. Anyone can go through the steps of a specific SAT but unless the application of SATs is approached with the right attitude and openness for creative and critical thinking, the techniques will only hide the analysts' exiting bias and subjectivity. Most importantly, however, accurate prediction is not the point, the crux of the matter is rather an integral use of different techniques to advance the analysis in situations of change and uncertainty. In other words, the effort should be on how to apply relevant scientific concepts to a non-scientific world where the wanted output is a close as possible proximity to the truth.

This research, in comparison, has broken new ground and shifts the paradigm on SATs research by investigating the effects of using a comprehensive SATs-methodology where different techniques are being applied in a layered and iterative manner by small analytical teams. This methodology furthermore combines different modes of reasoning, inductive, deductive, but above all abductive reasoning, with sensemaking and creative and critical thinking skills. In this process, the different SATs function as building blocks, applied to create both knowledge and understanding, and with a focus on using answers to secrets as indication and warning for mysteries and complexities. Unlike previous research, the data collection for this thesis has been gathered over a longer timeframe and involved both quantitative as well as qualitative data. In addition to having the research conducted in a controlled environment of a series of two-week SATs-courses, a follow-up study among serving intelligence analysts has been conducted to investigate long-time effects. Lastly, a comparative analysis has been conducted, involving the three main intelligence agencies in the Norwegian intelligence community. This research is consequently the most far-reaching research conducted on SATs to this date.

This research has revealed that Norwegian intelligence analysts experience that using SATs in a comprehensive, layered, and iterative manner in the estimative intelligence process increases analytic rigour and makes it easier to assess uncertainty. Using a creative, critical, and sensemaking logic-process increases analysts' analytic objectivity and integrity and thereby making them more confident in their key judgments and aware of the attached uncertainties. One can thus conclude that the overall effect of using this comprehensive SATs-methodology is reduced analytic uncertainty through increased analytic rigour, resulting in a 'best truth' that reduces the chances of discourse failure. Consequently, the use of a comprehensive SATs-methodology for estimative intelligence analysis fulfils two of the key requirements of the post-9/11 intelligence reform.

The research did, however, also reveal that intelligence analysts recognise that there is more to managing uncertainty than structuring the analysis. The realization that other factors also impact results in slightly lower experienced effect of using SATs on analysts' ability to assess uncertainty from an overstated pre-course expected effect. This reality check of SATs' impact seemingly continues over time, nevertheless without diminishing the lasting higher ability to assess uncertainty. A main reason for this finding is arguably that there as of today exists no SAT for improving the calculus of probability in estimative intelligence, still leaving it up to the analysts' judgmental heuristics. For the analytic rigour dimension, however, the experienced effect of using SATs was the same across the two surveys, indicating a solid, lasting effect. The research did nonetheless reveal that participants reporting other methods to be better than SATs also reported a lower SATs' effect on analytic rigour. A reasonable explanation can be found in another key result, significant differences in the use and the effects of SATs on account of agency cultures.

The research found significant differences in analytic culture between personnel from different Norwegian intelligence agencies, the Norwegian Intelligence Service, the Norwegian Intelligence Battalion, and the Norwegian Security Police. There were significant differences both in the use of SATs and in the communication of uncertainties across the different agencies. The research revealed that prior training and experience in using SATs are necessary conditions for continued use of SATs. Those who have SATs as part of their analytical DNA tend to continue to use SATs even in agencies where there is no culture for using such techniques. Even so, for SATs to become the analytical norm in an agency, endorsement of SATs usage by management and top leaders is vital.

The results of this research furthermore point towards the art-science debate of intelligence being a false dichotomy and that estimative intelligence analysis is a tradecraft that demands both scientific and "artistic" approaches. The choice of approach depends on many different factors, such as the intelligence problem at hand, timeframe, the quantity and quality of the available information, the type of assumptions needed, and not the least the type of methodological training one has received. However, the NORDIS SATs-methodology is no universal solution and the intra-agency cultural differences towards SATs also suggests it would be beneficiary to pursue a less doctrinal approach to a structured intelligence analysis methodology. The longer timeframe of an intelligence problem, the less need for a strict doctrinal methodology. Instead, intelligence analysis courses should make the analysts capable of creating their own procedures within a shared overall structured framework, where they can, based upon different factors and criteria, pick different suitable techniques that fits their problem sets, improving their ability to solve the task at hand. To do so, intelligence analysts need to learn to use far more techniques than what is normally done in contemporary SATs-courses.

This research has given a quite clear picture of the situation in Norway. The results of this research are of direct relevance for the intelligence community, and one is looking forward to the findings from similar research in allied intelligence communities. The nature of contemporary intelligence problems necessitates the intelligence community to investigate several interlinked factors in a structured manner to improve forecasts about the future. The findings should therefore have a large impact on future training of intelligence analysts, hopefully impacting on cultural and organisational issues in the long run. Subsequently, the research's key findings potentially impact on limiting the possibility for future intelligence failures, especially with reference to the renewed focus on strategic warning, intelligence estimates that no longer are limited to the military sphere.

Although this research has provided promising results, using a structured intelligence analysis methodology in estimative intelligence may not bring the intelligence community that much closer to prevent future intelligence failures. According to Betts, intelligence failures are natural and inevitable due to several basic barriers that can be difficult to mitigate.[1] Organizational inertia, judicial restraints, atrophy of reform, time-pressure, and lack of resources are just a few factors that in addition to ambiguous data and information can contribute to new surprises. Ben-Israel sums it up: 'intelligence, like science, deals with hypotheses, and their very nature is their fallibility. Errors, therefore, can never be avoided, no matter what method is used. The most that can be done is to localize and narrow the margin of error'.[2] Dahl, moreover, questions the strive towards encouraging a greater use of imagination, arguing it has a limited effect since policymakers are not likely to believe in occurrences of new or unexpected threats.[3] In other words, in the land of the blind, the one-eyed is king while the two-eyed is a heretic. As a consequence, consumers of intelligence need to understand the limits to intelligence analysis, that one cannot eliminate uncertainty, and that they therefore have to be able to apprehend any potential consequences of

---

[1] Betts, "Analysis, War, and Decision: Why Intelligence Failures Are Inevitable."
[2] Ben-Israel, "Philosophy and Methodology of Intelligence: The Logic of Estimate Process," 694.
[3] Dahl, "Missing the Wake-up Call: Why Intelligence Failures Rarely Inspire Improved Performance."

uncertain estimates themselves. What they should demand, however, is being presented with as robust and rigorous estimates as possible.

What the research presented in this thesis does suggest, however, is that using a structured analytic methodology which focuses on developing comprehensive past, current and alternative future models of an intelligence problem can render the current versus long-term debate pointless. Although the future always will be uncertain, identifying indicators for how different scenarios can evolve from the present can function as warnings to policymakers if key indicators manifest themselves. This combination of future scenarios and the attached indicators could satisfy both the need for more long-term strategic analysis as well as policymakers' need for current and actionable intelligence. Moreover, the scenarios and indicators work as focal points for a more relevant collection effort.

Consequently, a more comprehensive process of estimative intelligence analysis can in many ways be compared with an integrated social science methodology. Instead of limiting the debate of the utility of SATs to the effect of single techniques, measured against unattainable goals of accuracy or bias elimination, the development of intelligence theory would benefit from investigating any implications on how to pursue the process of estimative intelligence analysis within the overall intelligence process. Wheaton, for instance, suggests discarding the traditional intelligence cycle and replace it with a thinking in parallel process.[4] Since this research found that using a comprehensive SATs-methodology increases analytic rigour and makes it easier to assess uncertainty, a key question for future research should therefore be to investigate what implications using this or structured methodologies from other intelligence communities will have on the intelligence process, both in a current and long-term perspective.

[4] Wheaton, "Let's Kill the Intelligence Cycle."

Bibliography

Abdalla, Neveen S., and Philip H.J. Davies. "Intelligence, Policy, and the Mandate: A Third Form of Strategic Failure." *International Journal of Intelligence, Security, and Public Affairs* 23, no. 2 (2021): 105–24. https://doi.org/10.1080/23800992.2021.1881724.

Agrell, Wilhelm. "When Everything Is Intelligence – Nothing Is Intelligence." *The Sherman Kent Center for Intelligence Analysis Occasional Papers* 1, no. 4 (2002): 1–6.

Agrell, Wilhelm, and Gregory F. Treverton. *National Intelligence and Science: Beyond the Great Divide in Analysis and Policy*. Oxford: Oxford University Press, 2015.

Albers, Casper, and Daniël Lakens. "When Power Analyses Based on Pilot Data Are Biased: Inaccurate Effect Size Estimators and Follow-up Bias." *Journal of Experimental Social Psychology* 74, no. October 2017 (January 2018): 187–95. https://doi.org/10.1016/j.jesp.2017.09.004.

Álvarez, Asunción, and Tom Ritchey. "Applications of General Morphological Analysis: From Engineering Design to Policy Analysis." *Acta Morphologica Generalis* 4, no. 1 (2015): 1–40. http://www.amg.swemorph.com/pdf/amg-4-1-2015.pdf.

Arnauld, Antoine, and Pierre Nicole. *Logic or the Art of Thinking*. Edited by Jill Vance Buroker. Cambridge: Cambridge University Press, 1996.

Artner, Stephen, Richard Girven, and James Bruce. "Assessing the Value of Structured Analytic Techniques in the U.S. Intelligence Community." RAND Corporation, 2016. https://doi.org/10.7249/RR1408.

Babetski, Frank J. "Intelligence in Public Literature: Thinking, Fast and Slow." *Studies in Intelligence* 56, no. 2 (2012): 1–4.

Bar-Joseph, Uri. "Intelligence Failure and the Need for Cognitive Closure: The Case of Yom Kippur." In *Paradoxes of Strategic Intelligence: Essays in Honor of Michael I. Handel*, edited by Richard K. Betts and Thomas G. Mahnken, 166–89. Abingdon: Routledge, 2003.

———. "Israel's 1973 Intelligence Failure." In *Revisiting the Yom Kippur War*, edited by P. R. Kumaraswamy, 11–35. Hoboken: Taylor & Francis, 2013.

Barclay, Scott, Rex V. Brown, Clinton W. Kelly, Cameron R. Peterson, Lawrence D. Philips, and Judith Selvidge. *Handbook for Decisions Analysis*. McLean, Va.: Decision and Designs,Inc., 1977.

Barnes, Alan. "Making Intelligence Analysis More Intelligent: Using Numeric Probabilities." *Intelligence and National Security* 31, no. 3 (April 15, 2016): 327–44. https://doi.org/10.1080/02684527.2014.994955.

Beadle, Alexander William. "FFIs Prediksjonsturnering: Datagrunnlag Og Foreløpige Resultater. FFI-Rapport 21/00737." Kjeller, 2022. https://www.ffi.no/publikasjoner/arkiv/ffis-prediksjonsturnering-datagrunnlag-og-forelopige-resultater.

Ben-Israel, Isaac. "Philosophy and Methodology of Intelligence: The Logic of Estimate Process." *Intelligence and National Security* 4, no. 4 (1989): 660–718. https://doi.org/10.1080/02684528908432023.

Berkovitz, Joseph. "The Propensity Interpretation of Probability." *Erkenn*, 2015, 629–711. https://doi.org/10.1007/s10670-014-9716-8.

Berkowitz, Bruce D., and Allen A. Goodman. *Best Truth: Intelligence in the Information Age*. New Haven: Yale University Press, 2000.

Betts, Richard K. “Analysis, War, and Decision: Why Intelligence Failures Are Inevitable.” *World Politics* 31, no. 1 (October 13, 1978): 61–89. https://doi.org/10.2307/2009967.

———. *Enemies of Intelligence: Knowledge and Power in American National Security*. New York: Colombuia University Press, 2007.

———. “Policy-makers and Intelligence Analysts: Love, Hate or Indifference?” *Intelligence and National Security* 3, no. 1 (January 1988): 184–89. https://doi.org/10.1080/02684528808431934.

Bishop, Peter, Andy Hines, and Terry Collins. “The Current State of Scenario Development: An Overview of Techniques.” *Foresight* 9, no. 1 (2007): 5–25. https://doi.org/10.1108/14636680710727516.

Borg, Lars C. “Improving Intelligence Analysis: Harnessing Intuition and Reducing Biases by Means of Structured Methodology.” *The International Journal of Intelligence, Security, and Public Affairs* 19, no. 1 (January 2, 2017): 2–22. https://doi.org/10.1080/23800992.2017.1289747.

Bradfield, Ron, George Wright, George Burt, George Cairns, and Kees Van Der Heijden. “The Origins and Evolution of Scenario Techniques in Long Range Business Planning.” *Futures* 37, no. 8 (2005): 795–812. https://doi.org/10.1016/j.futures.2005.01.003.

Brasfield, A D. “Forecasting Accuracy and Cognitive Bias in the Analysis of Competing Hypotheses.” Mercyhurst College, 2009.

Brinkmann, Svend, and Steinar Kvale. *Doing Interviews*. Edited by Uwe Flick. London: SAGE Publications, Ltd, 2018.

Brun, Wibecke, and Karl Halvor Teigen. “Verbal Probabilities: Ambiguous, Context-Dependent, or Both?” *Organizational Behavior and Human Decision Processes* 41, no. 3 (1988): 390–404. https://doi.org/10.1016/0749-5978(88)90036-2.

Budescu, David V., Tzur M. Karelitz, and Thomas S. Wallsten. “Predicting the Directionality of Probability Words from Their Membership Functions.” *Journal of Behavioral Decision Making* 16, no. 3 (2003): 159–80. https://doi.org/10.1002/bdm.440.

Budescu, David V., Han-Hui Por, Stephen B. Broomell, and Michael Smithson. “The Interpretation of IPCC Probabilistic Statements around the World.” *Nature Climate Change* 4, no. 6 (2014): 508–12. https://doi.org/10.1038/nclimate2194.

Budescu, David V., Han Hui Por, and Stephen B. Broomell. “Effective Communication of Uncertainty in the IPCC Reports.” *Climatic Change* 113, no. 2 (2012): 181–200. https://doi.org/10.1007/s10584-011-0330-3.

Budescu, David V., and Thomas S. Wallsten. “On Two Complenetary Appraoches to the Study of Verbal Probabilities.” In *Perspectives on Thinking, Judging and Decision Making a Tribute to Karl Halvor Teigen*, edited by Wibecke Brun, Gideon Keren, Geir Kirkebøen, and Henry Montgomery, 36–46. Oslo: Universitetsforlaget, 2011.

Budescu, David V, Stephen Broomell, and Han-Hui Por. “Improving Communication of Uncertainty in the Reports of the Intergovernmental Panel on Climate Change.” *Psychological Science* 20, no. 3 (2009): 299–308. https://doi.org/10.1111/j.1467-9280.2009.02284.x.

Burson, Katherine A., Richard P Larrick, and Joshua Klayman. “Skilled or Unskilled, but Still Unaware of It: How Perceptions of Difficulty Drive Miscalibration in Relative Comparisons.” *Journal of Personality and Social Psychology* 90, no. 1 (2006): 60–77. https://doi.org/10.1037/0022-3514.90.1.60.

Canadian Forces Intelligence Command. “Aide Memoire on Intelligence Analysis Tradecraft.” Ottawa: Canadian Forces Intelligence Command, 2015.

Cardillo, Robert. “A Cultural Evolution.” *Studies in Intelligence* 54, no. 3 (2010): 1–8.
Central Intelligence Agency. “A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis,” 2009. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/Tradecraft Primer-apr09.pdf.
———. “CIA: Who We Are and What We Do.” SpyKids Briefing: Who we are and what we do, 2022. https://www.cia.gov/spy-kids/parents-teachers/docs/Briefing-who-we-are-and-what-we-do.pdf.
Central Intelligence Agency Office of Research and Development. “General Notice No. 91: Establishment of Analytical Methodology Research Division (AMR/ORD).” Washington, D.C.: Central Intelligence Agency, 1977. https://www.cia.gov/library/readingroom/docs/CIA-RDP88B00553R000100180034-2.pdf.
Chang, Welton, Elissabeth Berdini, David R. Mandel, and Philip E. Tetlock. “Restructuring Structured Analytic Techniques in Intelligence.” *Intelligence and National Security* 33, no. 3 (April 16, 2018): 337–56. https://doi.org/10.1080/02684527.2017.1400230.
Chang, Welton, Eva Chen, and Barbara Mellers. “Developing Expert Political Judgment: The Impact of Training and Practice on Judgmental Accuracy in Geopolitical Forecasting Tournaments.” *Judgment and Decision Making* 11, no. 5 (2016): 509–26.
Chang, Welton, and Philip E. Tetlock. “Rethinking the Training of Intelligence Analysts.” *Intelligence and National Security* 31, no. 6 (2016): 903–20. https://doi.org/10.1080/02684527.2016.1147164.
Chilcot, John. “The Report of the Iraq Inquiry: Executive Summary.” Vol. 1. London, 2016.
CIA Senior Review Panel. “Report on a Study of Intelligence Judgments Preceding Significant Historical Failures: The Hazards of Single-Outcome Forecasting.” Washington, D.C., 1983. https://www.cia.gov/library/readingroom/docs/CIA-RDP86B00269R001100100010-7.pdf.
Clark, Robert M. *Intelligence Analysis: A Target-Centric Approach*. 6th ed. Los Angeles: CQ Press, 2020.
Cogan, Charles. “Hunters Not Gatherers: Intelligence in the Twenty-First Century.” *Intelligence and National Security* 19, no. 2 (June 2004): 304–21. https://doi.org/10.1080/0268452042000302010.
Cohen J. *Statistical Power Analysis for the Behavioural Science*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.
Cohen, M R, and E Nagel. *An Introduction to Logic and Scientific Method*. New York: Harcourt, Brace, 1934.
Convertino, Gregorio, Dorrit Billman, Peter Pirolli, J. P. Massar, and Jeff Shrager. “The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis.” *Computer Supported Cooperative Work (CSCW)* 17, no. 4 (August 16, 2008): 353–93. https://doi.org/10.1007/s10606-008-9080-9.
Corbetta, Piergiorgio. *Social Research: Theory, Methods and Techniques*. London: SAGE Publications, Ltd, 2003. https://doi.org/10.4135/9781849209922.
Corkill, Jeff. “Evaluation a Critical Point on the Path to Intelligence.” *Journal of the Australian Institute of Professional Intelligence Officers* 16, no. 1 (2008): 3–11.
Cortina, Jose M. “What Is Coefficient Alpha? An Examination of Theory and Applications.” *Journal of Applied Psychology* 78, no. 1 (1993): 98–104. https://doi.org/10.1037/0021-9010.78.1.98.

Coulthart, Stephen. "Why Do Analysts Use Structured Analytic Techniques? An in-Depth Study of an American Intelligence Agency." *Intelligence and National Security* 31, no. 7 (November 9, 2016): 933–48. https://doi.org/10.1080/02684527.2016.1140327.

Coulthart, Stephen J. "An Evidence-Based Evaluation of 12 Core Structured Analytic Techniques." *International Journal of Intelligence and CounterIntelligence* 30, no. 2 (2017): 368–91. https://doi.org/10.1080/08850607.2016.1230706.

———. "Improving the Analysis of Foreign Affairs: Evaluating Structured Analytic Techniques." University of Pittsburgh, 2015.

Dahl, Erik J. *Intelligence and Surprise Attack: Failure and Success from Pearl Harbor to 9/11 and Beyond*. Washington, DC: Georgetown University Press, 2013.

———. "Missing the Wake-up Call: Why Intelligence Failures Rarely Inspire Improved Performance." *Intelligence and National Security* 25, no. 6 (December 16, 2010): 778–99. https://doi.org/10.1080/02684527.2010.537876.

Daston, Lorraine. *Classical Probability in the Enlightenment*. Princeton, N.J: Princeton University Press, 1988.

———. "How Probabilities Came to Be Objective and Subjective." *Historia Mathematica* 21, no. 3 (1994): 330–44. https://doi.org/10.1006/hmat.1994.1028.

Davies, Jack. "Strategic Warning: If Surprise Is Inevitable, What Role for Analysis?" *The Sherman Kent Center for Intelligence Analysis, Occasional Papers* 2, no. 1 (2003): 2–16.

Davies, Philip H. J. *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 1: Evolution of the U.S. Intelligence Community*. Oxford: Praeger, 2012.

———. *Intelligence and Government in Britain and the United States: A Comparative Perspective. Volume 2: Evolution of the UK Intelligence Community*. Oxford: Praeger, 2012.

———. "Jointery versus Tradecraft: The Brunel Analytical Simulation and Alternative Approaches to Intelligence Analysis and Analytical Professionalization in Postgraduate Academic Teaching." In *The Art of Intelligence: Simulations, Exercises, and Games*, edited by William J. Lahneman and Rubén Arcos, 200–218. Lanham, MD: Rowman & Littlefield Publishers, 2014.

Davies, Philip H.J. "Intelligence Culture and Intelligence Failure in Britain and the United States." *Cambridge Review of International Affairs* 17, no. 3 (October 21, 2004): 495–520. https://doi.org/10.1080/0955757042000298188.

Davis, Jack. "Defining the Analytic Mission: Facts, Findings, Forecasts, and Fortune-Telling." *Studies in Intelligence* 39, no. 3 (1995): 25–30.

———. "Improving CIA Analytic Performance : Strategic Warning." *The Sherman Kent Center for Intelligence Analysis Occasional Papers* 1, no. September (2002): 1–9. https://www.cia.gov/library/kent-center-occasional-papers/pdf/OPNo1.pdf.

———. "Sherman Kent and the Profession of Intelligence Analysis." *The Sherman Kent Center for Intelligence Analysis Occasional Papers* 1, no. 5 (March 2002). https://www.cia.gov/library/readingroom/docs/CIA-RDP98-00412R000300010016-9.pdf.

Defence Intelligence. "Quick Wins for Busy Analysts." London: Ministry of Defence, 2013.

Development Concepts and Doctrine Centre. "Joint Doctrine Publication 04: Understanding and Decision-Making." *Joint Doctrine Publication*. London: Ministry of Defence, 2016.

Dhami, Mandeep K., Ian K. Belton, and David R. Mandel. "The 'Analysis of Competing Hypotheses' in Intelligence Analysis." *Applied Cognitive Psychology* 33, no. 6 (November

30, 2019): 1080–90. https://doi.org/10.1002/acp.3550.
Dhami, Mandeep K., Ian K Belton, and Kathryn E Careless. “Critical Review of Analytic Techniques.” In *European Intelligence and Security Informatics Conference (EISIC) 2016*, 4, 2016. https://doi.org/10.1109/EISIC.2016.33.
Dhami, Mandeep K., and David R. Mandel. “Words or Numbers? Communicating Probability in Intelligence Analysis.” *American Psychologist* 76, no. 3 (April 2021): 549–60. https://doi.org/10.1037/amp0000637.
Dhami, Mandeep K., David R. Mandel, Barbara A. Mellers, and Philip E. Tetlock. “Improving Intelligence Analysis With Decision Science.” *Perspectives on Psychological Science* 10, no. 6 (November 17, 2015): 753–57. https://doi.org/10.1177/1745691615598511.
Dillman, Don A. *Mail and Internet Surveys: The Tailored Design Method, 2nd Ed. Mail and Internet Surveys: The Tailored Design Method, 2nd Ed.*, 2007.
Director of Central Intelligence. “National Intelligence Estimate 11-18-91: Implications of Alternative Soviet Futures.” Washington, D.C., 1991.
———. “National Intelligence Estimate 15-90: Yugoslavia Transformed.” Washington, D.C., 1990.
DISI Analysis Instruction Group. *Analysis Handbook: Theory and Methodology in Intellgience Analysis (HB 30-31)*. ’t Harde: Netherlands Defence Intelligence and Security Institute, 2010.
Doran, Charles F. “Why Forecasts Fail: The Limits and Potential of Forecasting in International Relations and Economics.” *International Studies Review* 1, no. 2 (September 1999): 11–41. https://doi.org/10.1111/1521-9488.00154.
Durbin, Brent. *The CIA and the Politics of US Intelligence Reform*. Cambridge: Cambridge University Press, 2017. https://doi.org/10.1017/9781316941317.
Eijk, Cees van der, and Jonathan Rose. “Risky Business: Factor Analysis of Survey Data – Assessing the Probability of Incorrect Dimensionalisation.” *PLOS ONE* 10, no. 3 (March 19, 2015): 1–31. https://doi.org/10.1371/journal.pone.0118900.
Eisinga, Rob, Manfred Te Grotenhuis, and Ben Pelzer. “The Reliability of a Two-Item Scale: Pearson, Cronbach, or Spearman-Brown?” *International Journal of Public Health* 58, no. 4 (2013): 637–42. https://doi.org/10.1007/s00038-012-0416-3.
Empericus, Sextus. *Outlines of Pyrrhonism*. Edited by R. G. Bury. Digital Lo. London: W. Heinemann, 1933.
Etterretningstjenesten. “FOKUS 2022: Etterretningstjenestens Vurdering Av Aktuelle Sikkerhetsutfordringer.” Oslo, 2022.
Facione, Peter A. “Critical Thinking: A Statement of Expert Consuncus for Purposes of Educational Assessment and Instruction.” *The Delphi Report - Executive Summary*, 1990.
Fastvold, Morten. “Socrates versus Peter Pan : The Difficulties of Thinking Critically,” 2011. http://www.fastvold-filopraksis.com/artikler_pdf/Socrat_vs_PetPan.pdf.
Field, Andy. *Discovering Statistics Using IBM SPSS Statistics*. 4th ed. London: SAGE Publications, Inc., 2013.
Fischhoff, Baruch, and Don MacGregor. “Subjective Confidence in Forecasts.” *Journal of Forecasting* 1, no. 2 (1982): 155–72. https://doi.org/10.1002/for.3980010203.
Fishbein, Warren, and Gregory F. Treverton. “Making Sense of Transnational Threaths.” *Sherman Kent Center Occasional Papers* 3, no. 1 (2004): 1–25.
Folker Jr, Robert D. “Intelligence Analysis in Theater Joint Intellgience Centers: An Experiment in Applying Structured Methods.” *Joint Military Intelligence College Occasional Papers* 7, no. 7 (2000): 1–43.

Forsvaret. "Etterretningsdoktrinen." Oslo: Forsvaret, 2013.
———. "Etterretningsdoktrinen." Oslo: Forsvaret, 2021.
Fox, Cr, and Gülden Ülkümen. "Distinguishing Two Dimensions of Uncertainty." In *Perspectives on Thinking, Judging, and Decision Making*, edited by W Brun, G Keren, G Kirkebøen, and H Montgomery, 21–35. Oslo: Universitetsforlaget, 2011.
Frassen, Bas C. van. *Laws and Symmetry*. Oxford: Oxford University Press, 1989.
Friedman, Jeffrey A., Joshua D. Baker, Barbara A. Mellers, Philip E. Tetlock, and Richard Zeckhauser. "The Value of Precision in Probability Assessment: Evidence from a Large-Scale Geopolitical Forecasting Tournament." *International Studies Quarterly* 62, no. 2 (2018): 410–22. https://doi.org/10.1093/isq/sqx078.
Friedman, Jeffrey a., and Richard Zeckhauser. "Assessing Uncertainty in Intelligence." *Intelligence and National Security* 27, no. 6 (2012): 1–24. https://doi.org/10.1080/02684527.2012.708275.
Friedman, Jeffrey A., and Richard Zeckhauser. "Analytic Confidence and Political Decision-Making: Theoretical Principles and Experimental Evidence From National Security Professionals." *Political Psychology* 39, no. 5 (October 2018): 1069–87. https://doi.org/10.1111/pops.12465.
———. "Handling and Mishandling Estimative Probability: Likelihood, Confidence, and the Search for Bin Laden." *Intelligence and National Security* 0, no. 0 (2014): 1–23. https://doi.org/10.1080/02684527.2014.885202.
Friedman, Jeffrey A. "Probability and Confidence," 2017.
———. *War and Chance: Assessing Uncertainty in International Politics*. New York: Oxford University Press, 2019.
Gardiner, L. Keith. "Squaring the Circle: Dealing with Intelligence-Policy Breakdowns." *Intelligence and National Security* 6, no. 1 (January 2, 1991): 141–53. https://doi.org/10.1080/02684529108432094.
———. "Squaring the Circle: Dealing with Intelligence-policy Breakdowns." *Intelligence and National Security* 6, no. 1 (January 2, 1991): 141–53. https://doi.org/10.1080/02684529108432094.
Gates, Robert M. "A Message to Analysts: Guarding Against Politicization." *Studies in Intelligence* 36, no. 1 (1992): 5–13. https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/volume-36-number-1/pdf/v36i1a01p.pdf.
———. "The CIA and American Foreign Policy." *Foreign Affairs* 66, no. 2 (1987): 215–30. https://doi.org/10.2307/20043370.
Gauch, Hugh G. Jr. *Scientific Method in Practise*. Cambridge: Cambridge University Press, 2003.
Gelder, Tim Van, Ashley Barnett, Tamar Primoratz, Richard de Rozario, Morgan Saletta, and Luke Thorburn. "Analytic Rigour in Intelligence." Melbourne, 2021.
Gelder, Tim van, Richards J. Heuer, and Stephen Marrin. "Van Gelder's 'How to Free Your Mind' Re ACH," 2010.
Gelfand, Lawrence Emerson. *The Inquiry: American Preparations for Peace, 1917-1919*. New Haven: Yale University Press, 1963.
Gelman, Andrew. "Objections to Bayesian Statistics." *Bayesian Analysis* 3, no. 3 (2008): 445–50. https://doi.org/10.1214/08-BA318.
Gentry, John A. "Intelligence Failure Reframed." *Political Science Quarterly* 123, no. 2 (2008): 247–70. http://www.jstor.org/stable/20203011.
George, Roger Z. "Fixing the Problem of Analytical Mind-Sets: Alternative Analysis."

International Journal of Intelligence and CounterIntelligence 17, no. 3 (September 11, 2004): 385–404. https://doi.org/10.1080/08850600490446727.
Gettier, Edmund L. “Is Justified True Belief Knowledge?” *Analysis* 23, no. 6 (1963): 121–23. http://www.jstor.org/stable/3326922.
Gigerenzer, Gerd. “The Adaptive Toolbox.” In *Bounded Rationality - The Adaptive Toolbox*, edited by Gerd Gigerenzer and R. Selten, 37–50. Cambridge, MA: The MIT Press, 2001. http://pubman.mpdl.mpg.de/pubman/item/escidoc:2102373/component/escidoc:2102372/GG_Adaptive_Toolbox_2001.pdf.
———. “Why Heuristics Work.” *Perspectives on Psychological Science* 3, no. 1 (January 1, 2008): 20–29. https://doi.org/10.1111/j.1745-6916.2008.00058.x.
Gill, Jeff. *Bayesian Methods: A Social and Behavioral Sciences Approach*. 3rd ed. Boca Raton: CRC Press, 2015. http://public.eblib.com/choice/publicfullrecord.aspx?p=1598081.
Gill, Peter, and Mark Phythian. “Developing Intelligence Theory.” *Intelligence and National Security* 33, no. 4 (June 7, 2018): 467–71. https://doi.org/10.1080/02684527.2018.1457752.
———. “From Intelligence Cycle to Web of Intelligence: Complexity and the Conceptualisation of Intelligence.” In *Understanding the Intelligence Cycle*, edited by Mark Phythian. London: Routledge, 2013. https://doi.org/10.4324/9780203558478.
———. *Intelligence in an Insecure World*. 2nd ed. Cambridge: Polity Press, 2012.
Gillies, Donald. *Philosophical Theories of Probability*. Philosophical Issues in Science. London: Routledge, 2000.
Glassner, Amnon, and Baruch B. Schwarz. “What Stands and Develops between Creative and Critical Thinking? Argumentation?” *Thinking Skills and Creativity* 2, no. 1 (April 2007): 10–18. https://doi.org/10.1016/j.tsc.2006.10.001.
Grabo, Cynthia. *Handbook of Warning Intelligence: Assessing the Threat to National Security*. Lanham, MD: Scarecrow Press, 2010.
Granåsen, Magdalena, and Maja Karasalo. “Methodology and Tool to Facilitate Structured Analysis of Multiple Hypotheses,” 2016, 52–59. https://doi.org/10.1109/EISIC.2016.46.
Gray, Colin D., and Paul R Kinnear. *IBM SPSS Statistics 19 Made Simple*. *IBM SPSS Statistics 19 Made Simple*. 4th ed. London: Taylor and Francis, 2011. https://www.taylorfrancis.com/books/9780203723524.
Greiff, Samuel, Christoph Niepel, and Sascha Wüstenberg. “21st Century Skills: International Advancements and Recent Developments.” *Thinking Skills and Creativity* 18 (December 2015): 1–3. https://doi.org/10.1016/j.tsc.2015.04.007.
Griffin, Dale, and Amos Tversky. “The Weighing of Evidence and the Determinants of Confidence.” *Cognitive Psychology* 24, no. 3 (July 1992): 411–35. https://doi.org/10.1016/0010-0285(92)90013-R.
Guilford, Joy Paul. *The Nature of Human Intelligence*. New York: McGraw-Hill, 1967.
Guillan, Amanda. “Epistemological Limits to Scientific Prediction: The Problem of Uncertainty.” *Open Journal of Philosophy* 04, no. 04 (2014): 510–17. https://doi.org/10.4236/ojpp.2014.44053.
Gustafson, Kristian. “Assumption... the Necessary Evil,” 2021.
———. “You Can’t Yell Loud Enough: Intelligence and Warning.” *International Studies Review* 17, no. 2 (June 2015): 329–31. https://doi.org/10.1111/misr.12215.
Guy, James Joy. *People, Politics and Government: A Canadian Perspective*. Scarborough, Ont: Prentice Hall, 2001.
Hacking, I. *The Emergence of Probability: A Philosophical Study of Early Ideas about*

Probability, Induction and Statistical Inference. Cambridge: Cambridge University Press, 1975. https://books.google.no/books?id=ja00AAAAQBAJ.
Hacking, Ian. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press, 2001. https://doi.org/10.1017/CBO9780511801297.
Halford, Graeme S., Rosemary Baker, Julie E. McCredden, and John D. Bain. "How Many Variables Can Humans Process?" *Psychological Science*, 2005. https://doi.org/10.1111/j.0956-7976.2005.00782.x.
Hammer, Mike. "How Accurate Are Weather Forecasts?" World Science Festival, 2014. https://www.worldsciencefestival.com/2014/02/how_accurate_is_the_weather_forecast/.
Handel, M I. *War, Strategy, and Intelligence*. Cass Series on Politics and Military Affairs in the Twentieth Century. New York: F. Cass, 1989.
Harding, Sandra G. *Can Theories Be Refuted? Essays on the Duhem-Quine Thesis*. Edited by Sandra G. Harding. Dordrecht: D. Reidel Publishing Company, 1976.
Harman, Gilbert H. "The Inference to the Best Explanation." *The Philosphical Review* 74, no. 1 (1965): 88–95. https://www.jstor.org/stable/2183532.
Harris, Adam J. L., and Adam Corner. "Communicating Environmental Risks: Clarifying the Severity Effect in Interpretations of Verbal Probability Expressions." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37, no. 6 (2011): 1571–78. https://doi.org/10.1037/a0024195.
Hastie, Reid. "Group Processes in Intelligence Analysis." In *Intelligence Analysis: Behavioral and Social Science Foundations*, 169–96. Washington, DC: The National Academies Press, 2011. https://doi.org/https://doi.org/10.17226/13062.
Hatlebrekke, K A. *The Problem of Secret Intelligence*. Edinburgh Scholarship Online. Edinburgh: Edinburgh University Press, 2019. https://books.google.no/books?id=WA5MzQEACAAJ.
Hatlebrekke, Kjetil Anders, and M. L.R. Smith. "Towards a New Theory of Intelligence Failure? The Impact of Cognitive Closure and Discourse Failure." *Intelligence and National Security* 25, no. 2 (2010): 147–82. https://doi.org/10.1080/02684527.2010.489274.
Hendrickson, Noel. "Critical Thinking in Intelligence Analysis." *International Journal of Intelligence and CounterIntelligence* 21, no. 4 (September 8, 2008): 679–93. https://doi.org/10.1080/08850600802254749.
Herbert, Matthew. "The Intelligence Analyst as Epistemologist." *International Journal of Intelligence and CounterIntelligence* 19, no. November 2014 (2006): 666–84. https://doi.org/10.1080/08850600600829890.
Herman, M. *Intelligence Power in Peace and War*. Cambridge: Cambridge University Press, 1996.
Heuer, Richards J. "Adapting Academic Methods and Models to Governmental Needs: The CIA Experience." Carlisle Barracks, PA, 1978.
———. "Adapting Academic Methods and Models to Governmental Needs." In *Quantitative Approaches to Political Intelligence: The CIA Experience*, 1–10. New York: Routledge, 2019. https://doi.org/10.4324/9780429303647-2.
———. *Psychology of Intelligence Analysis*. Washington, DC: CIA Center for the Study of Intelligence, 1999.
———. "Small Group Processes for Intelligence Analysis." Washington, D.C., 2008.
Heuer, Richards J., and Randolph H. Pherson. *Structured Analytic Techniques for Intelligence*

Analysis. 1st ed. Thousand Oaks: CQ Press, 2010.
———. *Structured Analytic Techniques for Intelligence Analysis*. 3rd ed. Thousand Oaks: CQ Press, 2020.
Hilsman, Roger. "Intelligence and Policy-Making in Foreign Affairs." *World Politics* 5, no. 1 (October 18, 1952): 1–45. https://doi.org/10.2307/2009086.
Hintikka, Jaakko. *Logic, Language-Games and Information: Kantian Themes in the Philosophy of Logic*. Oxford: Clarendon Press, 1973.
Hitchcock, David. "Critical Thinking as an Educational Ideal." In *On Reasoning and Argument: Essays in Informal Logic and on Critical Thinking*, 477–97. Dordrecht: Springer, 2017. https://doi.org/10.1007/978-3-319-53562-3_30.
Holbrook, Allyson L., Jon A. Krosnick, and Alison Pfent. "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms." In *Advances in Telephone Survey Methodology*, 499–528. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2007. https://doi.org/10.1002/9780470173404.ch23.
Holland, Kathlyn, Lyn Middleton, and Leana Uys. "Professional Confidence: A Concept Analysis." *Scandinavian Journal of Occupational Therapy* 19, no. 2 (2012): 214–24. https://doi.org/10.3109/11038128.2011.583939.
Horn, John L. "A Rationale and Test for the Number of Factors in Factor Analysis." *Psychometrika*, 1965. https://doi.org/10.1007/BF02289447.
Hume, David. *An Enquiry Concerning Human Understanding*. Edited by L. A Selby-Bigge. 2nd ed. Oxford: Project Gutenberg EBook, 1902. http://www.gutenberg.org/files/9662/9662-h/9662-h.htm.
Hutcheson, Graeme, and Nick Sofroniou. *The Multivariate Social Scientist*. SAGE Publications, Ltd, 1999. https://doi.org/10.4135/9780857028075.
Jeremiah, David. "Recommendations of the Jeremiah Report," 1998. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB187/IN38.pdf.
Johansen, Iver. "Scenario Modelling with Morphological Analysis." *Technological Forecasting and Social Change* 126, no. May 2017 (January 2018): 116–25. https://doi.org/10.1016/j.techfore.2017.05.016.
Johnson, B, and L Turner. "Data Collection Strategies in Mixed Methods Research." In *Handbook of Mixed Methods in Social and Behavioral Research*, edited by A Tashakkori and C Teddlie, 297–319. Thousand Oaks, California: SAGE Publications, Inc., 2003.
Johnson, Loch K. *America's Secret Power: The CIA in a Democratic Society*. New York: Oxford University Press, 1989.
Johnston, Judith. "Defining Analytic Rigor for Analysis in the Intelligence Community (Unpublished Report)." Raleigh, 2020.
Johnston, R. *Analytic Culture in the US Intelligence Community: An Ethnographic Study*. *The Center for the Study of Intelligence*. Washington, DC: Government Printing Office, 2005.
Johnston, Rob. "Developing a Taxonomy of Intelligence Analysis Variables." *Studies in Intelligence* 47, no. 3 (2003). https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol47no3/article05.html#_ftn5.
Joint Chiefs of Staff. "Joint Publication 2-0 - Joint Intelligence." Washington, D.C.: Department of Defense, 2013.
———. "Joint Publication 2-01.3: Joint Intelligence Preparation of the Operational Environment." Washington, D.C.: Joint Chiefs of Staff, 2014.
Jones, Morgan D. *The Thinker's Toolkit: Fourteen Powerful Techniques for Problem Solving*. New York: Three Rivers Press, 1998.

Jones, Reginald Victor. *Most Secret War*. London: Hamilton, 1978.
Kahneman, Daniel. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist* 58, no. 9 (2003): 697–720. https://doi.org/10.1037/0003-066X.58.9.697.
———. *Thinking Fast and Slow*. New York: Macmillan, 2011.
Kahneman, Daniel, and Amos Tversky. "Variants of Uncertainty." *Cognition* 11, no. 2 (1982): 143–57. https://doi.org/10.1016/0010-0277(82)90023-3.
Kant, Immanuel. *The Critique of Pure Reason*. Edited by J. M. D. Meiklejohn. Project Gutenberg EBook, 2003.
Kaplan, Abraham. *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler Publishing Company, 1964.
Karvetski, Christopher W., David R. Mandel, and Daniel Irwin. "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation." *Risk Analysis* 40, no. 5 (2020): 1040–57. https://doi.org/10.1111/risa.13443.
Karvetski, Christopher W., Kenneth C. Olson, Donald T. Gantz, and Glenn A. Cross. "Structuring and Analyzing Competing Hypotheses with {Bayesian} Networks for Intelligence Analysis." *EURO Journal on Decision Processes* 1, no. 3–4 (November 2013): 205–31. https://doi.org/10.1007/s40070-013-0001-x.
———. "Structuring and Analyzing Competing Hypotheses with Bayesian Networks for Intelligence Analysis." *EURO Journal on Decision Processes* 1, no. 3–4 (2013): 205–31. https://doi.org/10.1007/s40070-013-0001-x.
Kennedy, Robert. *Of Knowledge and Power: The Complexities of National Intelligence*. London: Praeger, 2008.
Kent, Sherman. "A Crucial Estimate Relived." *Studies in Intelligence* 35, no. 5 (1964): 11–119.
———. *Strategic Intelligence for American World Policy*. Princeton: Princeton University Press, 1949. https://doi.org/10.1515/9781400879151.
———. "Words of Estimative Probability." *Studies in Intelligence* 8, no. 4 (1964): 49–65.
Kerby, Dave S. "The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation." *Comprehensive Psychology* 3 (2014): 11.IT.3.1. https://doi.org/10.2466/11.it.3.1.
Kesselman, Rachel F. "Verbal Probability Expressions in National Intelligence Estimates : A Comprehensive Analysis of Trends from the Fifties through Post 9 /11." Mercyhurst College, 2008.
Keynes, John Maynard. *A Tract on Monetary Reform*. London: Macmillan and Co, Limited, 1924.
———. *A Treaties on Probability*. London: Macmillan and Co, Limited, 1921.
Khalsa, Sundri. "The Intelligence Community Debate over Intuition versus Structured Technique: Implications for Improving Intelligence Warning." *Journal of Conflict Studies* 29, no. 0 (2009): 75–95. http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=11988614&AN=91714468&h=nrTPIWmWKK9M4YzdHZYt8zyi8NInikWJ/SSTDbehkxnqs1znDuDnDTqyAOgIcE2YcFFlSZIeC3Pn2bjVwHYfWA==&crl=c.
Kindsvater, Larry C. "The Need to Reorganize the Inteffigence Community." *Studies in Intelligence* 47, no. 1 (2003): 33–37.
Kirk, Roger E. "Practical Significance: A Concept Whose Time Has Come." *Educational and Psychological Measurement* 56, no. 5 (1996): 746–59.
Kirkham, Richard L. "Does the Gettier Problem Rest on a Mistake ?" *Mind* 93, no. 372 (1984):

501–13. http://www.jstor.org/stable/2254258.
Knight, Frank. *Risk, Uncertainty and Profit*. Boston: Hart, Schaffner & Marx, 1921.
Koriat, Asher, Sarah Lichtenstein, and Baruch Fischhoff. "Reasons for Confidence." *Journal of Experimental Psychology: Human Learning & Memory* 6, no. 2 (1980): 107–18. https://doi.org/10.1037/0278-7393.6.2.107.
Kretz, Donald R., and Charles W. Granderson. "An Interdisciplinary Approach to Studying and Improving Terrorism Analysis." In *2013 IEEE International Conference on Intelligence and Security Informatics*, 157–59. IEEE, 2013. https://doi.org/10.1109/ISI.2013.6578808.
Kretz, Donald R., B. J. Simpson, and Colonel Jacob Graham. "A Game-Based Experimental Protocol for Identifying and Overcoming Judgment Biases in Forensic Decision Analysis." In *2012 IEEE Conference on Technologies for Homeland Security (HST)*, 439–44. IEEE, 2012. https://doi.org/10.1109/THS.2012.6459889.
Krizan, Lisa. "Intelligence Essentials for Everyone." *Joint Military Intelligence College Occasional Papers*, no. 6 (1999): 1–81. http://www.au.af.mil/au/awc/awcgate/dia/intel_essentials_krizan.pdf.
Krummel, John W M. "Imagination, Formation, World, and Place: An Ontology." In *Imagination : Cross-Cultural Philosophical Analyses*, 1st ed., 163–86. London: Bloomsbury Academic, 2019. http://www.bloomsburycollections.com/book/imagination-cross-cultural-philosophical-analyses/ch10-imagination-formation-world-and-place-an-ontology/.
Kuhn, Kristine M., and Janet a. Sniezek. "Confidence and Uncertainty in Judgmental Forecasting: Differential Effects of Scenario Presentation." *Journal of Behavioral Decision Making* 9, no. 4 (1996): 231–47. https://doi.org/10.1002/(SICI)1099-0771(199612)9:4<231::AID-BDM240>3.0.CO;2-L.
Kuhns, Woodrow J. "Intelligence Failures: Forecasting and the Lessons of Epistemology." In *Paradoxes of Strategic Intelligence: Essays in Honor of Michael I. Handel*, edited by Richard K. Betts and Thomas G. Mahnken. London: Routledge, 2003. https://doi.org/10.4324/9780203508640.
Kuipers, Theo A. F. "Inference to the Best Theory, Rather than Inference to the Best Explanation — Kinds of Abduction and Induction." In *Induction and Deduction in the Sciences*, 25–51. Dordrecht: Springer, 2004. https://doi.org/10.1007/978-1-4020-2196-1_3.
Kuipers, Theo A.F. "Empirical Progress and Truth Approximation by the 'Hypothetico-Probabilistic Method.'" *Erkenntnis* 70, no. 3 (2009): 313–30. https://doi.org/10.1007/s10670-008-9126-x.
Labovitz, Sanford. "Some Observations on Measurement and Statistics." *Social Forces* 46, no. 2 (1967): 151–60. https://www.jstor.org/stable/2574595.
Lakatos, Imre. *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*. Edited by John Worall and Gregory Currie. Cambridge: Cambridge University Press, 1978.
Laplace, Pierre Simon Marquis de. *A Philosophical Essay on Probabilities*. Translated by Fredrick Wilson Truscott and Fredrick Lincoln Emory. New York: John Wiley & Sons, 1902.
Laquer, Walter. *World of Secrets: The Uses and Limits of Intelligence*. New York: Basic Books, 1985.
Lee, Allen S. "A Scientific Methodology for MIS Case Studies." *MIS Quarterly* 13, no. 1 (March

1989): 33. https://doi.org/10.2307/248698.
Lehner, Paul Edward, Leonard Adelman, Brant A. Cheikes, and Mark J. Brown. “Confirmation Bias in Complex Analyses.” *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans* 38, no. 3 (2008): 584–92. https://doi.org/10.1109/TSMCA.2008.918634.
Lewis, David K. “Elusive Knowledge.” *Australasian Journal of Philosophy* 74, no. 4 (1996): 549–67.
Liddell Hart, Basil Henry. *Strategy*. New York: Praeger, 1967.
Lissitz, Robert W., and Samuel B. Green. “Effect of the Number of Scale Points on Reliability: A Monte Carlo Approach.” *Journal of Applied Psychology* 60, no. 1 (1975): 10–13. https://doi.org/10.1037/h0076268.
Løhre, Erik, and Karl Halvor Teigen. “How Fast Can You (Possibly) Do It, or How Long Will It (Certainly) Take? Communicating Uncertain Estimates of Performance Time.” *Acta Psychologica* 148 (2014): 63–73. https://doi.org/10.1016/j.actpsy.2014.01.005.
———. “There Is a 60% Probability, but I Am 70% Certain: Communicative Consequences of External and Internal Expressions of Uncertainty.” *Thinking & Reasoning* 6783, no. November (2015): 1–28. https://doi.org/10.1080/13546783.2015.1069758.
Lord Butler, Robin. “Review of Intelligence on Weapons of Mass Destruction.” *House of Commons Committee Report*. London, 2004.
Lowenthal, Mark M. *Intelligence: From Secrets to Policy*. 8th ed. Thousand Oaks, California: CQ Press, 2020.
M. Isaksen, Bjørn Gunnar, and Ken R. McNaught. “Uncertainty Handling in Estimative Intelligence – Challenges and Requirements from Both Analyst and Consumer Perspectives.” *Journal of Risk Research* 22, no. 5 (May 4, 2019): 643–57. https://doi.org/10.1080/13669877.2018.1474245.
MacCallum, Robert C C, Keith F F Widaman, Sehee Hong, and Shaobo Zhang. “Sample Size in Factor Analysis.” *Psychological Methods* 4, no. 1 (1999): 84–99.
MacEachin, Douglas J. “The Tradecraft of Analysis: Challenge and Change in the CIA’s Directorate of Intelligence.” *General CIA Records*. Washington, D.C., 1994. https://www.cia.gov/library/readingroom/docs/CIA-RDP98-00412R000300010005-1.pdf.
MacKay, R. Bradley, and Peter McKiernan. *Scenario Thinking: A Historical Evolution of Strategic Foresight*. Elements in Business Strategy. Cambridge: Cambridge University Press, 2018. https://doi.org/DOI: 10.1017/9781108571494.
Mandel, David R. “Communicating Uncertainty, Assessing Information Quality and Risk, and Using Structured Techniques in Intelligence Analysis.” *System Analysis and Studies Panel 114*. Brussels: NATO, 2018.
Mandel, David R., and Alan Barnes. “Accuracy of Forecasts in Strategic Intelligence.” *Proceedings of the National Academy of Sciences* 111, no. 30 (July 29, 2014): 10984–89. https://doi.org/10.1073/pnas.1406138111.
———. “Geopolitical Forecasting Skill in Strategic Intelligence.” *Journal of Behavioral Decision Making* 31, no. 1 (2018): 127–37. https://doi.org/10.1002/bdm.2055.
Mandel, David R., Alan Barnes, and Karen Richards. “A Quantitative Assessment of the Quality of Strategic Intelligence Forecasts.” Toronto, 2014. http://www.researchgate.net/profile/David_Mandel3/publication/260833937_A_quantitative_assessment_of_the_quality_of_strategic_intelligence_forecasts/links/0a85e5327588234819000000.pdf.

Mandel, David R., Christopher W. Karvetski, and Mandeep K. Dhami. “Boosting Intelligence Analysts’ Judgment Accuracy: What Works, What Fails?” *Judgment and Decision Making* 13, no. 6 (2018): 607–21.
Marchio, James. “‘ If the Weatherman Can ...’: The Intelligence Community ’ s Struggle to Express Analytic Uncertainty in the 1970s.” *CIA- Studies in Intelligence* 58, no. 4 (2014): 31–42. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol-58-no-4/pdfs/If the Weatherman Can.pdf.
Marchio, Jim. “Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis.” *Intelligence and National Security*. Taylor & Francis, 2014. https://doi.org/10.1080/02684527.2012.746415.
———. “‘How Good Is Your Batting Average?’ Early IC Efforts To Assess the Accuracy of Estimates.” *Studies in Intelligence* 60, no. 4 (2016): 3–14.
Marquardt-Bigman, Petra. “Behemoth Revisited: The Research and Analysis Branch of the Office of Strategic Services in the Debate of Us Policies towards Germany, 1943–46.” *Intelligence and National Security* 12, no. 2 (April 1997): 91–100. https://doi.org/10.1080/02684529708432415.
Marrin, Stephen. “Analytic Objectivity and Science: Evaluating the US Intelligence Community’s Approach to Applied Epistemology.” *Intelligence and National Security* 35, no. 3 (April 15, 2020): 350–66. https://doi.org/10.1080/02684527.2019.1710806.
———. “Evaluating Intelligence Theories: Current State of Play.” *Intelligence and National Security* 33, no. 4 (June 7, 2018): 479–90. https://doi.org/10.1080/02684527.2018.1452567.
———. *Improving Intelligence Analysis: Bridging the Gap Between Scholarship and Practice*. London: Routledge, 2011.
———. “Is Intelligence Analysis an Art or a Science?” *International Journal of Intelligence and CounterIntelligence* 25, no. 3 (September 2012): 529–45. https://doi.org/10.1080/08850607.2012.678690.
———. “Why Strategic Intelligence Analysis Has Limited Influence on American Foreign Policy.” *Intelligence and National Security* 32, no. 6 (September 19, 2017): 725–42. https://doi.org/10.1080/02684527.2016.1275139.
Marrin, Stephen, and Jonathan D. Clemente. “Improving Intelligence Analysis by Looking to the Medical Profession.” *International Journal of Intelligence and CounterIntelligence* 18, no. 4 (December 2005): 707–29. https://doi.org/10.1080/08850600590945434.
Marrin, Stephen, and Efren Torres. “Improving How to Think in Intelligence Analysis and Medicine.” *Intelligence and National Security* 32, no. 5 (July 29, 2017): 649–62. https://doi.org/10.1080/02684527.2017.1311472.
Mastrandrea, Michael D, Christopher B Field, Thomas F Stocker, Ottmar Edenhofer, Kristie L Ebi, David J Frame, Hermann Held, et al. “Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties,” 2010. https://www.ipcc.ch/site/assets/uploads/2017/08/AR5_Uncertainty_Guidance_Note.pdf.
McCloskey, Joseph F. “British Operational Research in World War II.” *Operations Research* 35, no. 3 (June 1987): 453–70. https://doi.org/10.1287/opre.35.3.453.
McEwan, Desmond, Geralyn R. Ruissen, Mark A. Eys, Bruno D. Zumbo, and Mark R. Beauchamp. “The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions.” Edited by Nico W. Van Yperen. *PLOS ONE* 12, no. 1 (January 13, 2017):

1–23. https://doi.org/10.1371/journal.pone.0169604.
Mcmullin, Ernan. "The Inference That Makes Science." *Zygon* 48, no. 1 (2013): 143–91.
Mellers, Barbara, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, Michael M. Bishop, Michael Horowitz, Ed Merkle, and Philip Tetlock. "The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics." *Journal of Experimental Psychology: Applied* 21, no. 1 (2015): 1–14. https://doi.org/10.1037/xap0000040.
Mellers, Barbara, Eric Stone, Terry Murray, Angela Minster, Nick Rohrbaugh, Michael Bishop, Eva Chen, et al. "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions." *Perspectives on Psychological Science* 10, no. 3 (2015): 267–81. https://doi.org/10.1177/1745691615577794.
Miller, George A. "The Magical Number Seven, Plus Minus Two: Some Limits on Our Capacity for Processing Information." *The Psychological Review* 63, no. 2 (August 5, 1956): 81–97.
Mises, Richard Von. *Probability, Statistics, and Truth*. 2d rev. En. New York: Dover Publications, 1957.
———. *Probability, Statistics and Thruth*. Edited by Hilda Geiringer. 2nd Rev. E. New York: Dover Publications, 1981.
Mohd Razali, Nornadiah, and Yap Bee Wah. "Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests." *Journal of Statistical Modeling and Analytics* 2, no. 1 (2011): 21–33. http://instatmy.org.my/downloads/e-jurnal 2/3.pdf%0Ahttps://www.nrc.gov/docs/ML1714/ML17143A100.pdf.
Moore, David T. "Critical Thinking and Intelligence Analysis." Vol. Occasional. Occational Papers. Washington, D.C., 2007.
———. *Sensemaking: A Structure for an Intelligence Revolution*. Washington, D.C.: National Defense Intelligence College Press, 2011.
Morris, Alan. *A Practical Introduction to In-Depth Interviewing*. London: SAGE Publications, Ltd, 2015.
Moser, Paul K. *The Oxford Handbook of Epistemology*. *The Oxford Handbook of Epistemology*, 2009. https://doi.org/10.1093/oxfordhb/9780195301700.001.0001.
National Commission on Terrorist Attacks Upon the United States. "The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States." Washington, D.C., June 2004.
National Intelligence Council. "National Intelligence Estimate January 2007 - Prospects for Iraq's Stability: A Challenging Road Ahead." Washington, D.C.: Office of the Director of the National Intelligence, 2007.
———. "National Intelligence Estimate July 2007 - The Terrorist Threat to the US Homeland." Washington, D.C.: Office of the Director of the National Intelligence, 2007.
———. "National Intelligence Estimate November 2007 - Iran: Nuclear Intentions and Capabilities." Washington, D.C.: Office of the Director of the National Intelligence, 2007.
National Research Council. *Field Evaluation in the Intelligence and Counterintelligence Context: Workshop Summary*. Washington, D.C.: National Academies Press, 2010. https://doi.org/10.17226/12854.
———. *Intelligence Analysis for Tomorrow: Advances from the Behavioral and Social Sciences*. Edited by Baruch Fischhoff and Cherie Chauvin. Washington, DC: The National Academies Press, 2011. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc8&NEWS=N&AN=2

012-00579-000.
NATO. *AJP-2(B) Allied Joint Doctrine for Intelligence, Counter-Intelligence and Security*. Edition B. Brussels: NATO Standardization Office, 2020.
———. “Alternative Analysis Handbook.” Mons; Norfolk, VA: Supreme Allied Commanders, Europe and Transformation, 2012.
———. “NATO Standard AIntP-18(A) Intelligence Processing.” Brussels: NATO Standardization Office, 2020.
———. “NATO Standard AJP-2.1(B): Allied Joint Doctrine for Intelligence Procedures.” Brussels: NATO Standardization Office, 2016.
Neary, Patrick C. “Intelligence Reform, 2001–2009: Requiesat in Peace?” *Studies in Intelligence* 54, no. 1 (2010): 1–16.
Neumann, Peter R., and Michael L.R. M.L.R. Smith. “Missing the Plot? Intelligence and Discourse Failure.” *Orbis* 49, no. 1 (December 2005): 95–107. https://doi.org/10.1016/j.orbis.2004.10.008.
Njå, Ove, Øivind Solberg, and Geir Sverre Braut. “Uncertainty-Its Ontological Status and Relation to Safety.” In *The Illusion of Risk Control: What Does It Take to Live With Uncertainty?*, edited by Gilles Motet and Corinne Bieder, 5–21. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-32939-0_2.
Norwegian Defence Intelligence School. “Study Programme Bachelor of Arts in Intelligence Studies.” Oslo: NORDIS, 2013.
Office of the Director of National Intelligence. “Intelligence Community Assessment 2017-01D: Assessing Russian Activities and Intentions in Recent US Elections.” Washington, D.C., 2017.
———. “Intelligence Community Directive 203: Analytic Standards.” Washington, D.C.: Office of the Director of National Intelligence, 2007.
———. “Intelligence Community Directive 203: Analytic Standards,” 2015.
———. “Intelligence Community Directive 203: Analytic Standards.” Washington, D.C.: Office of the Director of National Intelligence, 2015.
Olson, Michael J, and David V Budescu. “Patterns of Preference for Numerical and Verbal Probabilities.” *Journal of Behavioral Decision Making* 10, no. September 1996 (1997): 117–32. https://doi.org/10.1002/(SICI)1099-0771(199706)10:2<117::AID-BDM251>3.0.CO;2-7.
Omand, David. *Securing the State*. New York: Oxford University Press, 2010.
———. “The Future of Intelligence: What Are the Threats, the Challenges and the Opportunities?” In *The Future of Intelligence: Challenges in the 21st Century*, 14–26, 2014. https://doi.org/10.4324/9780203071472.
Oxford English Dictionary. “Estimate.” Oxford English Dictionary, 2021. https://www.oed.com/view/Entry/64581?rskey=n0LPjF&result=1#eid.
Paik, Jaehyon, Peter L. Pirolli, Christian Lebiere, and Matthew Rutledge-Taylor. “Cognitive Biases in a Geospatial Intelligence Analysis Task: An ACT-R Model.” *Proceedings of the Annual Meeting of the Cognitive Sciences Society*, 2012, 2168–73.
Pallant, Julie. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. 6th editio. Maidenhead New York: McGraw Hill Education, 2016.
Pedersen, Tore, and Christian Johansen. “Behavioural Artificial Intelligence: An Agenda for Systematic Empirical Studies of Artificial Inference.” *AI & SOCIETY* 35, no. 3 (September 18, 2020): 519–32. https://doi.org/10.1007/s00146-019-00928-5.
Pendleton-Jullian, A M, and J S Brown. *Pragmatic Imagination: Single from Design Unbound*.

Blurb (self-publishing platform), 2016.
Peterson, D, and G Pitz. "Confidence, Uncertainty, and the Use of Information." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, no. 1 (1988): 85–92. https://doi.org/10.1037/0278-7393.14.1.85.
Peterson, Joshua J. "Appropriate Factors to Consider When Assessing Analytic Confidence in Intelligence Analysis," no. March (2008): 112. http://c.ymcdn.com/sites/www.scip.org/resource/resmgr/White_Papers/Peterson-Appropriate-Factors.pdf.
Pherson, Katherine Hibbs, and Randolph H. Pherson. *Critical Thinking for Strategic Intelligence*. Thousand Oaks, California: CQ Press, 2013.
Phythian, Mark. "Intelligence Analysis and Social Science Methods: Exploring the Potential for and Possible Limits of Mutual Learning." *Intelligence and National Security* 32, no. 5 (2017): 600–612. https://doi.org/10.1080/02684527.2017.1310972.
Pidd, M. *Tools for Thinking: Modelling in Management Science*. 3rd ed. West Sussex: Wiley, 2009.
Pierce, Charles Sanders. "Harvard Lectures on Pragmatism." In *Collected Papers Vol. 5: Pragmatism and Pramaticism*, 1903.
———. "On the Natural Classification of Arguments." In *Proceedings of the American Academy of Arts and Sciences*, 7:261–87. Boston: Welsch, Bigelow & Co, 1867.
Pincione, Guido, and Fernando R. Tesón. *Rational Choice and Democratic Deliberation: A Theory of Discourse Failure*. Cambridge: Cambridge University Press, 2006. https://doi.org/10.1017/CBO9780511720178.
Plano Clark, Vicki L, and Nataliya V Ivankova. "Mixed Methods Research: A Guide to the Field." Thousand Oaks, California: SAGE Publications, Inc., 2016. https://doi.org/10.4135/9781483398341.
Politiets sikkerhetstjeneste (PST). "Etterretningsdoktrine for PST." Oslo: PST, 2017.
———. "National Threat Assessment 2020." Oslo, 2020. https://urldefense.proofpoint.com/v2/url?u=https-3A__www.vsd.lt_wp-2Dcontent_uploads_2020_02_2020-2DGresmes-2DEn.pdf&d=DwMBaQ&c=KveGjKEXiH4bMFgGs-LRbCbewnnyGW6-rJ0JK7ViA_E&r=TLJquTB5XdBVyMhwHD8t8akdwfsjcokLy3EbJBDDfes&m=lgE-H0cq7HiWvg8lPro1R1u5DfWzNzdclTa.
———. "National Threat Assessment 2021." Oslo, 2021. https://urldefense.proofpoint.com/v2/url?u=https-3A__www.vsd.lt_wp-2Dcontent_uploads_2020_02_2020-2DGresmes-2DEn.pdf&d=DwMBaQ&c=KveGjKEXiH4bMFgGs-LRbCbewnnyGW6-rJ0JK7ViA_E&r=TLJquTB5XdBVyMhwHD8t8akdwfsjcokLy3EbJBDDfes&m=lgE-H0cq7HiWvg8lPro1R1u5DfWzNzdclTa.
Pope, Simon, and Audun Jøsang. "Analysis of Competing Hypotheses Using Subjective Logic." *Tenth International Command and Control Research and Technology Symposium*, no. June (2005): 29.
Popper, Karl R. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge, 1963.
———. "The Propensity Interpretation of Probability." *The British Journal for the Philosophy of Science* 10, no. 37 (1959): 25–42. http://www.jstor.org/stable/685773.
Professional Head of Defence Intelligence Analysis. "Quick Wins for Busy Analysts." London: UK Ministry of Defence, 2016. https://doi.org/10.32388/tfm82d.

Professional Head of Intelligence Assessment. “Professional Development Framework for All-Source Intellgience Assessment.” London: Crown, 2019.
Ramsey, Frank P. “Truth and Probability.” *The Foundations of Mathematics and Other Logical Essays* Ch. VII, no. 1926 (1931): 156–98. https://doi.org/10.1007/978-3-319-20451-2_3.
Rea, Louis M, and Richard A. Parker. *Designing and Conducting Survey Research : A Comprehensive Guide*. 4th ed. John Wiley & Sons, Inc., 2014. https://ebookcentral.proquest.com/lib/brunelu/detail.action?docID=1757959.
Reichenbach, H. *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability*. 2nd ed. California Library Reprint Series. Berkeley: University of California Press, 1971. https://books.google.no/books?id=WUm_2lAvj2cC.
Rejón Altable, Carlos. “Logic Structure of Clinical Judgment and Its Relation to Medical and Psychiatric Semiology.” *Psychopathology* 45, no. 6 (2012): 344–51. https://doi.org/10.1159/000337968.
Rescher, Nicholas. *Predicting the Future: An Introduction to Theory of Forecasting*. Albany, NY: Sate University of New York Press, 1998.
Richards, Julian. *The Art and Science of Intelligence Analysis*. Oxford: Oxford University Press, 2010.
Rieber, Steven. “Intelligence Analysis and Judgmental Calibration.” *International Journal of Intelligence and CounterIntelligence* 17, no. 1 (2004): 97–112. https://doi.org/10.1080/08850600490273431.
Riste, Olav, and Arnfinn Moland. *Strengt Hemmelig: Norsk Etterretningstjeneste 1945-1970*. Oslo: Universitetsforlaget, 1997.
Rittel, Horst W. J., and Melvin M. Webber. “Dilemmas in a General Theory of Planning.” *Policy Sciences* 4, no. 2 (June 1973): 155–69. https://doi.org/10.1007/BF01405730.
Rosenberg, Joab. “The Interpretation of Probability in Intelligence Estimation and Strategic Assessment.” *Intelligence and National Security* 23, no. 2 (2008): 139–52. https://doi.org/10.1080/02684520801977238.
Rosenthal, James A. “Qualitative Descriptors of Strength of Association and Effect Size.” *Journal of Social Service Research*, 1996. https://doi.org/10.1300/J079v21n04_02.
Rumsfeld, Donald. “DoD News Briefing - Secretary Rumsfeld and Gen. Myers,” 2002. http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636.
Salas, Eduardo, Nancy J. Cooke, and Michael A. Rosen. “On Teams, Teamwork, and Team Performance: Discoveries and Developments.” *Human Factors* 50, no. 3 (2008): 540–47. https://doi.org/10.1518/001872008X288457.
Sands, Robert R. Greene. “Thinking Differently: Unlocking the Human Domain in Support of the 21st Century Intelligence Mission.” *Small Wars Journal*, no. 20 August 2013 (2013).
Sawyer, R. Keith. “Educating for Innovation.” *Thinking Skills and Creativity* 1, no. 1 (April 2006): 41–48. https://doi.org/10.1016/j.tsc.2005.08.001.
Schmitt, Gary J. “Truth to Power? Rethinking Intelligence Analysis.” In *The Future of American Intellgience*, edited by Peter Berkowitz, 41–64. Stanford, CA: Hoover Institution Press, 2005.
Schrage, Michael. “What Percent Is ’Slam Dunk ’? Give Us Odds on Those Estimates.” *The Washington Post*. February 20, 2005. http://www.washingtonpost.com/wp-dyn/articles/A37115-2005Feb19.html.
Schwartz, Alan. “Scenarios for the Insurgency in Iraq.” *United States Institute of Peace - Special Report*, no. 174 (2006). https://www.usip.org/sites/default/files/SRoct06_3.pdf.

Secretary of State for Foreign and Commonwealth Affairs. “Review of Intelligence on Weapons of Mass Destruction: Implementations of Its Conclusions.” London, 2005.
Sherman Kent School Kent Center for Analytic Tradecraft. “A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis.” *Tradecraft Review*. Washington, D.C.: Central Intelligence Agency, 2005.
Shuffler, Marissa L., Deborah Diazgranados, M. Travis Maynard, and Eduardo Salas. “Developing, Sustaining, and Maximizing Team Effectiveness: An Integrative, Dynamic Perspective of Team Development Interventions.” *Academy of Management Annals* 12, no. 2 (June 2018): 688–724. https://doi.org/10.5465/annals.2016.0045.
Shulsky, Abram M., and Gary James Schmitt. *Silent Warfare: Understanding the World of Intelligence, 2nd Ed*. *Armed Forces & Society*, 2002.
Sieck, Winston, and J.Frank Yates. “Exposition Effects on Decision Making: Choice and Confidence in Choice.” *Organizational Behavior and Human Decision Processes* 70, no. 3 (1997): 207–19. https://doi.org/10.1006/obhd.1997.2706.
Silberman, Laurence H., and Charles S. Robb. “The Commission on the Intelligence Capabilities Weapons of Mass Destruction: Report to the President of the United States.” Washington, DC, 2005.
Simon, Herbert A. “Rational Choice and the Structure of the Environment.” *Psychological Review* 63, no. 2 (1956): 129–38.
Smith, Abbot E. “On the Accuracy of National Intelligence Estimates.” *Studies in Intelligence* 13, no. 4 (1969): 25–35.
Snowden, David. “Complex Acts of Knowing: Paradox and Descriptive Self-Awareness.” *Journal of Knowledge Management* 6, no. 2 (2002): 100–111. https://doi.org/10.1108/13673270210424639.
Stacey, Ralph D. *Complex Responsive Processes in Organizations: Learning and Knowledge Creation*. London: Routledge, 2001.
Stenslie, Stig. “The Necessity of Experts.” In *Intelligence Analysis in the Digital Age*, 65–83. Oslo: Routledge, 2021.
Stromer-Galley, Jennifer, Patricia Rossini, Kate Kenski, Brian McKernan, Benjamin Clegg, James Folkestad, Carsten Østerlund, et al. “Flexible versus Structured Support for Reasoning: Enhancing Analytical Reasoning through a Flexible Analytic Technique.” *Intelligence and National Security* 00, no. 00 (November 2, 2020): 1–20. https://doi.org/10.1080/02684527.2020.1841466.
Strong, Kenneth. *Intelligence at the Top: The Recollections of an Intellgience Officer*. London: Cassel, 1968.
Tabachnick, Barbara G, and Linda S Fidell. *Using Multivariate Statistics: Pearson New International Edition*. 6th ed. Harlow: Pearson Education UK, 2013.
Taleb, Nassim Nicholas. *The Black Swan: The Impact of the Highly Improbable*. London: Penguin, 2010.
Tang, Jeffrey. “How Do We Know? What Intelligence Analysis Can Learn from the Sociology of Science.” *Intelligence and National Security* 32, no. 5 (2017): 663–74. https://doi.org/10.1080/02684527.2017.1311473.
Tashakkori, Abbas, and Charles Teddlie. *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., 2010. https://doi.org/10.4135/9781506335193.
Teddlie, Charles, and Abbas Tashakkori. “Overview of Contemporary Issues in Mixed

Methods Research." In *SAGE Handbook of Mixed Methods in Social & Behavioral Research*, edited by Charles Teddlie and Abbas Tashakkori, 2nd ed., 1–42. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., 2010. https://doi.org/10.4135/9781506335193.n1.

Teigen, Karl Halvor. "Risk Communication in Words and Numbers." In *Narratives of Risk*, edited by Karen Patrick Knutsen, Sigmund Kvam, Peter Langemeyer, Anastasia Parianou, and Kåre Solfjeld, 240–54. Münster: Waxmann, 2012.

———. "The Language of Uncertainty." *Acta Psychologica* 68, no. 1–3 (1988): 27–38. https://doi.org/10.1016/0001-6918(88)90043-1.

———. "Variants of Subjective Probabilities: Concepts, Norms, and Biases." In *Subjective Probability.*, edited by George Wright and Peter Ayton, 211–38. Chichester: John Wiley & Sons, 1994.

Teigen, Karl Halvor, and Wibecke Brun. "The Directionality of Verbal Probability Expressions: Effects on Decisions, Predictions, and Probabilistic Reasoning." *Organizational Behavior and Human Decision Processes* 80, no. 2 (November 1999): 155–90. https://doi.org/10.1006/obhd.1999.2857.

———. "Verbal Probabilities: A Question of Frame?" *Journal of Behavioral Decision Making* 16, no. 1 (January 2003): 53–72. https://doi.org/10.1002/bdm.432.

Teigen, Karl Halvor, and Petra Filkuková. "Can>Will: Predictions of What Can Happen Are Extreme, but Believed to Be Probable." *Journal of Behavioral Decision Making* 26, no. 1 (January 2013): 68–78. https://doi.org/10.1002/bdm.761.

Teigen, Karl Halvor, Marie Juanchich, and Anine H. Riege. "Improbable Outcomes: Infrequent or Extraordinary?" *Cognition* 127, no. 1 (2013): 119–39. https://doi.org/10.1016/j.cognition.2012.12.005.

Tetlock, Philip E. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton: Princeton University Press, 2005. http://www.degruyter.com/view/books/9781400888818/9781400888818/9781400888818.xml.

———. "Reading Tarot on K Street." *The National Interest*, no. 103 (2009): 57–67. https://www.jstor.org/stable/42896251.

Tetlock, Philip E., and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. New York: Crown, 2015.

The Norwegian Intelligence Service. "Focus 2016: The Norwegian Intelligence Service's Assessment of Current Security Challenges." Oslo, 2016.

———. "Focus 2017: The Norwegian Intelligence Service's Assessment of Current Security Challenges." Oslo, 2017.

———. "Focus 2018: The Norwegian Intelligence Service's Assessment of Current Security Challenges." Oslo, 2018.

———. "Focus 2019: The Norwegian Intelligence Service's Assessment of Current Security Challenges." Oslo, 2019.

———. "Focus 2020: The Norwegian Intelligence Service's Assessment of Current Security Challenges." Oslo, 2020.

———. "Focus 2022: The Norwegian Intelligence Service's Assessment of Current Security Challenges." Oslo, 2022.

Toosi, Nahal, Daniel Lippman, and Dan Diamond. "Before Trump's Inauguration, a Warning: 'The Worst Influenza Pandemic since 1918.'" *Politico*, March 16, 2020.

Treverton, Gregory, Seth Jones, Steven Boraz, and Phillip Lipscy. *Toward a Theory of*

Intelligence: Workshop Report. *Toward a Theory of Intelligence: Workshop Report*. Arlington: RAND Corporation, 2006. https://doi.org/10.7249/CF219.
Turner, Brandon M, Mark Steyvers, Edgar C Merkle, David V Budescu, and Thomas S Wallsten. "Forecast Aggregation via Recalibration." *Machine Learning* 95, no. 3 (June 12, 2014): 261–89. https://doi.org/10.1007/s10994-013-5401-4.
Tversky, A., and D. Kahneman. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185, no. 4157 (1974): 1124–31. https://doi.org/10.1126/science.185.4157.1124.
Ülkümen, Gülden, Craig R. Fox, and Bertram F. Malle. "Two Dimensions of Subjective Uncertainty: Clues from Natural Language." *Journal of Experimental Psychology: General* 145, no. 10 (2016): 1280–97. https://doi.org/10.1037/xge0000202.
United States Congress. Intelligence reform and terrorism prevention act of 2004, 108–458 Public Law 108-458 § (2004). https://doi.org/10.1007/s13398-014-0173-7.2.
United States Department of Justice. "Common Competencies for State, Local, and Tribal Intelligence Analysis." Washington, D.C.: United States Department of Justice, 2010.
United States Senate Committe on Governmental Affairs. "Summary of Intelligence Reform and Terrorism Prevention Act of 2004." Washington, D.C.: GPO, 2004. http://www.gpo.gov/fdsys/pkg/PLAW-108publ458/pdf/PLAW-108publ458.pdf.
United States Senate Select Committee on Intelligence. "Senate Report 110- Report on Whether Public Statements Regarding Iraq by U.S. Government Officials Were Substanciated by Intelligence Information." Washington, D.C., 2008.
US Air Force, and United States Air Force. "Air Force Handbook 14-133- Intelligence Analysis," 2017. https://fas.org/irp/doddir/usaf/afh14-133.pdf.
Vandepeer, Charles. *Applied Thinking for Intelligence Analysis: A Guide for Practitioners*, 2014. www.airforce.gov.au/airpower.
———. *Applied Thinking for Intelligence Analysis*, n.d.
Vrist Rønn, Kira. "(Mis-) Informed Decisions? On Epistemic Reasonability of Intelligence Claims." *International Journal of Intelligence and CounterIntelligence* 27, no. 2 (June 7, 2014): 351–67. https://doi.org/10.1080/08850607.2014.842813.
Walker, W.E., P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, and M.P. Krayer von Krauss. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." *Integrated Assessment* 4, no. 1 (March 2003): 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.
Walker, Warren E., Marjolijn Haasnoot, and Jan H. Kwakkel. "Adapt or Perish: A Review of Planning Approaches for Adaptation under Deep Uncertainty." *Sustainability (Switzerland)* 5, no. 3 (2013): 955–79. https://doi.org/10.3390/su5030955.
Walker, Warren E, Robert J Lempert, and Jan H Kwakkel. "Deep Uncertainty." In *Encyclopedia of Operations Research and Management Science*, edited by Saul I Gass and Michael C Fu, 395–402. Boston, MA: Springer US, 2013. https://doi.org/10.1007/978-1-4419-1153-7_1140.
Wallsten, T S, and D V Budescu. "A Review of Human Linguistic Probability Processing - General-Principles and Empirical-Evidence." *Knowledge Engineering Review* 10, no. 1 (1995): 43. isi:A1995QW27100004.
Wallsten, Thomas S., David V. Budescu, Amnon Rapoport, Rami Zwick, and Barbara Forsyth. "Measuring the Vague Meanings of Probability Terms." *Journal of Experimental Psychology: General* 115, no. 4 (1986): 348–65. https://doi.org/10.1037/0096-3445.115.4.348.
Walter, Laqueur. "The Question of Judgment: Intelligence and Medicine,." *The Journal of*

Contemporary History 18, no. 4 (1983): 533.
Walton, Douglas. *Abductive Reasoning*. Tuscaloosa: The University of Alabama Press, 2005.
Wark, David L. "The Definitions of Some Estimative Expressions." *Studies in Intelligence/ CIA*, 1964. https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol8no4/html/v08i4a07p_0001.htm.
Warner, Michael. "Wanted: A Definition of 'Intelligence.'" In *Secret Intelligence: A Reader*, edited by Christopher Andrew, Richard J. Aldrich, and Wesley K. Wark, 2nd ed., 4–12. Abingdon: Routledge, 2020. https://doi.org/10.4324/9780429029028-2.
Warner, Michael, and J. Kenneth Mcdonald. "US Intelligence Community Reform Studies Since 1947." *Studies in Intelligence*. Washington, D.C.: Center for the Study of Intelligence, 2005. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/US Intelligence Community Reform Studies Since 1947.pdf.
Weathers, Danny, Subhash Sharma, and Ronald W. Niedrich. "The Impact of the Number of Scale Points, Dispositional Factors, and the Status Quo Decision Heuristic on Scale Reliability and Response Accuracy." *Journal of Business Research* 58, no. 11 SPEC. ISS. (2005): 1516–24. https://doi.org/10.1016/j.jbusres.2004.08.002.
Weick, Karl E. *Sensemaking in Organizations*. Thousand Oaks: SAGE Publications, 1995.
———. *The Social Psychology of Organizing*. 2nd ed. New York: McGraw-Hill, 1979.
Weick, Karl E, Kathleen M Sutcliffe, and David Obstfeld. "Organization Science and the Process of Sensemaking." *Organization Science* 16, no. 4 (2005): 409–21. http://web.ebscohost.com.ez.sun.ac.za/ehost/pdfviewer/pdfviewer?vid=3&hid=15&sid=f0a088d8-d323-4d85-9864-1372c542c0ec%40sessionmgr11.
Weihrich, Heinz. "The TOWS Matrix - A Tool for Situational Analysis." *Long Range Planning* 15, no. 2 (1982): 54–66.
Weiss, Charles. "Communicating Uncertainty in Intelligence and Other Professions." *International Journal of Intelligence and CounterIntelligence* 21, no. 1 (2007): 57–85. https://doi.org/10.1080/08850600701649312.
Westad, Odd Arne. "Concerning the Situation in 'A': New Russian Evidence on the Soviet Intervention in Afghanistan." *COLD WAR INTERNATIONAL HISTORY PROJECT BULLETIN* Winter, no. 8–9 (1996): 128–32.
Wheaton, Kristan J. "Let's Kill the Intelligence Cycle." *Competitive Intelligence* 15, no. 2 (2012): 9–24.
———. "The Revolution Begins on Page Five: The Changing Nature of NIEs." *International Journal of Intelligence and CounterIntelligence* 25, no. 2 (2012): 330–49. https://doi.org/10.1080/08850607.2012.652548.
Whitesmith, Martha. "The Efficacy of ACH in Mitigating Serial Position Effects and Confirmation Bias in an Intelligence Analysis Scenario." *Intelligence and National Security* 34, no. 2 (February 23, 2019): 225–42. https://doi.org/10.1080/02684527.2018.1534640.
Wilensky, Harold L. *Organizational Intelligence: Knowledge and Policy InGovernment and Industry*. New York: Basic Books, 1967.
Williamson, Jon. "Philosophies of Probability : Objective Bayesianism and Its Challenges." *Philosophy*, 2006, 1–41.
Williamson, Timothy. "ARMCHAIR PHILOSOPHY, METAPHYSICAL MODALITY AND COUNTERFACTUAL THINKING." *Proceedings of the Aristotelian Society (Hardback)* 105, no. 1 (June 2005): 1–23. https://doi.org/10.1111/j.0066-7373.2004.00100.x.

Windschitl, Paul D., and Elke U. Weber. “The Interpretation of ‘Likely’ Depends on the Context, but ‘70%’ Is 70% - Right? The Influence of Associative Processes on Perceived Certainty.” *Journal of Experimental Psychology: Learning Memory and Cognition* 25, no. 6 (1999): 1514–33. https://doi.org/10.1037/0278-7393.25.6.1514.

Wohlstetter, Roberta. *Pearl Harbor: Warning and Decision*. Stanford, CA: Stanford University Press, 1962.

Zegart, Amy B. *Spying Blind: The CIA, the FBI, and the Origins of 9/11*. Princeton, N.J: Princeton University Press, 2007.

Zelik, Daniel J., Emily S. Patterson, and David D. Woods. “Measuring Attributes of Rigor in Information Analysis.” In *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*, edited by Emily S. Patterson and Janet E. Miller, 65–83. London: Taylor & Francis, 2010.

Zlotnick, Jack. “Bayes’ Theorem for Intelligence Analysis.” *Studies in Intelligence* 16, no. 2 (1972): 43–52. https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol16no2/html/v16i2a03p_0001.htm.

Appendices

Appendix 1: Ethical approval

University Research Ethics Committee
Brunel University London
Kingston Lane
Uxbridge
UB8 3PH
United Kingdom

www.brunel.ac.uk

1 September 2017

LETTER OF APPROVAL

Applicant: Major Lars Borg

Project Title: SAT methodology and analytic confidence

Reference: 6855-SS-Aug/2017- 8163-2

Dear Major Lars Borg,

The Research Ethics Committee has considered the above application recently submitted by you.

The Chair, acting under delegated authority has agreed that there is no objection on ethical grounds to the proposed study. Approval is given on the understanding that the conditions of approval set out below are followed:

- The agreed protocol must be followed. Any changes to the protocol will require prior approval from the Committee by way of an application for an amendment.

Please note that:

- Research Participant Information Sheets and (where relevant) flyers, posters, and consent forms should include a clear statement that research ethics approval has been obtained from the relevant Research Ethics Committee.
- The Research Participant Information Sheets should include a clear statement that queries should be directed, in the first instance, to the Supervisor (where relevant), or the researcher. Complaints, on the other hand, should be directed, in the first instance, to the Chair of the relevant Research Ethics Committee.
- Approval to proceed with the study is granted subject to receipt by the Committee of satisfactory responses to any conditions that may appear above, in addition to any subsequent changes to the protocol.
- The Research Ethics Committee reserves the right to sample and review documentation, including raw data, relevant to the study.
- You may not undertake any research activity if you are not a registered student of Brunel University or if you cease to become registered, including abeyance or temporary withdrawal. As a deregistered student you would not be insured to undertake research activity. Research activity includes the recruitment of participants, undertaking consent procedures and collection of data. Breach of this requirement constitutes research misconduct and is a disciplinary offence.

Kind regards,

Peter R Hobson

Professor Peter Hobson

Chair, University Research Ethics Committee

Brunel University London

Appendix 2: Survey 1 questionnaires

Pre-Course Survey Structured Analytic Techniques						
	Age (alder)	20-25; 26-30; 31-35; 36-40; 41-45; 46-50; 50+				
	Gender (Kjønn)	Male	Female			
	Unit (avdeling)	Armed Forces	Intel Service		Other	
	Years of service (År i tjeneste)	>1 year	2-4 years	5-9 years	10+ years	
	Years of intel experience (år med e-erfaring)	>1 year	2-4 years	5-9 years	10+ years	
	Educational level (Utdanningsnivå)	High School, eqv.		Bachelor	Master	PhD
	Service (Forsvarsgren)	Army	Navy	Air Force	Civilian	
Words of Estimative Probability (WEP)						
1	In general, I find it easy to ascribe a probability to my estimative intelligence assessments. (Jeg finner det enkelt å tillegge en sannsynlighet til mine e-vurderinger.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
2	Please rank the WEP in order of usage in your estimative assessments (1 = most used, 5 = least/rarely used):	WEP:			Rank:	
		Highly likely (Meget sannsynlig)				
		Likely (sannsynlig)				
	(Ranger sannsynlighetsordene fra mest brukt (1) til minst brukt (5) i dine e-vurderinger:)	Even Chance (Mulig)				
		Unlikely (Lite sannsynlig)				
		Highly unlikely (Svært lite sannsynlig)				
3	I think using structured analytic techniques (SATs) will make it easier to ascribe a probability to my estimative intelligence assessments. (Jeg tror det å bruke strukturerte analyseteknikker vil gjøre det enklere å tillegge en sannsynlighet til mine e-vurderinger.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
Analytic Confidence						
4	In general, I have high analytic confidence in my estimative intelligence assessments. (Jeg pleier å være sikker i mine e-vurderinger.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
5	Analytic confidence is a difficult concept. (Analytisk usikkerhet er et vanskelig konsept.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)

6	In general, I find it easy to state an analytic confidence in my estimative intelligence assessments. (Jeg synes det er enkelt å angi min grad av usikkerhet i mine e-vurderinger.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
7	I find it easy to see that WEPs and analytic confidence are two different concepts. (Det er enkelt å forstå at sannsynlighetsord og analytisk usikkerhet er to ulike konsepter.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
8	Properly distinguishing between underlying information, the assumptions, and assessments in the analysis process reduces analytic uncertainty. (Å skille tydelig mellom informasjon, antakelser og vurderinger i analyseprosessen minsker den analytiske usikkerheten.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
9	Credibility/quality of available information is an important factor when rating analytic confidence. (Informasjonens riktighet er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
10	Using a structured and traceable methodology in estimative intelligence analysis is an important factor when rating analytic confidence. (Å bruke en strukturert og sporbar analysemetode i utarbeidelse av e-vurderinger er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
11	Level of assumptions made in estimative intelligence assessments an important factor when rating analytic confidence. (Type/antall antakelser til grunn i en e-vurdering er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
12	How much assessments/ predictions might shift in response to new information is an important factor when rating analytic confidence. (Hvordan e-vurderingen står seg dersom vi får inn ny informasjon er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)

13	Level of analyst subject matter expertise is an important factor when rating analytic confidence. (Graden av fagkompetanse er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
14	Level of analyst agreement is an important factor when rating analytic confidence. (Samsvarhet mellom analytikere er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
15	The cost/difficulty of acquiring new information is an important factor when rating analytic confidence. (Kostnaden/vanskelighetsgraden i å innhente ny informasjon er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
16	I think the use of structured analytic techniques will help me in stating analytic confidence. (Jeg tror å bruke strukturerte analyseteknikker vil hjelpe meg i å uttrykke analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
17	Please rank the following factors as to their importance for deciding analytic confidence, from 1 being most important to 4 being least important.	Factor			Importance	
		The assessment's responsiveness to change in view of new information				
	(Ranger faktorene ut ifra graden av viktighet for å angi analytisk usikkerhet, der 1 er mest viktig og 4 minst viktig.)	The credibility/ quality of the information at hand				
		The cost of acquiring new information				
		The assessment's dependency on key assumptions				

Post-Course Survey: Structured Analytic Techniques						
	Age	20-25; 26-30; 31-35; 36-40; 41-45; 46-50; 50+				
	Gender	Male	Female			
	Unit	Armed Forces		Intel Service		Other
	Years of service	<1 year	2-4 years	5-9 years	10+ years	
	Years of intel experience	<1 year	2-4 years	5-9 years	10+ years	
	Educational level	High School, eqv.		Bachelor	Master	PhD
	Service	Army	Navy	Air Force	Civilian	
Words of Estimative Probability (WEP)						
1	Using the SATs-methodology made it easier to ascribe a probability to my assessments. (SAT-kurs metoden gjorde det enklere å tillegge en sannsynlighet til mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
2	Using the SATs-methodology made it easier to adjust my estimative intelligence assessments/ judgements (SAT-kurs metoden gjorde det enklere a justere mine e-vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
3	Using the SATs-methodology made it easier to use the whole WEP scale (SAT-kurs metoden gjorde det enklere å utnytte hele sannsynlighetsskalaen)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
4	ACH as a technique is helpful in ascribing a level of probability to my assessments (ACH-teknikken er nyttig for å tillegge en sannsynlighet til mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
5	Which other technique is helpful in ascribing a level of probability to your assessments? (Hvilken annen teknikk er til hjelp for å tillegge en sannsynlighet til mine e-vurderinger?)					
Analytic Confidence						
6	Using the SATs-methodology improved my analytic integrity (SAT-kurs metoden økte min analytiske integritet)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
7	Using the SATs-methodology improved my analytic objectivity (SAT-kurs metoden økte min analytiske objektivitet)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)

8	Using the SAT methodology made it easier to ascribe a level of analytic confidence to my assessments (SAT-kurs metoden gjorde det enklere å tillegge en en verdi til min analytiske usikkerhet)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
9	ACH as a technique is helpful in ascribing a level of confidence to my assessments. (ACH-teknikken er til hjelp for å tillegge en verdi på analytisk usikkerhet på en e-vurdering)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
10	Which other technique is helpful in ascribing a level of confidence to your assessments? (Hvilken annen teknikk er til hjelp for å tillegg en verdi til din analytiske usikkerhet i dine e-vurderinger?)					
11	I find it easy to see that WEPs and analytic confidence are two different concepts. (Det er enkelt å forstå at sannsynlighetsord og analytisk usikkerhet er to ulike konsepter)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
12	Properly distinguishing between underlying information, the assumptions, and assessments in the analysis process reduces analytic uncertainty. (Å skille tydelig mellom informasjon, antakelser og vurderinger i analyseprosessen minsker den analytiske usikkerheten.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
13	Credibility/quality of available information is an important factor when rating analytic confidence. (Informasjonens riktighet/kvalitet er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
14	Using a structured and traceable methodology in estimative intelligence analysis is an important factor when rating analytic confidence. (Å bruke en strukturert og sporbar analysemetode i utarbeidelse av e-vurderinger er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
15	Level of assumptions made in estimative intelligence assessments an important factor when rating analytic confidence. (Type/antall antakelser til grunn i en e-vurdering er en viktig	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)

	faktor når man skal vurdere analytisk usikkerhet.)					
16	How much assessments/predictions might shift in response to new information is an important factor when rating analytic confidence. (Hvordan e-vurderingen står seg dersom vi får inn ny informasjon er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
17	Level of analyst subject matter expertise is an important factor when rating analytic confidence. (Graden av fagkompetanse er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
18	Level of analyst agreement is an important factor when rating analytic confidence. (Samsvarhet mellom analytikere er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
19	The cost/difficulty of acquiring new information is an important factor when rating analytic confidence. (Kostnaden/vanskelighetsgraden i å innhente ny informasjon er en viktig faktor når man skal vurdere analytisk usikkerhet.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
20	The NORDIS SATs-methodology helps identify causes of uncertainties and helps identify indicators that can alter the level of uncertainty in analytic judgements. (FEHs SAT-metode hjelper en til å idenfisere årsaker til analytisk usikkerhet samt identifisere indikatorer som kan endre graden av denne usikkerheten.)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)
21	Please rank the following factors as to their importance for deciding analytic confidence, from 1 being most imporant to 4 being least important.	Factor			Importance	
		The assessment's responsiveness to change in view of new information				
		The credibility/ quality of the information at hand				
	(Ranger faktorene ut i fra graden av viktighet for å angi analytisk usikkerhet, der 1 er mest viktig og 4 minst viktig.)	The cost of acquiring new information				
		The assessment's dependency on key assumptions				

Appendix 3: Survey 2 questionnaire

Follow-up survey: Structured Analytic Techniques							
	Age (alder):	20-25; 26-30; 31-35; 36-40; 41-45; 46-50; 50+					
	Gender (Kjønn):	Male	Female				
	Unit (avdeling):	Armed Forces		Intel Service		Other	
	Years of service (År i tjeneste):	<1 year	2-4 years	5-9 years	10+ years		
	Years of intel experience (år med e-erfaring):	<1 year	2-4 years	5-9 years	10+ years		
	Educational level (Utdanningsnivå):	High School, eqv		Bachelor	Master	PhD	
	Service (Forsvarsgren):	Army	Navy	Air Force	Civilian		
	Time since SATs-course (Når gjennomført SAT-kurs)	0-6 months	7-12 months	1-2 years	3-5 years	5+ years	Not attended
1	To what extent do you use the NORDIS Structured Analytic Techniques (SATs) model on a regular basis in your work? (I hvor stor grad bruker du FEHs modell for strukturerte analyseteknikker (SAT) i din daglige jobb?)	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
2	To what extent do you use single Structured Analytic Techniques (SATs) on a regular basis in your work? (I hvor stor grad bruker du enkeltvise strukturerte analyseteknikker (SAT) i din daglige jobb?)	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
3	To what extent do you use the following techniques? (I hvilken grad bruker du følgende teknikker?)						
	ACH	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
	Key Assumptions Check	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
	Alternative futures/scenario cross	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
	SWOT/TOWS	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
	Operational design/ Course of Action development	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
	Other techniques (Andre teknikker):						

4	What are the reasons for not using SATs? Mark the one(s) appropriate to you. (Hva er grunnen til at du eventuelt ikke bruker SAT?) (Marker alle som gjelder for deg)						
	Time demanding (Tidkrevende)						
	Difficult (Vanskelig)						
	Not part of organizational culture (Ikke en del av organisasjonskulturen)						
	Not relevant (Ikke relevant)						
	Other methods are better (Andre metoder er bedre)						
5	To what extent do you use standardized verbal probabilities in your intelligence estimates? (I hvilken grad bruker du standardiserte sannsynlighetsord i dine etterretningsvurderinger?)	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
6	Using SATs makes it easier to ascribe a probability to my assessments. (Bruk av SAT gjør det enklere å tillegge en sannsynlighet til mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
7	Using SATs makes it easier to adjust my estimative intelligence assessments. (Bruk av SAT gjør det enklere a justere estimatene i mine e-vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
8	Using ACH is makes it easier to ascribe a level of probability to my assessments. (Bruk av ACH gjør det enklere å tillegge en sannsynlighet til mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
9	Using ACH makes it easier to adjust my estimative intelligence assessments. (Bruk av ACH gjør det enklere a justere estimatene i mine e-vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
10	Using Key Assumptions Check makes it easier to ascribe a level of probability to my assessments. (Bruk av Key Assumptions Check gjør det enklere å tillegge en sannsynlighet til mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant

11	Using Key Assumptions Check makes it easier to adjust my estimative intelligence assessments. (Bruk av Key Assumptions Check gjør det enklere a justere estimatene i mine e-vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
12	To what extent are you explicitly stating your analytic confidence in your intelligence estimates? (I hvilken grad uttrykker du analytisk konfidens i dine etterretningsvurderinger?)	Almost always (5)	Often (4)	Sometimes (3)	Seldom (2)	Never (1)	Not relevant
13	Using SATs improves my analytical confidence. (Bruk av SAT øker min analytiske konfidens)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
14	Using SATs makes it easier to ascribe a level of analytic confidence to my estimates (low-med.-hi) (Bruk av SAT gjør det enklere å nivåsette min analytiske usikkerhet i mine vurderinger (lav-med.-høy))	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
15	Using ACH helpful in ascribing a level of analytic confidence to my estimates. (Bruk av ACH er til hjelp for å beskrive et nivå på analytisk konfidens på mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
16	Using Key Assumptions Check (KAC) helpful in ascribing a level of analytic confidence to my estimates. (Bruk av KAC er til hjelp for å beskrive et nivå på analytisk konfidens på mine vurderinger)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
17	Verbal probabilities expressions (WEP) and analytic confidence are two different concepts. (Verbale sannsynlighetsord og analytisk konfidens er to ulike konsepter)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
18	Using SATs improves my analytic integrity. (Bruk av SAT bedrer min analytiske integritet)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant
19	Using SATs improves my analytic objectivity. (Bruk av SAT bedrer min analytiske objektivitet)	Agree (5)	Somewhat agree (4)	Neither agree or disagree (3)	Somewhat disagree (2)	Disagree (1)	Not relevant

		Factor	Importance
20	Please rank the following factors as to their importance for deciding analytic confidence, from 1 being most important to 4 being least important.	The assessment's responsiveness to change in view of new information	
		The credibility/ quality of the information at hand	
	(Ranger faktorene ut i fra graden av viktighet for å angi analytisk usikkerhet, der 1 er mest viktig og 4 minst viktig.)	The cost of acquiring new information	
		The assessment's dependency on key assumptions	

Appendix 4: Interview questions

Age:
Gender:
Service:
Unit:
Years of service:
Years of intelligence experience:
Educational level:

1. When did you take the NORDIS SATs-course?
2. How much do you now use what you learned at the SATs-course in your job as an analyst?
3. Do you follow the school's comprehensive SATs-model in your work?
4. Are there any techniques you use more than others?
5. If you use SATs, how does it impact on your analytic integrity?
6. If you use SATs, how does it impact on your analytic objectivity?
7. How does using SATs impact on your willingness to adjust your assessments/judgements?
8. How important is it to assess and manage uncertainty related to estimative intelligence products rather than to reduce the uncertainty?
9. If you do not use SATs, what are the main reasons?
10. What is your understanding of the usage of words of estimative probability and analytic confidence?
11. Do you find that using the school's comprehensive SATs-methodology helps you to ascribe a probability to your estimative intelligence assessments?
12. Which technique do you find most useful when ascribing a probability to your estimative intelligence assessments?
13. Do you find that using the school's comprehensive SATs-methodology helps you ascribing a level of analytic confidence to your estimative intelligence assessments?
14. Which technique do you find most useful when ascribing a level of analytic confidence to your estimative intelligence assessments?
15. Are you explicitly stating your analytic confidence in your estimative intelligence products?
16. Which factor is most useful in helping you to ascribe your analytic confidence in your estimative intelligence products? (More than one factor?)
17. In what way do you think working with different scenarios/possible future outcomes affect your analytic confidence?

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.770
Bartlett's Test of Sphericity	Approx. Chi-Square	605.697
	df	136
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	4.735	27.850	27.850	4.735	27.850	27.850	3.876
2	1.938	11.399	39.249	1.938	11.399	39.249	1.906
3	1.531	9.006	48.255	1.531	9.006	48.255	1.646
4	1.337	7.863	56.118	1.337	7.863	56.118	2.151
5	1.111	6.533	62.651	1.111	6.533	62.651	2.963
6	.971	5.711	68.362				
7	.815	4.797	73.158				
8	.697	4.100	77.259				
9	.683	4.019	81.278				
10	.559	3.287	84.565				
11	.510	3.000	87.565				
12	.467	2.745	90.310				
13	.407	2.397	92.707				
14	.365	2.145	94.852				
15	.343	2.020	96.872				
16	.278	1.636	98.508				
17	.254	1.492	100.000				

Pattern Matrix[a]

	Component				
	1	2	3	4	5
Post-course ACH helpful in ascribing a level of analytic confidence	.852				
Post-course SATs made ascribing a level of analytic confidence easier	.757				
Post-course ACH helpful in ascribing a level of probability	.732				
Post-course SATs made ascribing probability easier	.538				
Post-course SATs made adjusting assessments easier	.501				
Post-course SATs made using the whole WEP-scale easier	.500				
Post-course SATs help adjusting level of uncertainty	.446		.445		
Pre-course Easiness of ascribing probability		.794			
Pre-course High analytic confidence in assessments		.755			
Pre-course Easiness of stating analytic confidence		.731			
Pre-course Assumption level important factor for analytic confidence			.716		
Post-course Assumption level important factor for analytic confidence			.677		
Post-course SATs improved analytic objectivity				.807	
Post-course SATs improved analytic integrity				.749	
Pre-course SATs will make it easier to ascribe probability					.816
Pre-course SATs will help state analytic confidence					.805
Pre-course Structured methodology important factor for analytic confidence					.533

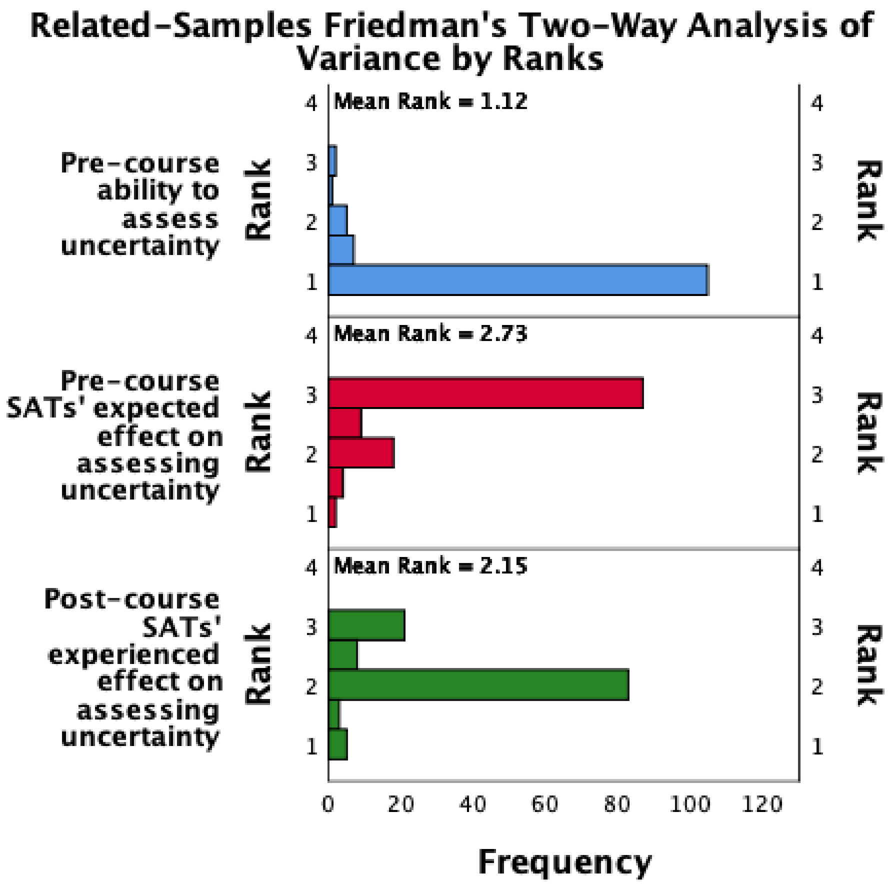
Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.[a]
a. Rotation converged in 9 iterations.

Appendix 6: Survey 1 The effect of using SATs on analysts' ability to assess uncertainty

A: Assessing uncertainty repeated measures

	Pre-course ability to assess uncertainty	Pre-course SATs' expected effect on assessing uncertainty	Post-course SATs' experienced effect on assessing uncertainty
Median	3.0000	4.6667	4.2857
Mean	2.9750	4.5931	4.2471
Std. Deviation	.82351	.55374	.56274
N	120	120	120

Total N	120
Test Statistic	167.257
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000



Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Pre-course ability to assess uncertainty-Post-course SATs' experienced effect on assessing uncertainty	-1.037	.129	-8.036	.000	.000
Pre-course ability to assess uncertainty-Pre-course SATs' expected effect on assessing uncertainty	-1.612	.129	-12.490	.000	.000
Post-course SATs' experienced effect on assessing uncertainty-Pre-course SATs' expected effect on assessing uncertainty	.575	.129	4.454	.000	.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

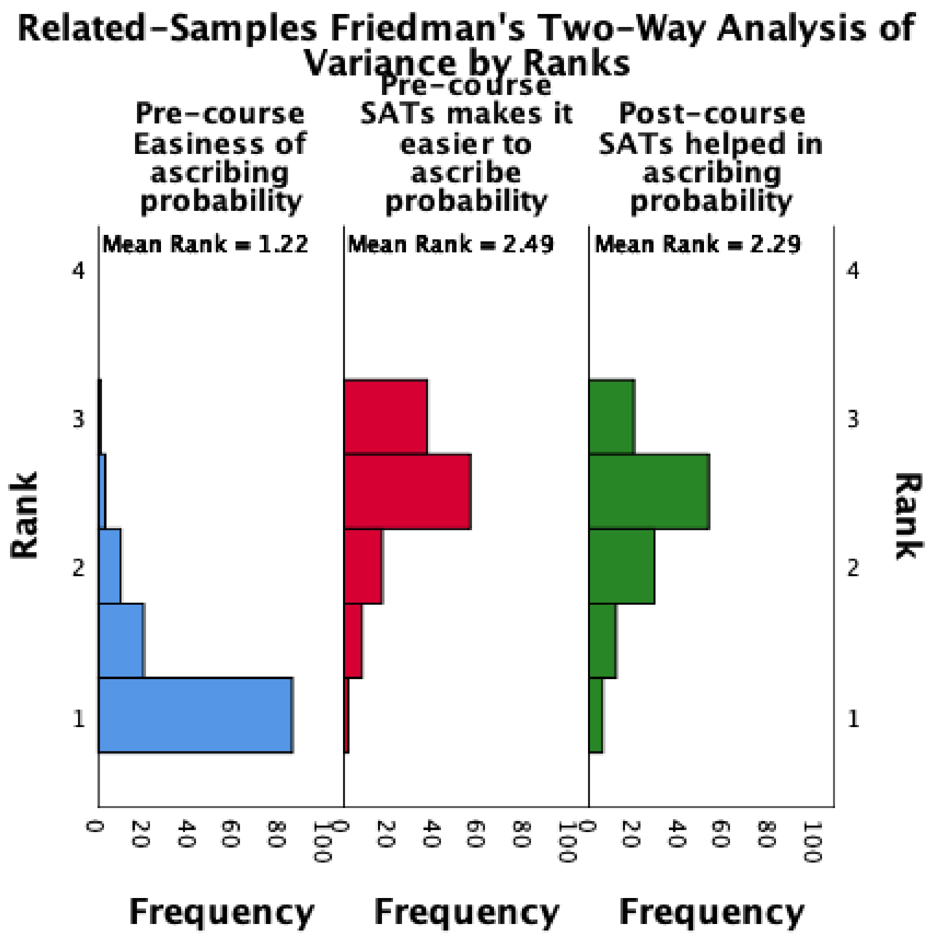
Related-Samples Kendall's Coefficient of Concordance Summary

Total N	120
Kendall's W	.697
Test Statistic	167.257
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000

B: Probability repeated measures

	Pre-course Easiness of ascribing probability	Pre-course SATs makes it easier to ascribe probability	Post-course SATs helped in ascribing probability
Median	3.00	5.00	4.50
N	120	120	120

Total N	120
Test Statistic	141.195
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000



Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Pre-course Easiness of ascribing probability-Post-course SATs helped in ascribing probability	-1.067	.129	-8.262	.000	.000
Pre-course Easiness of ascribing probability-Pre-course SATs makes it easier to ascribe probability	-1.271	.129	-9.844	.000	.000
Post-course SATs helped in ascribing probability-Pre-course SATs makes it easier to ascribe probability	.204	.129	1.581	.114	.341

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

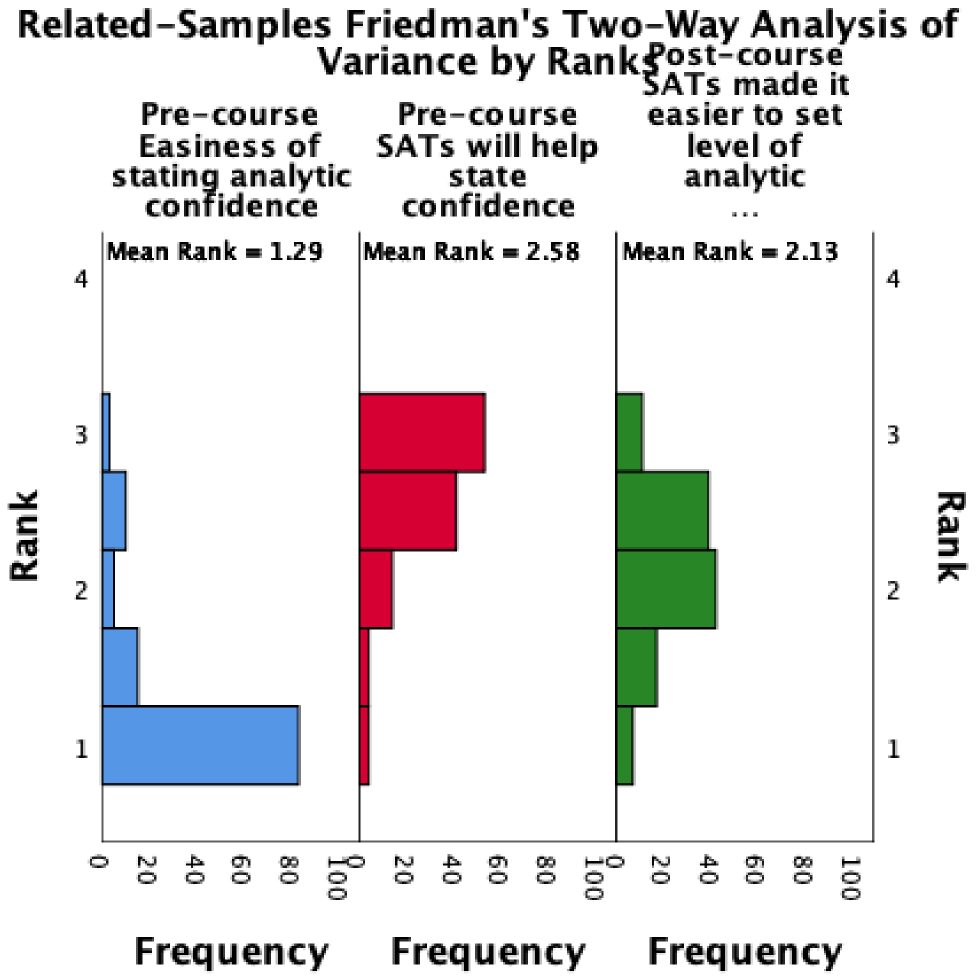
Related-Samples Kendall's Coefficient of Concordance Summary

Total N	120
Kendall's W	.588
Test Statistic	141.195
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000

C: Analytic confidence repeated measures

	Pre-course Easiness of stating analytic confidence	Pre-course SATs will help state confidence	Post-course SATs made it easier to set level of analytic confidence
Median	3.00	5.00	4.00
N	120	119	117

Total N	116
Test Statistic	120.390
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000



Pairwise Comparisons

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
Pre-course Easiness of stating analytic confidence-Post-course SATs made it easier to set level of analytic confidence	-.841	.131	-6.401	.000	.000
Pre-course Easiness of stating analytic confidence-Pre-course SATs will help state confidence	-1.293	.131	-9.848	.000	.000
Post-course SATs made it easier to set level of analytic confidence-Pre-course SATs will help state confidence	.453	.131	3.447	.001	.002

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Related-Samples Kendall's Coefficient of Concordance Summary

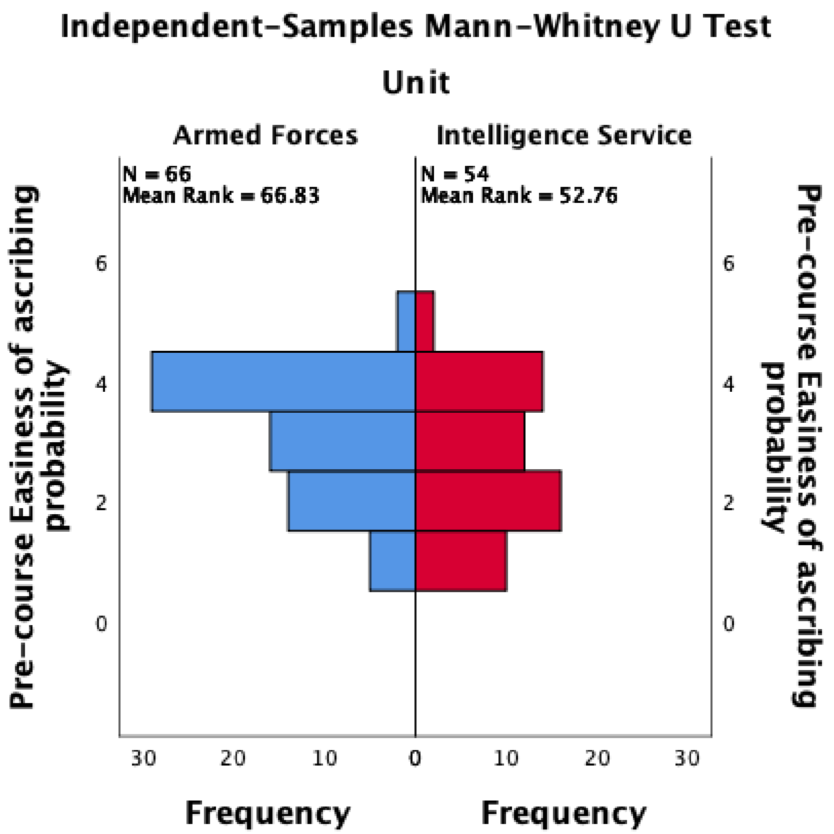
Total N	116
Kendall's W	.519
Test Statistic	120.390
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000

D: Easiness of ascribing probability across Unit

Pre-course

Unit	Median	N
Armed Forces	3.00	66
Intelligence Service	3.00	54
Total	3.00	120

Total N	120
Mann-Whitney U	1364.000
Wilcoxon W	2849.000
Test Statistic	1364.000
Standard Error	182.198
Standardized Test Statistic	-2.294
Asymptotic Sig.(2-sided test)	.022



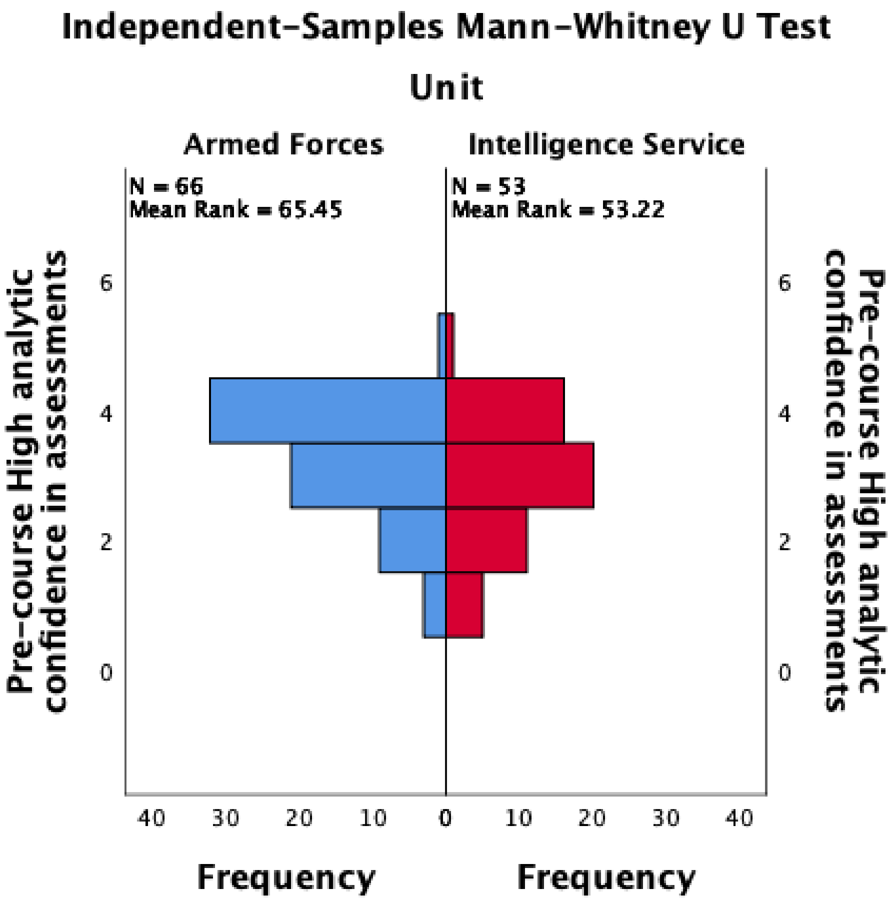
Post-course

Total N	120
Mann-Whitney U	1611.500
Wilcoxon W	3096.500
Test Statistic	1611.500
Standard Error	169.362
Standardized Test Statistic	-1.007
Asymptotic Sig.(2-sided test)	.314

E: Analytic confidence across unit

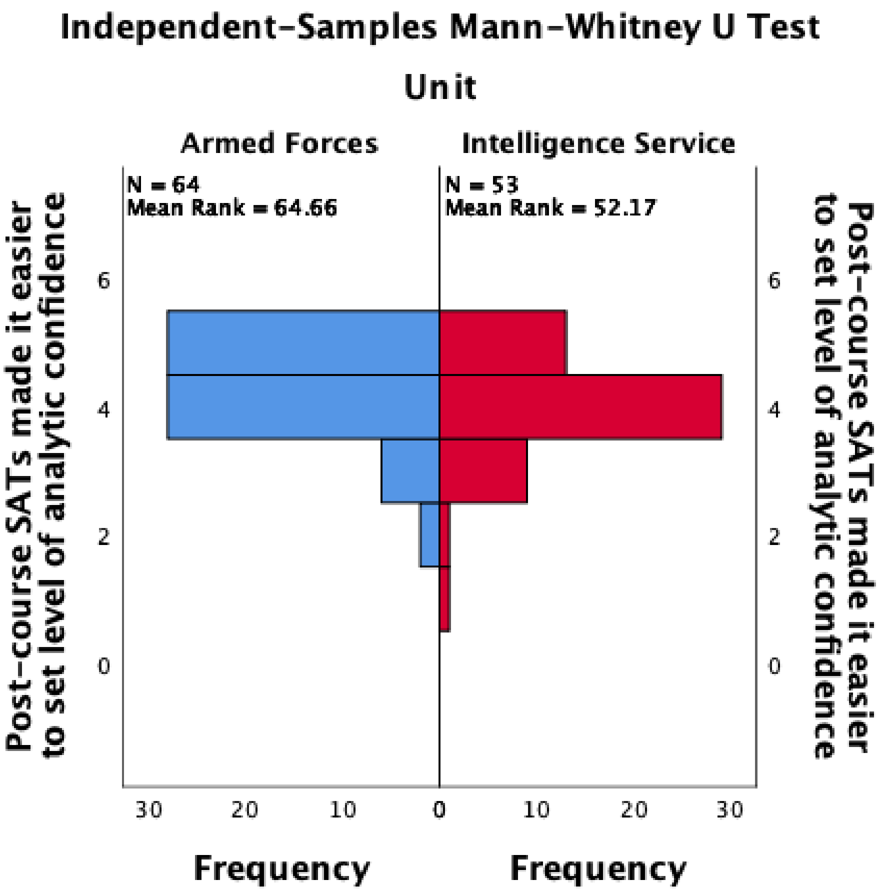
Pre-course High analytic confidence in assessments across Unit

Total N	119
Mann-Whitney U	1389.500
Wilcoxon W	2820.500
Test Statistic	1389.500
Standard Error	176.293
Standardized Test Statistic	-2.039
Asymptotic Sig.(2-sided test)	.041



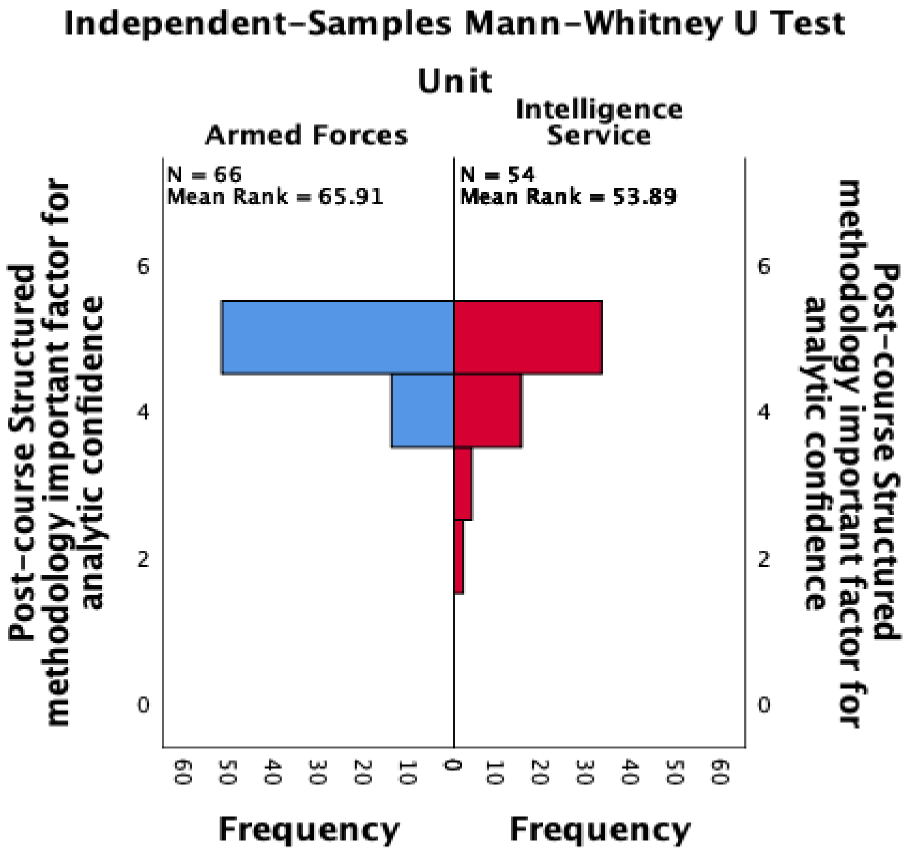
Post-course SATs made it easier to set level of analytic confidence across Unit

Total N	117
Mann-Whitney U	1334.000
Wilcoxon W	2765.000
Test Statistic	1334.000
Standard Error	167.313
Standardized Test Statistic	-2.164
Asymptotic Sig.(2-sided test)	.030



Post-course Structured methodology important factor for analytic confidence across Unit

Total N	120
Mann-Whitney U	1425.000
Wilcoxon W	2910.000
Test Statistic	1425.000
Standard Error	150.526
Standardized Test Statistic	-2.372
Asymptotic Sig.(2-sided test)	.018



Unit		Pre-course High analytic confidence in assessments	Post-course SATs made it easier to set level of analytic confidence	Post-course Structured methodology important factor for analytic confidence
Armed Forces	Median	3.50	4.00	5.00
	N	66	64	66
Intelligence Service	Median	3.00	4.00	5.00
	N	53	53	54
Total	Median	3.00	4.00	5.00
	N	119	117	120

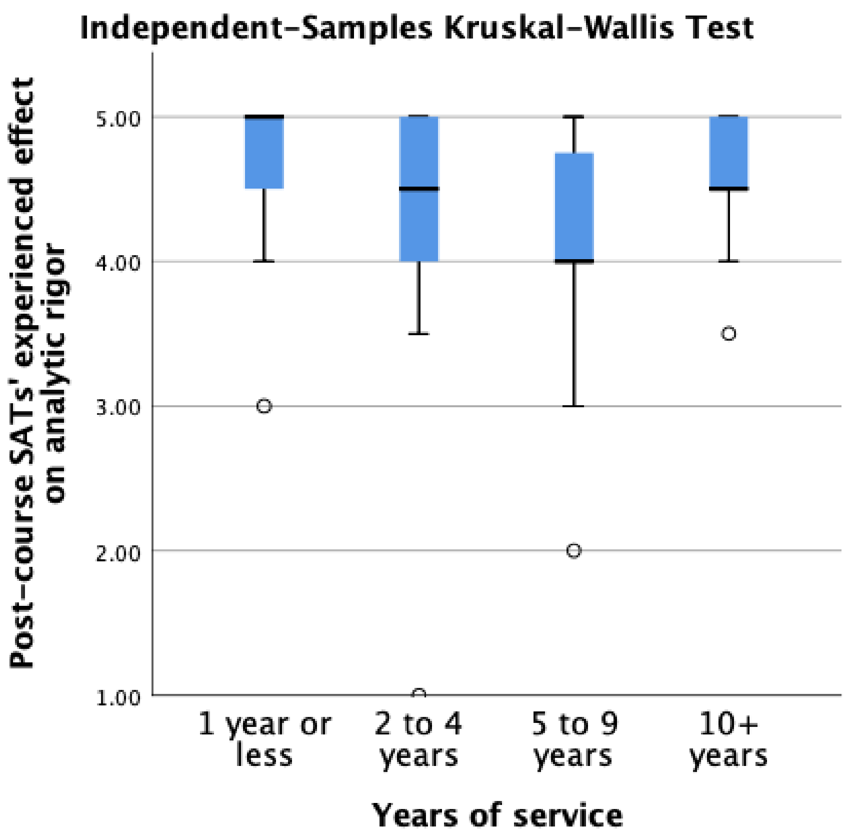
Appendix 7: Survey 1 SATs' effect on analytic rigor

Years of service	Median	N
1 year or less	5.0000	19
2 to 4 years	4.5000	42
5 to 9 years	4.0000	19
10+ years	4.5000	38
Total	4.5000	118

Independent-Samples Kruskal-Wallis Test Summary

Total N	118
Test Statistic	9.777[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.021

a. The test statistic is adjusted for ties.

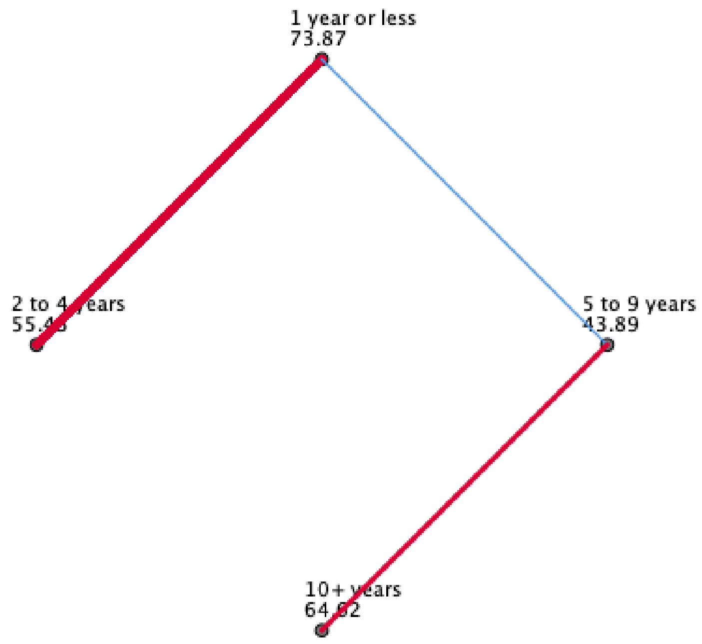


Pairwise Comparisons of Years of service

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
5 to 9 years-2 to 4 years	11.534	8.948	1.289	.197	1.000
5 to 9 years-10+ years	-20.724	9.094	-2.279	.023	.136
5 to 9 years-1 year or less	29.974	10.501	2.854	.004	.026
2 to 4 years-10+ years	-9.190	7.246	-1.268	.205	1.000
2 to 4 years-1 year or less	18.440	8.948	2.061	.039	.236
10+ years-1 year or less	9.250	9.094	1.017	.309	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Pairwise Comparisons of Years of service



Each node shows the sample average rank of Years of service.

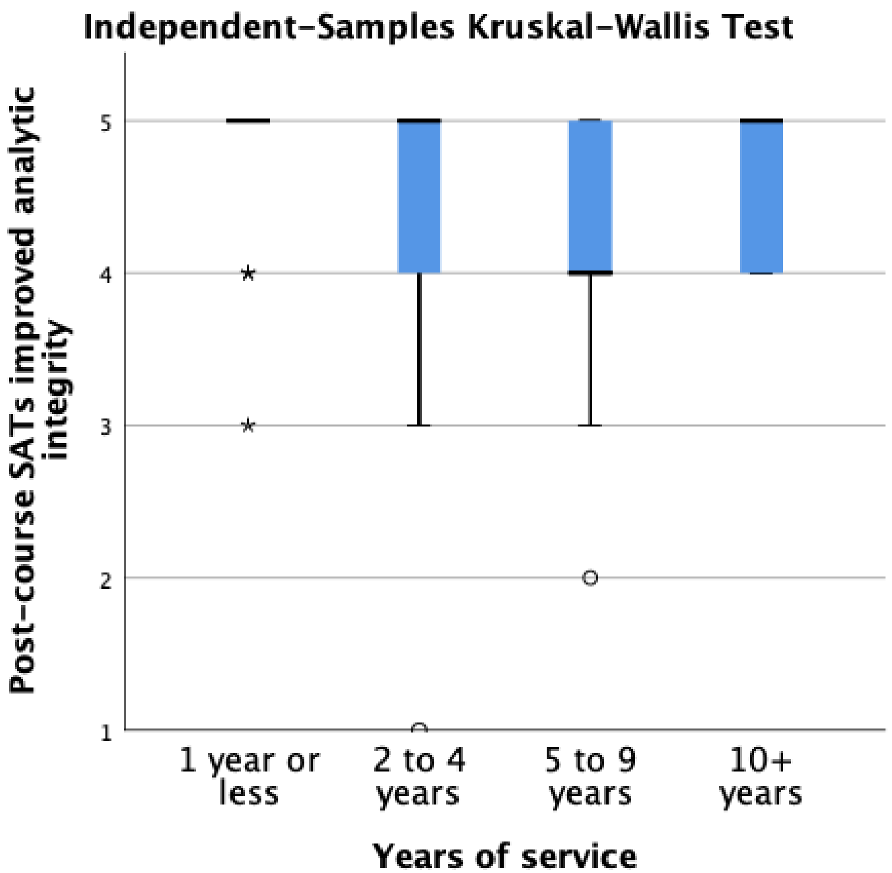
SATs improved analytic integrity across Years of service

Years of service	Median	N
1 year or less	5.00	19
2 to 4 years	5.00	42
5 to 9 years	4.00	19
10+ years	5.00	38
Total	5.00	118

Independent-Samples Kruskal-Wallis Test Summary

Total N	118
Test Statistic	9.883[a]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	.020

a. The test statistic is adjusted for ties.

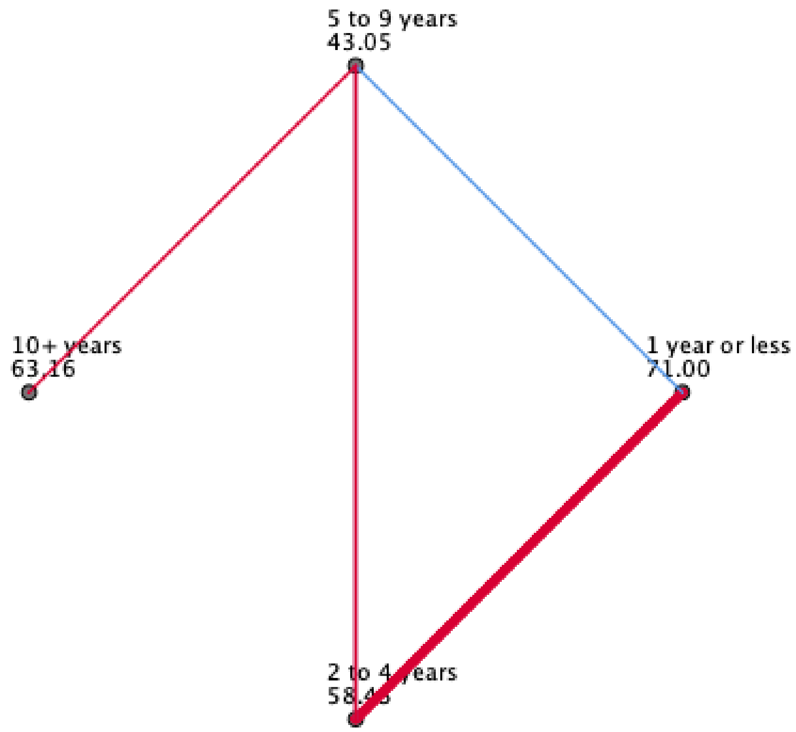


Pairwise Comparisons of Years of service

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
5 to 9 years-2 to 4 years	15.376	7.969	1.930	.054	.322
5 to 9 years-10+ years	-20.105	8.098	-2.483	.013	.078
5 to 9 years-1 year or less	27.947	9.351	2.989	.003	.017
2 to 4 years-10+ years	-4.729	6.453	-.733	.464	1.000
2 to 4 years-1 year or less	12.571	7.969	1.578	.115	.688
10+ years-1 year or less	7.842	8.098	.968	.333	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.
a. Significance values have been adjusted by the Bonferroni correction for multiple tests.

Pairwise Comparisons of Years of service



Each node shows the sample average rank of Years of service.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.865
Bartlett's Test of Sphericity	Approx. Chi-Square	1012.390
	df	136
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	7.880	46.356	46.356	7.880	46.356	46.356	7.140
2	2.581	15.182	61.537	2.581	15.182	61.537	5.459
3	1.262	7.423	68.961				
4	1.018	5.986	74.947				
5	.920	5.412	80.358				
6	.622	3.658	84.017				
7	.448	2.634	86.651				
8	.406	2.390	89.041				
9	.340	2.001	91.042				
10	.281	1.653	92.694				
11	.257	1.511	94.205				
12	.234	1.377	95.582				
13	.197	1.157	96.739				
14	.178	1.050	97.789				
15	.149	.878	98.667				
16	.123	.726	99.393				
17	.103	.607	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

Pattern Matrix[a]

	Component 1	Component 2
SATs make it easier to ascribe analytic confidence	.872	
SATs make it easier to adjust assessments	.831	
ACH helpful in ascribing analytic confidence	.812	
SATs improve analytic confidence	.811	
SATs make is easier to ascribe probability	.809	
KAC helpful in ascribing analytic confidence	.785	
ACH makes it easier to adjust assessments	.782	
ACH makes it easier to ascribe probability	.683	
KAC makes it easier to adjust assessments	.683	
KAC makes is easier to ascribe probability	.658	
Use of Key Assumptions Check		.860
Use of Operational Design/Course of Action development		.787
Use of SWOT/TOWS		.778
Use of single SATs		.756
Use of ACH		.753
Use of NORDIS SATs-methodology		.748
Use of Alternative futures/scenario cross		.602

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.[a]
a. Rotation converged in 4 iterations.

Appendix 9: Survey 2 The effect of using SATs on analysts' ability to assess uncertainty

A: SATs' effect on assessing uncertainty across units

Independent-Samples Kruskal-Wallis

Total N	90
Test Statistic	5.918[a,b]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.052

Oneway

Descriptives

SATs' effect on assessing uncertainty

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
IntBn	28	4.2589	.48896	.09241	4.0693	4.4485	3.20	5.00
NIS	45	3.7969	.89840	.13393	3.5269	4.0668	1.00	5.00
PST	17	4.0235	.68331	.16573	3.6722	4.3749	2.20	5.00
Total	90	3.9834	.77284	.08146	3.8216	4.1453	1.00	5.00

ANOVA

SATs' effect on assessing uncertainty

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.719	2	1.860	3.272	.043
Within Groups	49.439	87	.568		
Total	53.158	89			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: SATs' effect on assessing uncertainty

	(I) Unit	(J) Unit	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Bonferroni	IntBn	NIS	.46208*	.18145	.038	.0191	.9050
		PST	.23540	.23178	.938	-.3304	.8012
	NIS	IntBn	-.46208*	.18145	.038	-.9050	-.0191
		PST	-.22668	.21461	.881	-.7506	.2972
	PST	IntBn	-.23540	.23178	.938	-.8012	.3304
		NIS	.22668	.21461	.881	-.2972	.7506
Hochberg	IntBn	NIS	.46208*	.18145	.037	.0206	.9035
		PST	.23540	.23178	.672	-.3285	.7993
	NIS	IntBn	-.46208*	.18145	.037	-.9035	-.0206
		PST	-.22668	.21461	.645	-.7488	.2954
	PST	IntBn	-.23540	.23178	.672	-.7993	.3285
		NIS	.22668	.21461	.645	-.2954	.7488
Games-Howell	IntBn	NIS	.46208*	.16271	.016	.0725	.8517
		PST	.23540	.18975	.441	-.2361	.7069
	NIS	IntBn	-.46208*	.16271	.016	-.8517	-.0725
		PST	-.22668	.21308	.542	-.7464	.2931
	PST	IntBn	-.23540	.18975	.441	-.7069	.2361
		NIS	.22668	.21308	.542	-.2931	.7464

*. The mean difference is significant at the 0.05 level.

B: The use of standardised verbal probabilities across unit

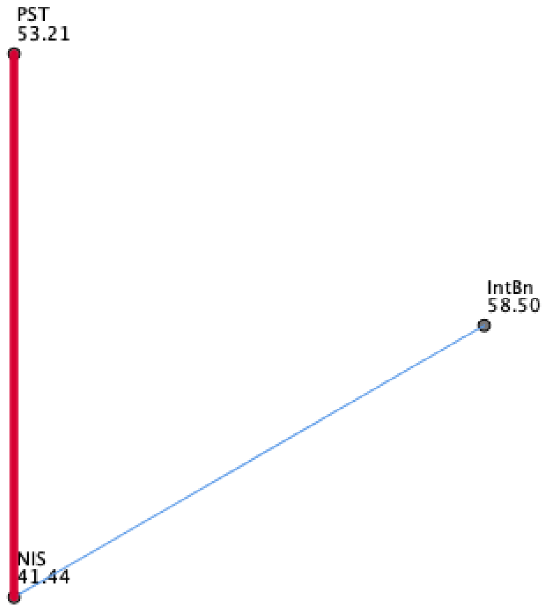
Independent-Samples Kruskal-Wallis

Total N	96
Test Statistic	14.714[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.001

Pairwise Comparisons of Unit

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
NIS-PST	-11.765	5.521	-2.131	.033	.099
NIS-IntBn	17.059	4.636	3.679	.000	.001
PST-IntBn	5.294	6.061	.873	.382	1.000

Pairwise Comparisons of Unit



Each node shows the sample average rank of Unit.

Unit	N	Median
IntBn	28	5.00
NIS	51	5.00
PST	17	5.00
Total	96	5.00

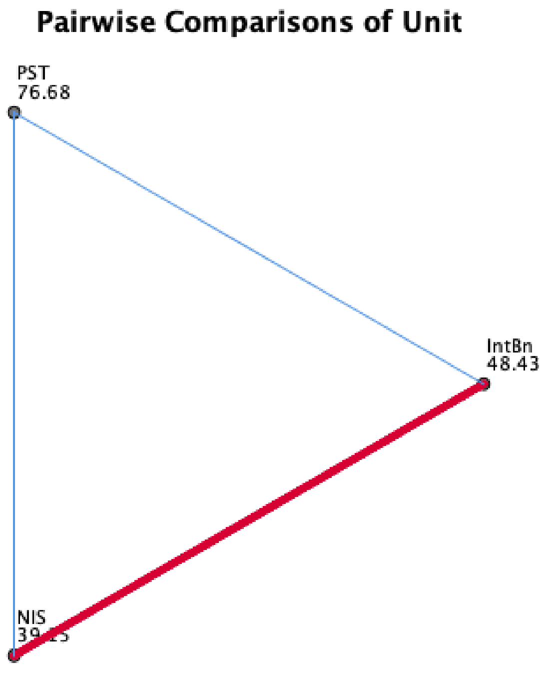
C: Degree of explicitly stating analytic confidence in assessments across unit

Independent-Samples Kruskal-Wallis

Total N	96
Test Statistic	24.646[a]
Degree Of Freedom	2
Asymptotic Sig.(2-sided test)	.000

Pairwise Comparisons of Unit

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.[a]
NIS-IntBn	9.282	6.349	1.462	.144	.431
NIS-PST	-37.529	7.560	-4.964	.000	.000
IntBn-PST	-28.248	8.300	-3.403	.001	.002



Each node shows the sample average rank of Unit.

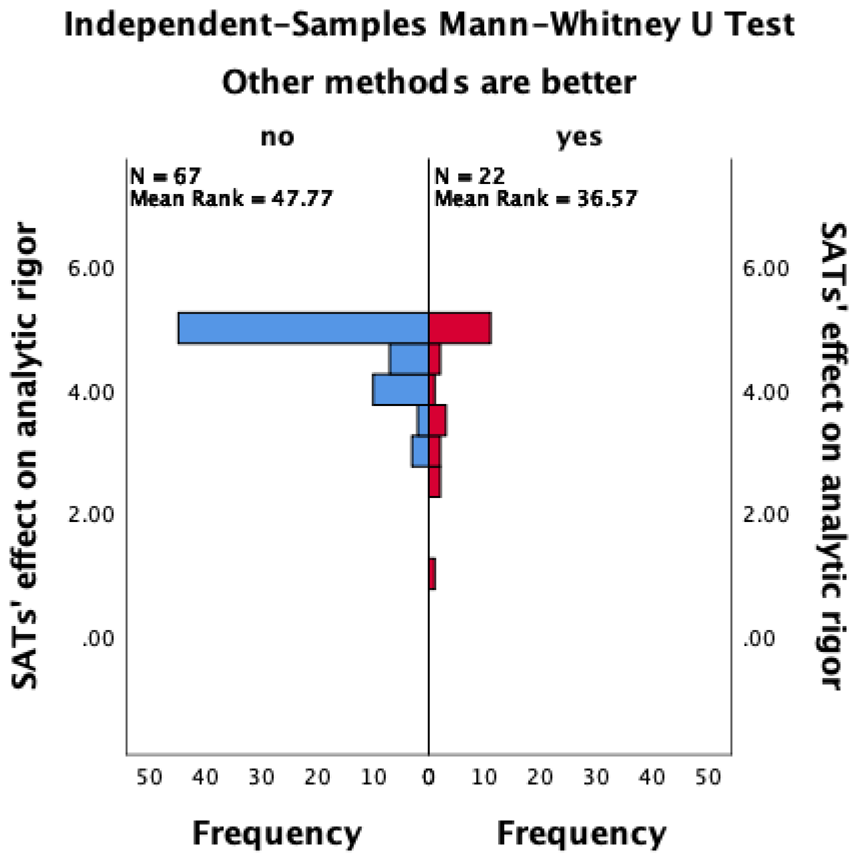
Unit	N	Median
IntBn	28	2.00
NIS	51	2.00
PST	17	4.00
Total	96	2.00

Appendix 10: Survey 2 SATs' effect on analytic rigour

A: SATs' effect on analytic rigour across other methods are better

Independent-Samples Mann-Whitney U

Total N	89
Mann-Whitney U	922.500
Wilcoxon W	3200.500
Test Statistic	922.500
Standard Error	90.916
Standardized Test Statistic	2.040
Asymptotic Sig.(2-sided test)	.041

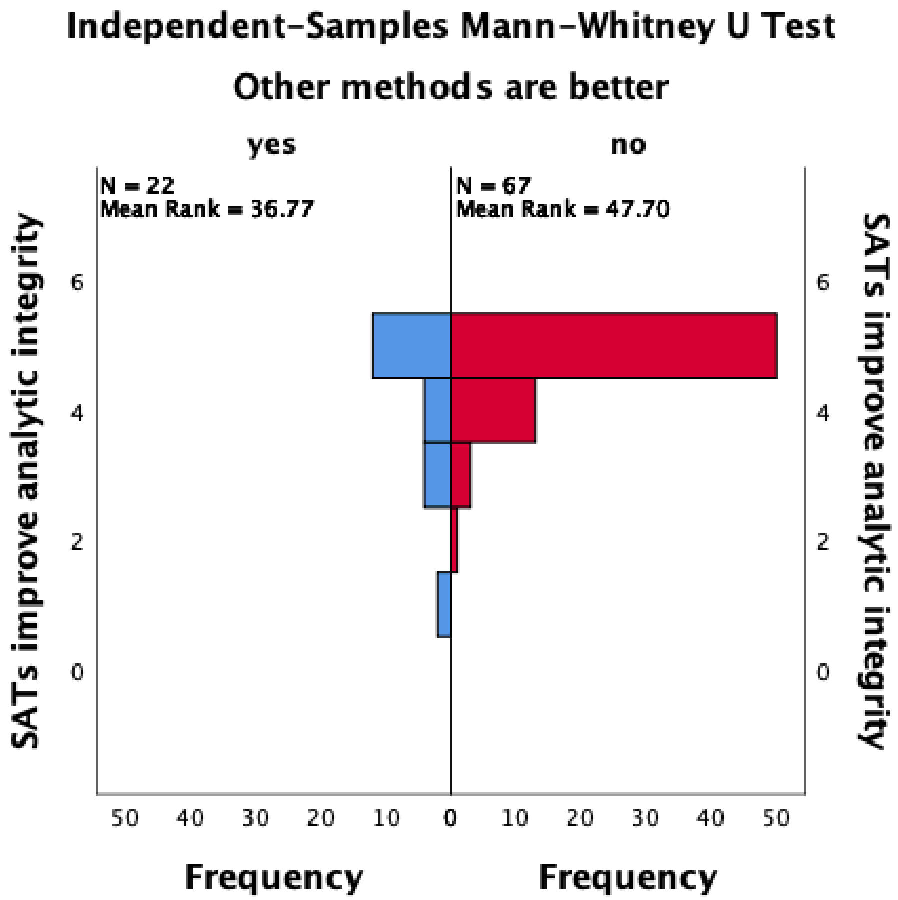


Other methods are better	Median	N
yes	4.7500	22
no	5.0000	67
Total	5.0000	89

B: SATs improve analytic integrity across other methods are better

Independent-Samples Mann-Whitney U

Total N	89
Mann-Whitney U	918.000
Wilcoxon W	3196.000
Test Statistic	918.000
Standard Error	85.065
Standardized Test Statistic	2.128
Asymptotic Sig.(2-sided test)	.033



Other methods are better	Median	N
yes	5.00	22
no	5.00	67
Total	5.00	89

Appendix 11: Survey 2 SATs usage

A: SATs usage across categories unit and personnel

Descriptives

SATs usage

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
IntBn	28	3.4753	.69433	.13122	3.2061	3.7446	1.83	4.71
NIS	51	2.0756	.73254	.10258	1.8696	2.2817	1.00	3.57
PST	17	2.7899	.47142	.11434	2.5475	3.0323	2.00	3.71
Total	96	2.6104	.91469	.09336	2.4250	2.7957	1.00	4.71

ANOVA

SATs usage

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	36.080	2	18.040	38.655	.000
Within Groups	43.403	93	.467		
Total	79.483	95			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: SATs-usage

	(I) Unit	(J) Unit	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Hochberg	IntBn	NIS	1.39971*	.16068	.000	1.0093	1.7902
		PST	.68542*	.21005	.005	.1750	1.1958
	NIS	IntBn	-1.39971*	.16068	.000	-1.7902	-1.0093
		PST	-.71429*	.19132	.001	-1.1792	-.2494
	PST	IntBn	-.68542*	.21005	.005	-1.1958	-.1750
		NIS	.71429*	.19132	.001	.2494	1.1792

*. The mean difference is significant at the 0.05 level.

T-Test SATs usage across personnel category

	Personnel category	N	Mean	Std. Deviation	Std. Error Mean
SATs usage	Armed Forces	57	2.8200	.96292	.12754
	Civilian	39	2.3040	.75035	.12015

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
SATs usage	Equal variances assumed	2.177	.143	2.811	94	.006	.51594	.18353
	Equal variances not assumed			2.944	92.333	.004	.51594	.17522

B: Reasons for not using SATs across unit

	No culture for SATs-usage	N	Mean	Std. Deviation	Std. Error Mean
SATs-usage	yes	37	2.3012	.76694	.12608
	no	59	2.8043	.95188	.12392

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
SATs-usage	Equal variances assumed	1.229	.270	-2.709	94	.008	-.50312	.18572
	Equal variances not assumed			-2.846	88.112	.006	-.50312	.17679

Unit * No culture for SATs-usage Crosstabulation

			No culture for SATs-usage		
			yes	no	Total
Unit	IntBn	Count	3	25	28
		% within Unit	10.7%	89.3%	100.0%
		Adjusted Residual	-3.6	3.6	
	NIS	Count	25	26	51
		% within Unit	49.0%	51.0%	100.0%
		Adjusted Residual	2.2	-2.2	
	PST	Count	9	8	17
		% within Unit	52.9%	47.1%	100.0%
		Adjusted Residual	1.3	-1.3	
Total		Count	37	59	96
		% within Unit	38.5%	61.5%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	13.005	2	.001
Likelihood Ratio	14.740	2	.001
Linear-by-Linear Association	10.015	1	.002
N of Valid Cases	96		

Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.368	.001
	Cramer's V	.368	.001
N of Valid Cases		96	

Unit * SATs not relevant Crosstabulation

			SATs not relevant		
			yes	no	Total
Unit	IntBn	Count	10	18	28
		% within Unit	35.7%	64.3%	100.0%
		Adjusted Residual	-.6	.6	
	NIS	Count	27	24	51
		% within Unit	52.9%	47.1%	100.0%
		Adjusted Residual	2.6	-2.6	
	PST	Count	2	15	17
		% within Unit	11.8%	88.2%	100.0%
		Adjusted Residual	-2.7	2.7	
Total		Count	39	57	96
		% within Unit	40.6%	59.4%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.357	2	.009
Likelihood Ratio	10.351	2	.006
Linear-by-Linear Association	1.170	1	.279
N of Valid Cases	96		

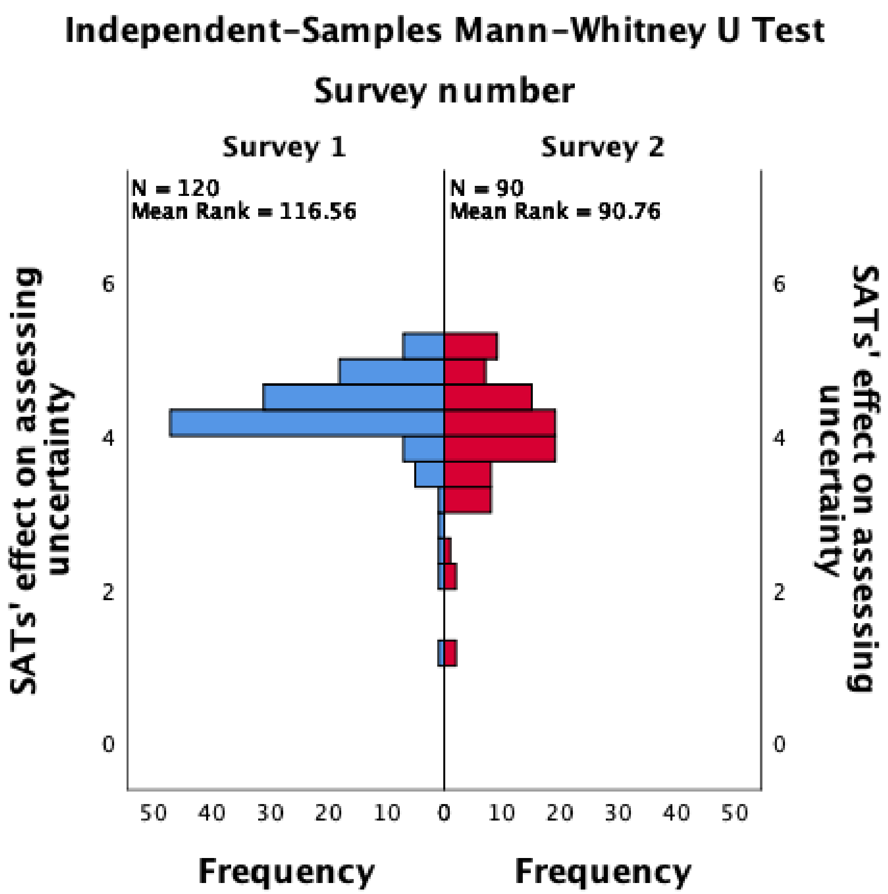
Symmetric Measures

		Value	Approximate Significance
Nominal by Nominal	Phi	.312	.009
	Cramer's V	.312	.009
N of Valid Cases		96	

Appendix 12: SATs' effect on assessing uncertainty across surveys

Independent-Samples Mann-Whitney U

Total N	210
Mann-Whitney U	4073.000
Wilcoxon W	8168.000
Test Statistic	4073.000
Standard Error	434.943
Standardized Test Statistic	-3.051
Asymptotic Sig.(2-sided test)	.002



SATs' effect on assessing uncertainty

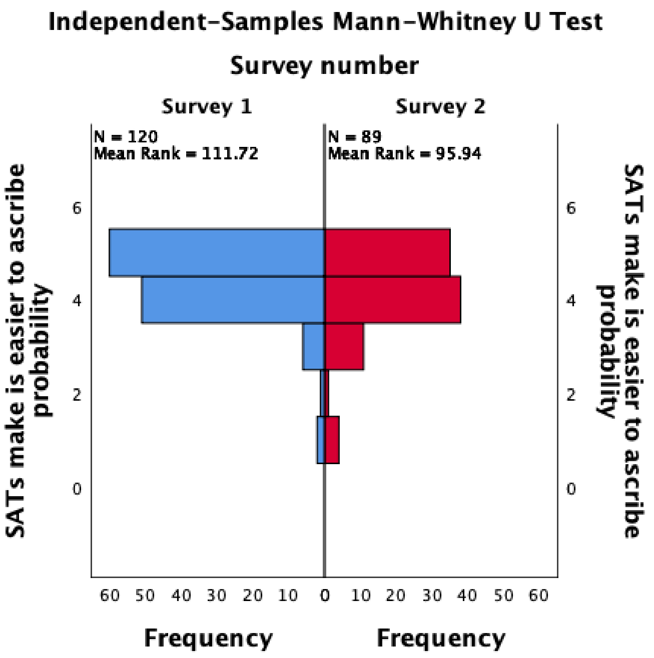
Survey number	N	Median	Mean	Std. Deviation
Survey 1	120	4.29	4.25	.563
Survey 2	90	4.00	3.98	.773
Total	210	4.18	4.13	.672

SATs make is easier to ascribe probability across surveys

Survey number		SATs make is easier to ascribe probability	SATs make it easier to ascribe analytic confidence
Survey 1	N	120	117
	Median	4.50	4.00
	Mean	4.38	4.15
	Std. Deviation	.769	.802
Survey 2	N	89	83
	Median	4.00	4.00
	Mean	4.11	3.98
	Std. Deviation	.982	.937
Total	N	209	200
	Median	4.00	4.00
	Mean	4.27	4.07
	Std. Deviation	.874	.862

Independent-Samples Mann-Whitney U

Total N	209
Mann-Whitney U	4533.500
Wilcoxon W	8538.500
Test Statistic	4533.500
Standard Error	393.463
Standardized Test Statistic	-2.050
Asymptotic Sig.(2-sided test)	.040



Appendix 13: SATs' effect on analytic rigour across surveys

Independent-Samples Mann-Whitney U

Total N	207
Mann-Whitney U	6067.500
Wilcoxon W	10072.500
Test Statistic	6067.500
Standard Error	396.244
Standardized Test Statistic	2.061
Asymptotic Sig.(2-sided test)	.039

T-Test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
SATs' effect on analytic rigour	Equal variances assumed	6.861	.009	-.508	205	.612	-.04928	.09701
	Equal variances not assumed			-.492	163.898	.623	-.04928	.10014

SATs' effect on analytic rigour

Survey number	N	Median	Mean	Std. Deviation
Survey 1	118	4.5000	4.4788	.61808
Survey 2	89	5.0000	4.5281	.77738
Total	207	4.5000	4.5000	.68973

Appendix 14: List of interviews

Interview A: Male, NCO, senior analyst and SATs instructor, IntBn

Interview B: Male, NCO, analyst, IntBn

Interview C: Male, officer, senior analyst, IntBn

Interview D: Male, civilian, analyst, NIS

Interview E: Male, officer, analyst, NIS

Interview F: Male, officer, senior analyst, NIS

Interview G: Male, civilian, senior analyst and SATs instructor, PST

Interview H: Female, senior SATs instructor, NLD

Interview I: Male, senior SATs instructor, NLD

Interview J: Female, senior staff officer, UK DI

Interview K: Male, senior SATs instructor, UK DI