

Analytic Rigour Literature Excerpts

About this document

To support participants in the Expert Panel on Analytic Rigour, the Hunt Lab has been compiling this selection of useful or interesting excerpts from the relevant literature.

Entries are organised under the three main topics in the Analytic Rigour project:

1. What is the nature of analytic rigour?
2. What are the factors that impact upon it?
3. What are the opportunities for enhancing analytic rigour?

The document is a **work-in-progress**.

Literature review

This document is related to the literature review on analytic rigour being conducted by the Hunt Lab as part of its Analytic Rigour project. Over 800 works were surfaced through searches of relevant databases, which were then filtered down to about 280 works deemed relevant via 2-3 independent applications of exclusion criteria. We are currently in the process of synthesising the findings from this reduced set of relevant works. If you are interested, the full list can be found in this [spreadsheet](#).

Updates

We will be updating this document during the Expert Panel process. You can access the latest version from the portal page.

Comments and suggestions

We welcome any suggestions for additions or other improvements. Just email us at hunt-lab@unimelb.edu.au.

The Hunt Laboratory for Intelligence Research

University of Melbourne

Parkville VIC 3010 Australia

huntlab.science.unimelb.edu.au

Contents

1	The Nature of Analytic Rigour	4
1.1	What analytic rigour is	4
1.1.1	Avoiding shallow analysis	4
1.1.2	Adherence to tradecraft standards	5
1.2	What Analytic Rigour is Not	6
1.3	Rigour and communication	6
2	Factors Impacting Analytic Rigour	8
2.1	Cognitive factors	8
2.1.1	Information overload	8
2.1.2	Inherent difficulties of analytic thinking	8
2.1.3	Pressure of the job	8
2.1.4	Tension between rigour and insight	9
2.2	Methodological factors	9
2.2.1	Not enough use of SATs	9
2.2.2	Too much use of SATs	10
2.2.3	Inappropriate use of SATs	11
2.2.4	Inadequate methods of assessing sources	11
2.3	Organisational factors	13
2.3.1	Lack of evaluation	13
2.3.2	Tolerance of poor analysis	13
2.3.3	Not enough focus on standards	14
2.3.4	Too much focus on standards	14
2.3.5	Not enough of a shared understanding of standards or analytic rigour	14
2.3.6	Consistency of emphasis on quality of analysis	16
2.3.7	Organisational and bureaucratic pressure and misguided emphasis	16
2.3.8	Communication of reasoning, rigour, and tradecraft	17
2.3.9	Tradecraft advocates	17
2.3.10	Ineffective training	18
2.3.11	Objectivity and subjectivity	18
2.3.12	Multiple sources of bias	19
2.4	Technological factors	20
2.4.1	Poor tools	20
3	Opportunities to Enhance Analytic Rigour	21
3.1	Standards	21
3.1.1	Improve adherence to standards	21
3.2	Methods	21
3.2.1	Introduce numeric estimative language	21
3.2.2	Discuss rigour more in briefings	22
3.2.3	Better guidance on the evaluation of evidence and sources	23
3.2.4	Adopt a sensemaking approach to analysis	24
3.2.5	Make analytic reasoning more explicit	24
3.3	Evaluation	25
3.3.1	Better evaluation of analytic products and rigour	25
3.3.2	Measure analytic rigour	26
3.3.3	Measure accuracy of assessments	26



3.3.4	Adopt elements of rigour from the social sciences (and other related disciplines)	27
3.3.5	Evaluate SATs	28
3.4	Training.....	28
3.4.1	Better training	28
3.5	Recruitment.....	29
3.5.1	Recruit differently.....	29

1 The Nature of Analytic Rigour

1.1 What analytic rigour is

Zelik et al. did some of the most important work on analytic rigour. Here they compare two types of definitions of analytic rigour.

Zelik, D. J., Patterson, E. S. and Woods, D. D. (2010) 'Measuring Attributes of Rigor in Analysis', in MacroCognition Metrics and Scenarios: Design and Evaluation for Real-World Teams.

“Across domains, conventional perspectives for measuring analytical rigor focus on identifying gaps between what was actually done versus a prescribed or “standard” method...

Unfortunately, such definitions suggest a conceptualization of analytical activity that is neither particularly likely to reflect rigorous analysis work as practiced...

Consequently, rather than on a standards-based notion of rigor, our measurement approach focuses on how the risk of shallow analysis is reduced via analyst-initiated strategies that are opportunistically employed throughout the analysis process. These strategies are alternatively conceptualized as “broadening” checks (Elm et al., 2005) insofar as they tend to slow the production of analytic product and make explicit the sacrifice of efficiency in pursuit of accuracy, a central tenet of the framework”. (p.66)

Zelik, D., Patterson, E. S. and Woods, D. D. (2007) 'Understanding Rigor in Information Analysis', in. Proceedings of the Eighth International NDM Conference.

“In information analysis, judgement of rigor reflects a relationship in the appropriateness of fit between analytic processes and contextual requirements. Thus, as supported by this and other research, rigor is more meaningfully viewed as an assessment of degree of sufficiency, rather than degree of adherence to an established analytic procedure.” (p.1)

1.1.1 Avoiding shallow analysis

Zelik et al. 2010 discuss analytic rigour as avoiding shallow analysis:

“Of particular concern is the potential inherent in all analytical activity that an analysis process is prematurely concluded and is subsequently of inadequate depth relative to the demands of a situation—a vulnerability we characterize as the risk of shallow analysis.

This chapter examines rigor as a measure of sensemaking in information analysis activity that captures how well an analytical process reduces this risk. It describes eight attributes of analytical rigor that were synthesized from empirical studies with professional intelligence analysts from a variety of specialties.” (p.65)

The eight attributes they list are:

- Hypothesis Exploration
- Information Search
- Information Validation
- Stance Analysis
- Sensitivity Analysis
- Specialist Collaboration

- Information Synthesis
- Explanation Critique

A recent effort to characterize analytic rigour comes from the Laboratory for Analytic Sciences.

Johnston, J. (2020). Defining Analytic Rigor for Analysis in the Intelligence Community [Unpublished report]. Laboratory for Analytic Sciences, North Carolina State University.

“What does the analyst need to have to be able to confidently state that intelligence conclusions have been rigorously determined? And, how does a decision-maker determine that they can rely on those analytic conclusions? Before answering these questions, we need to identify something that does not clearly exist in the Intelligence Community (IC) today, a useable *definition of analytic rigor...* (p.4)

For the present study, the conceptual framework is a candidate operational definition of analytic rigor. This definition was determined through discussions with seasoned analysts, review of related professional policy, and subject matter experts’ contributions. The underlined terms in the definition are those intended to be operationalized through further study and research. These terms represent measurable actions or processes, with the intent that the findings of this study and future studies will determine specifically how they will be measured... (p.6)

Conceptual Framework: Candidate Operational Definition of Analytic Rigor

Rigor is an effort by an analyst or researcher to be as complete as possible in order to arrive at the most accurate assessment/results possible in conducting an analysis with integrity. This is achieved by employing methods and techniques meant to support a variety of indicators of sufficiency. Indicators of sufficiency include:

- Objectivity
- Thoroughness
- Replicability, reliability, validity
- Transparency (in analysis and analytic decision-making)
- Credibility
- Relevance” (p.7)

1.1.2 Adherence to tradecraft standards

Rigour is sometimes defined as (or largely as) adherence to tradecraft standards

Rojas, J. T. (2019) ‘Masters of Analytical Tradecraft: Certifying the Standards and Analytic Rigor of Intelligence Products’.

“The term analytic rigor is used to convey the level of analytic tradecraft application, which is sometimes robust and other times weak.” (p.1)

One of the most influential set of analytic standards is found in ICD 203.

Intelligence Community Directive (ICD) 203, Analytic Standards (2015). Washington, DC: Office of the Director of National Intelligence.

Analytic tradecraft standards:

1. Proper description of sources, data and methodologies;
2. Proper expression of uncertainties;

3. Proper distinguishing of information from assumptions and judgements;
4. Incorporating analysis of alternatives;
5. Demonstrating customer relevance and addressing implications;
6. Use of clear and logical argumentation;
7. Explanation of change to or consistency of judgements with previous reports;
8. Making accurate judgements or assessments;
9. Incorporating effective visual information where appropriate.

Conversely, the UK PHIA identifies being rigorous as one of the 8 high level common standards intelligence should meet.

Professional Development Framework for all-source intelligence assessment (2019), London, U.K. Professional Head of Intelligence Assessment (PHIA)

"Rigorous: Analysts should use processes, methods, tools and techniques appropriate to the intelligence requirement in order to be able to show logical and coherent reasoning upon which the resulting judgements are based. Analysts should identify and systematically evaluate differing hypotheses, especially when judgements contain significant levels of uncertainty or complexity (such as forecasting future trends), or when low probability outcomes would have high impact results. This activity should be recorded in a discoverable format for the audit trail." (p. 28)

1.2 What Analytic Rigour is Not

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. (2005). Report to the President.

Under the heading of "Lack of rigorous analysis" the WMD commission report claims that:

"The scope and quality of analysis has eroded badly in the Intelligence Community and it must be restored. In part, this is a matter of tradecraft and training; in part, too, it is a matter of expertise. Analytic "tradecraft"—the way analysts think, research, evaluate evidence, write, and communicate—must be strengthened. In many instances, we found finished intelligence that was loosely reasoned, ill-supported, and poorly communicated. Perhaps most worrisome, we found too many analytic products that obscured how little the Intelligence Community actually knew about an issue and how much their conclusions rested on inference and assumptions. We believe these tendencies must be reversed if decision makers are to have confidence in the intelligence they receive. And equally important, analysts must be willing to admit what they don't know in order to focus future collection efforts." (p.12)

1.3 Rigour and communication

Rigour is often claimed to be important for transparency, accountability and communication.

Rojas, J. T. (2019) 'Masters of Analytical Tradecraft: Certifying the Standards and Analytic Rigor of Intelligence Products'.

"Decision makers should be provided "with actionable data backed by analytical rigor." When evaluating a problem set, an intelligence consumer's perceived analytical rigor (based on perceived cues) ideally should meet or exceed effective rigor (based on analytical process). Thus, decision makers should not be left to their own perceptions in judging the analytic rigor of a given intelligence product. They need to be

provided fidelity on the level of analytical rigor—or the robustness of the analytical process used in creating the product—so that they can meter their trust accordingly.” (p.1)

Zelik, D., Patterson, E. S. and Woods, D. D. (2007) ‘Understanding Rigor in Information Analysis’, in. Proceedings of the Eighth International NDM Conference.

“...poor representation hides and obscures the process being represented. Thus, rigor represents a mechanism for revealing the observability of an analysis process... This finding is extended through the acknowledgment that all forms of analytic products, to some extent, obscure the processes that produced them—a perspective that guides our exploration of rigor in information analysis.” (p.2)

Corkill, J. (2008) ‘Evaluation a critical point on the path to intelligence’, Journal of the Australian Institute of Professional Intelligence Officers., 16(1).

“Given intelligence products are generally based on incomplete data sets there always exists a significant margin for error. Poor evaluation has potential to significantly compound that margin. It may therefore be argued that evaluation is the critical juncture from which either good or poor quality intelligence emerges.” (p.3)

Marchio, J. (2014) ‘Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis’, *Intelligence and National Security*, 29(2).

“Lastly, this review reminds us that the critical thinking and the rigor underlying IC analysis – the community’s analytic tradecraft – remain its ‘center of gravity’. IC consumers in the wake of the Iraq WMD failure have rightly demanded to see more of what lies behind analytic judgments – be it the quality of the sources used, the number and nature of intelligence gaps surrounding the issue, or the underlying assumptions. While getting a judgment ‘right’ is what ultimately matters most, the recipients of IC analytic products recognize that strong analytic tradecraft is more likely to result in assessments that are relevant and rigorous – what they need and value most. Past as well as the future national security challenges confronting IC consumers suggest the need to employ strong analytic tradecraft will endure.” (p.182)

2 Factors Impacting Analytic Rigour

2.1 Cognitive factors

2.1.1 Information overload

Zelik, D., Patterson, E. S. and Woods, D. D. (2007) 'Understanding Rigor in Information Analysis', in. Proceedings of the Eighth International NDM Conference.

“In the face of significant production pressures... and rapidly proliferating data availability...—and resulting data overload deluging the professional analyst—it is increasingly easy for analysts to be trapped by shallow, low-rigor analysis. Given similar pressures, it is also increasingly difficult for decision makers to recognize when an analysis is not of sufficient rigor for a given decision.” (p.2)

2.1.2 Inherent difficulties of analytic thinking

Hoffman, R. et al. (2011) 'Reasoning difficulty in analytical activity', Theoretical Issues in Ergonomics Science, 12(3)

“What makes analysis difficult? It can certainly be said that analytical work is difficult because it is a form of critical thinking, and critical thinking is difficult ... Research and theory on critical thinking in the field of psychology highlights the fact that critical thinking is difficult (for people in general), the fact that critical thinking is not put to use nearly as often as it should be (by people in general), and the fact that it links strongly to curiosity, a motive that varies considerably in strength from person to person.(p.227)... Humans have a remarkable ability to learn to perceive meaningful patterns, and this can be disrupted and inhibited by the application of reductive approaches. The attempts to impose quantification of complex variables and patterns – typically seen as the only method for enhancing rigor – often end in a frustrating exercise of 'jury-rig-oring', where final numerical answers are nothing more than the aggregation of make-shift calculations that mask other judgments and caveats.” (p.233)

Hedley, J. H. (2005) 'Learning from Intelligence Failures', International Journal of Intelligence and CounterIntelligence, 18(3)

“...the Intelligence Community's failure to predict the Soviet invasion of Afghanistan soon brought new controversy. The IC had, in fact, accurately estimated the advantages and disadvantages of intervention. But, once again, the IC, having assessed that the disadvantages outweighed the advantages, concluded that the Soviets would act rationally, in accordance with U.S. perceptions of Soviet self-interest. The CIA's Douglas MacEachin, who was directly involved in the assessments, recalls that one of the dark humor jokes circulating around the Agency after the invasion was that the analysts had gotten it right, and the Soviets got it wrong.” (p.442)

2.1.3 Pressure of the job

Hoffman, R. et al. (2011) 'Reasoning difficulty in analytical activity', Theoretical Issues in Ergonomics Science, 12(3)

“The cognitive difficulty of critical thinking and sensemaking are amplified by the broader context of the work, that is, the ways in which jobs are specified. Researchers who have studied intelligence analysis and related domains of information synthesis and analysis (e.g. emergency response), show a clear

consensus that analysis is difficult because of a cluster of factors including workload, time pressure, high stakes and high uncertainty...

On top of all this, intelligence work often focuses on events that transpire quickly, and so analysis leaves little or no time for introspection or reflection on the processes of intelligence collection and production..."(p.228)

2.1.4 Tension between rigour and insight

Klein, G. (2011) 'Critical thoughts about critical thinking', *Theoretical Issues in Ergonomics Science*, 12(3)

"...performance depends on increasing insights and on reducing mistakes. Both of these activities are important. Analysts who make accurate speculations but also become known for uncritical acceptance of flawed data, or for making erroneous arguments, will lose their credibility. On the other hand, a purely defensive strategy that documents all assumptions and evaluates all data sources but fails to notice connections and implications is not very helpful either. Avoiding mistakes is not the same as gaining insights. Increased amounts of rigor do not result in valuable discoveries. Yet senior intelligence officials admit that new analysts are often told that they need not worry about making incorrect judgements as long as they followed procedures and documented their sources, identified assumptions and estimated uncertainty values. The products that analysts generate are evaluated for failure to adhere to standards; no one tracks whether the intelligence products contained valuable insights...(p.211)

One possibility is that reducing mistakes will also increase insights. I am not aware of any evidence that this is the case. A second possibility is that they are independent and unrelated. The most disturbing possibility is that the actions taken to reduce mistakes can interfere with insights. In discussions with many experienced Intelligence Analysts, this third possibility seems to be the most likely, and it worries them. One highly regarded analyst explained that 'In almost all useful products I've seen or done, there was some violation of tradecraft'. In the IC, the concept of 'tradecraft' is often used as a catch-all term for systematic analytic techniques. In some uses, it is used as a synonym for 'critical thinking', emphasising valid arguments, careful vetting of sources, tracking of assumptions, ensuring logical conclusions and so forth. Tradecraft emphasises the down arrow, the reduction of mistakes. Reducing mistakes is critical and important, but the way it is implemented may pose difficulties." (p. 213)

2.2 Methodological factors

2.2.1 Not enough use of SATs

Rojas, J. T. (2019) 'Masters of Analytical Tradecraft: Certifying the Standards and Analytic Rigor of Intelligence Products'.

"Advanced analysis brings a high level of analytical rigor and applied tradecraft to fulfill sufficiency to the above-stated measurables. To achieve rigor, both critical thinking and SATs should be applied. Structured analytic techniques are used to challenge assessments, identify cognitive bias, stimulate creativity, and measure uncertainty. SATs include mental network analysis, structured brainstorming, analysis of competing hypothesis, red hat analysis, pre-mortem analysis, devil's advocacy, and structured debate.³⁷ These techniques fall into broad categories like visualization, decomposition, challenge analysis, idea generation, and hypothesis generation/testing. SATs are a process driven approach that when combined with intuition, critical thinking, and subject matter expertise reduce analytical error and promote quality, effective, and rigorous analysis." (p.7)

2.2.2 Too much use of SATs

Chang, W. et al. (2018) 'Restructuring structured analytic techniques in intelligence', *Intelligence and National Security*, 33(3)

Chang et al. questions the claim that "SATs instill rigor and make analysts' thought processes transparent, thus more logically sound".

"Two core reasons why SATs are unlikely to deliver on promised benefits: bias bipolarity and noise neglect.

Unfortunately, little is known about whether structured analytic techniques improve, have no effect on or even degrade analysis because there is scant scientific research on their effectiveness. We propose that the root problems with SATs derive from the failure to deploy best practices in coping with the two fundamental sources of error that bedevil all efforts to improve human judgment: systematic bias and random noise in the processes by which people generate, test and update their hunches about the world. Specifically, SATs fail to address (a) the inherently bipolar nature of cognitive biases and the omnipresent risk that well-intentioned attempts to reduce one bias, say, over-confidence, will amplify its opposing bias, under-confidence; (b) the cumulative nature of error in multi-stage assessments, in which the noise in the conclusions from the one stage gets fed into the next—and the risk that well-intentioned efforts to reduce noise by decomposing complex judgments into sequences of simpler ones will make those judgments less, not more, consistent—an oversight we call noise neglect.

Both problems stem from the lack of sustained efforts to subject SATs to scientific tests of efficacy and scrutinizing the processes for logical validity. The net result is that it is difficult to know when SATs are sparing us from serious mistakes, when they are creating more problems than they are solving, or when they are just ineffective... (p.340)

Neglecting noise: Do SATs reduce noise in judgments?

In addition to slighting bias bipolarity, we believe that certain SATs unnecessarily introduce noise into an already noisy process, causing inconsistent judgments to become even more inconsistent (in scientific parlance, unreliable). Reliability of assessments is a necessary but not sufficient condition of their validity. Many intelligence analysts are highly skilled, attaining expertise through years of experience and specialized training. If nothing has changed, we should expect an expert analyst examining the same evidence using the same SATs to reach the same judgments at different times—that is, to be both internally consistent and temporally stable. However, several types of SATs rely on decomposing and separately addressing the components of a problem (e.g., evidence, hypotheses, conclusions).

Unfortunately there is a lot of subjectivity in the decomposition process, in part, due to the ambiguity of the SAT-prescribed processes. This sets the stage for an assumed strength of SATs—proceduralizing thinking processes as an alternative to 'mere intuition'—to impair intelligence analysis.

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. The inconsistent handling of the decomposed parts—hypotheses, evidence, and hypothesis-evidence linkages—is especially worrisome." (p.344)

2.2.3 Inappropriate use of SATs.

Chang, W. and Tetlock, P. E. (2016) 'Rethinking the training of intelligence analysts', *Intelligence and National Security*, 31(6)

"Current training is anchored in a mid-twentieth century understanding of psychology that focuses on checking over-confidence and rigidity but ignores the problems of under-confidence and excessive volatility." (p.903)

Table 1. Balancing opposing errors.

Current training emphasizes avoiding...	Symptom	But de-emphasizes...	Symptom
Over-confidence	Exaggerating the strength of a conclusion (e.g. Iraq WMD)	Under-confidence	Minimizing the strength of evidence towards a conclusion (e.g. the rise of ISIL)
Under-stating chances of change	Minimizing status quo changes (e.g. Russian ground incursion into Eastern Europe)	Over-stating chances of change	Exaggerating status quo changes (e.g. collapse of North Korea after a leader's death)
Overweighting the consensus argument	Blindly following the group (e.g. optimism about democratization post-Arab Spring)	Underweighting the consensus argument	Adopting a contrarian stance pro forma (e.g. routine warnings of a major North Korean attack)

(p.905)

"The confusion deepens when we get to the penultimate chapter of Psychology of Intelligence Analysis, which discusses 'base-rates', a term for the relative frequency of events in a reference class. The question is, in judging the risk of a coup in an African country, how much weight should be given to how often coups occur in that region versus country-specific conditions? The guidance from research on base-rate neglect is that analysts should over-weight base-rates in estimating likelihoods. The net effect is to make analysts much more cautious about predicting near-term changes to the status quo.

But this advice contradicts the warning against over-relying on mental models that blind us to looming change. Which bias – belief-perseverance or base-rate neglect – should analysts worry about more? Tension also exists between warnings against neglecting base rates and over-relying on precedents. Base-rates are simply sets of historical precedents....

In summary, traditional training fails to confront how biases can either reinforce or neutralize each other – and thus also misses the need to fashion prescriptions for managing these tensions." (p.907)

2.2.4 Inadequate methods of assessing sources

Corkill, J. (2008) 'Evaluation a critical point on the path to intelligence', *Journal of the Australian Institute of Professional Intelligence Officers.*, 16(1).

"Those analysts whose thoughts and opinions have contributed to this paper recognise the need to evaluate sources and information however formal processes of evaluation are rarely implemented. It is surprising just how many analysts appear to base their evaluation of information and sources on their immediate intuition and how often information is not tested if it appears to conform to what is expected. Furthermore the Admiralty Scale is identified as being a means of evaluating information rather than as a means of coding information to demonstrate analyst confidence in the information. Analysts are aware of the sorts of criteria that may assist in the effective evaluation of sources and information however rarely give any significant consideration to these factors excepting in cases where

there is a specific question raised about a specific element of information. It is not surprising then that many analysts simply evaluate information against the Admiralty Scale itself. Moreover a surprising number of analysts further limit themselves to classifying their confidence in information to either A1 of F6. That said there is a general view that it is other analysts who are guilty of this approach to evaluation. Generally speaking there appears to be very little intellectual rigour applied to the evaluation of information in particular where the information appears to confirm what is already known. It appears that for a significant number of analysts evaluation is an intuitive process undertaken once, at the initial exposure to that information. Information is rarely re-evaluated except in those circumstances when intelligence goes wrong.” (p.8)

The following excerpt is also discussing the Admiralty Scale.

Irwin, D. and Mandel, D. R. (2019) ‘Improving information evaluation for intelligence production’, *Intelligence and National Security*, 34(4)

“In their review of spot reports completed during a US Army field exercise, Baker, McKendry and Mace found that A1 and B2 represented 80 per cent of all reliability/credibility ratings, with B2 alone comprising 74 per cent of ratings. This tendency to treat B2 as a ‘sweet spot’ is particularly concerning given findings that decision makers who receive highly rated information are less likely to seek additional information prior to making an initial decision.

...A rating of B2 likely represents the accountability sweet spot because it uses terms with positive linguistic directionality that optimistically signal the value of the information from a trustworthy source, yet without sounding overconfident or insufficiently critical. In fact, intelligence analysts show a distinct bias towards using verbal probabilities with positive directionality in strategic intelligence forecasts. Thus, evaluators may prefer to avoid ratings worse than B2 if they anticipate being challenged for doing so. Conversely, evaluators may view ratings higher than B2 as running the risk of seeming too confident or uncritical. Note as well that as scale levels fall into disuse, they become increasingly untenable because there are strong normative pressures to avoid them. That is, the expectation is that one ought to be especially well justified to use something that is an abnormal practice. (p. 506)

... Much like the scales for evaluating source reliability, information credibility scales suffer from an inherent lack of clarity. Information credibility generally incorporates confirmation ‘by other independent sources’ as a key determinant. However, evaluators are given no guidance on how many independent sources must provide confirmation for information to be judged credible. (p.507)

... The lack of guidance regarding confirmation (specifically, what level of corroboration warrants each rating) could also facilitate overconfidence stemming from information volume. Beyond an early point in the information gathering process, predictive accuracy plateaus, while confidence continues to rise, creating substantial confidence-accuracy discrepancies. Failing to adjust for their cognitive limitations, judges often become overconfident in the face of surplus information, despite being unable to assimilate it effectively. Thus, without adequate guidance, evaluators may overvalue a piece of information confirmed six times, when an item confirmed three times has an equal probability of being accurate. In other words, the amount of confirmation on its own is a fallible indicator of information accuracy. Beyond confirmation, most of the information credibility scales examined incorporate consideration of whether an item is ‘logical in itself’. Current methods do not specify whether this simply refers to the extent that information conforms to the analyst’s current assessment. Furthermore – and not without a touch of irony – the use in certain standards of ‘not illogical’ as a level between ‘logical in itself’ and ‘illogical in itself’ is nonsensical, as ‘not illogical’ effectively means ‘logical’ (in itself).”(p. 509)

2.3 Organisational factors

2.3.1 Lack of evaluation

Chang, W. (2012) 'Getting It Right: Assessing the Intelligence Community's Analytic Performance', *American Intelligence Journal*, 30(2).

"...how often does the United States Intelligence Community (IC) "get it right"? We simply do not know. Why cannot an enterprise with a roughly \$75 billion budget answer this question? Despite reform and oversight efforts since 9/11 and myriad commissions examining intelligence failures, the IC has not developed a way to determine when, how often, and why it makes the right or wrong assessments." (p.99)

Marchio, J. (2014) 'Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis', *Intelligence and National Security*, 29(2).

"The IC effort to evaluate the quality of its finished intelligence analysis was another area marked by intermittent emphasis and a failure to sustain momentum and reach full potential."(p.173)

Hedley, J. H. (2005) 'Learning from Intelligence Failures', *International Journal of Intelligence and CounterIntelligence*, 18(3)

"... the post-mortem program lapsed in the mid-1970s for a variety of reasons: some of its principal sponsors moved on; some intelligence officers saw it as an unnecessary exercise in self flagellation; and it became embroiled in the House of Representatives Intelligence Oversight Committee's effort to exploit public controversy over the CIA's real and alleged abuses over the course of two decades." (p.441)

2.3.2 Tolerance of poor analysis

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. (2005). Report to the President.

"Long after the Community's assessment of Iraq had begun to fall apart, one of the main drafters of the NIE told us that, if he had to grade it, he would still give the NIE an "A." By that, he presumably meant that the NIE fully met the standards for analysis that the Community had set for itself. That is the problem." (p.12)

Borek, J. J. (2019) 'Developing a Conceptual Model of Intelligence Analysis', *International Journal of Intelligence and CounterIntelligence*, 32(4)

"Douglas MacEachin, a career analyst at the CIA who served as Deputy Director for Intelligence from 1993 – 1995, reportedly told a colleague in 1994 that after reading a number of published intelligence assessments designed to support policymakers "roughly a third of the papers...had no discernible argumentation to bolster the credibility of intelligence judgments and another third suffered from flawed argumentation" (p.5)

2.3.3 Not enough focus on standards

Marchio, J. (2014) 'Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis', *Intelligence and National Security*, 29(2)

"Gates ... argued that the best defense against politicization and ensuring objectivity was strong tradecraft. Gates went on to make the case for many of today's IC's Analytic Standards. He observed that 'distortion of analysis is much less likely, and much easier to spot, if there is a concerted effort at all levels to observe basic standards'. He went on to stress that 'We must make explicit what is known and clearly distinguish between fact, inference, and judgment' and emphasized the need for alternative analysis, stating: 'We should provide an outlet for different interpretations, theories, or predictions in our mainline publications, not just in a staff note or a piece at the back of the monthly'. Gates supported his push for more rigorous tradecraft by urging the DDI to 'develop a DI tradecraft manual and work with the Office of Training and Education to enhance the tradecraft training that analysts receive in formal courses'. He also asked the DDI to appoint a full-time ombudsman, take steps to include 'well-reasoned, relevant, and factually supported alternative views in mainline products', and 'mandate wider dissemination of studies by the Product Evaluation Staff focused on politicization and the use of alternative analysis'."(p.177)

2.3.4 Too much focus on standards

Lowenthal, M. M. (2008) 'Towards a Reasonable Standard for Analysis: How Right, How Often on Which Issues?', *Intelligence and National Security*, 23(3)

"The emphasis on new analytical standards on sourcing is useful to a point, but runs the very real risk of becoming an intellectual straitjacket, especially for analyses that may depend more on the analyst's "gut" than on available intelligence sources. The ODNI has promulgated a series of directives on analytic standards and sourcing requirements for analytic products. Again, these are useful in a general sense but they appear to be having an unexpected effect on analysts. The directive on sourcing seems to be especially problematic, with analysts attempting to adhere so closely to its strictures that they allow themselves only narrow bands within which they feel free to write, especially if they do not have "all" of the sources or cannot document their every thought. This behavior stems, again, from the fear of being wrong and of being called out for not following all of the rules for analysis. Analysts are reduced to being transmitters of received intelligence with little original thinking, which cannot—after all—be sourced." (p.36)

2.3.5 Not enough of a shared understanding of standards or analytic rigour

Aid Memoire on Intelligence Analysis Tradecraft, Canadian Forces Intelligence Command (2015)

"Analytic rigour" is a term that is widely used, but few analysts or managers can actually describe what it entails –not very helpful!" (P. 13)

Gentry, J. A. (2015) 'Has the ODNI Improved U.S. Intelligence Analysis?', *International Journal of Intelligence and CounterIntelligence*, 28(4)

"First, an examination of the content and impact of ICD 203 is in order. While not bad, these standards are basic and reflect poor adherence to normal communication practice in the United States—by the drafters of the IRTPA. For example, the five core standards are a mixture of words and partial sentences.

Tradecraft Standard 2 uses “caveat” as a verb, not a noun—conventional colloquial practice but not normal in reputable written communication. Tradecraft Standard 3 evidently is intended to distinguish raw intelligence reporting from analysts’ judgment, but does so awkwardly. Most of the linguistic flaws are fixed in the revised ICD 203 of January 2015. Tradecraft Standards 5 and 7 are so basic as to beg the questions: did the authors know what intelligence is about, or did they presume that serving IC analysts do not know? Increasingly apparent is that the ODNI authors of ICD 203 drafted verbiage for an analytic workforce they recognized to be deficient in basic skills—a reasonable observation in 2007, after the hiring binge dramatically reduced the collective skills of the IC’s analyst corps.

The IRTPA requires the ODNI to submit an annual report to Congress, and the ODNI/AIS reports periodically in other ways on the extent to which IC products meet its standards. AIS in turn asks the agencies for data. It, and the agencies, sample the universe of finished analysis and assign the four grades to seven of the eight tradecraft standards, reporting percentages of scores in each category and tracking these over time. But the samples are small and taken from a variety of publications and topical areas that vary from year to year. AIS makes no judgment about what scores are acceptable or not, leaving doubts as to whether the IC is doing well or not. Annotated ratings of specific products appear to go to analyst authors and their reviewer/managers only episodically, limiting opportunities for them to help improve analysis. (p. 644)

Borek, J. J. (2019) ‘Developing a Conceptual Model of Intelligence Analysis’, *International Journal of Intelligence and CounterIntelligence*, 32(4)

“Tradecraft: A Black Boxed Concept

Within the IC and academic communities meaningful definitions of tradecraft or even a scope of what analytic tradecraft encompasses, are rare. For example, a 54 page government funded assessment of analytic tradecraft, including ten suggestions for improvement and a separate chapter of recommendations, fails to define or describe the term (Treverton & Gabbard, 2008). Gannon’s (1997) definition of “the special skills and methods required to do their business” (p. v) and Johnston’s (2005) more general “practiced skill in a trade or art” (pp. 17-18) are decidedly vague. The WMD Commission’s definition of tradecraft as “...the way analysts think, research, evaluate evidence, write, and communicate...” may be the most comprehensive but still lacks necessary precision (Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction [WMD Commission,] 2005, p. 12). Johnston states that the term has become a “catchall for the often idiosyncratic methods and techniques” (p. 17) of analysts and believes the term purposely “obfuscates and complicates the reality of their [analysts] work.” (p. 18). While Bruce and George (2008) identify seven essential skills for an analyst, combining the professional traits of “historian, journalist, research methodologist, collection manager, and professional skeptic” (p. 3); current analytic standards (Director of National Intelligence, 2007; Pigg, 2009) focus on conveying analytic judgments rather than the process of arriving at them, resulting in uncertainty about the role and scope of analysts...(p.2)

Addressing analytic tradecraft in the IRTPA represented a new approach to legislative oversight of intelligence. The assumption is that improving analytic tradecraft will result in fewer intelligence failures. While this may be a valid premise, achieving those improvements without an understanding of what constitutes analytic tradecraft is unlikely. The community’s understanding of the process of analysis itself is still immature, compounding the uncertainty underlying analytic tradecraft and limiting the effectiveness of any potential reforms.” (p.7)

2.3.6 Consistency of emphasis on quality of analysis

Marchio, J. (2014) 'Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis', *Intelligence and National Security*, 29(2)

“This review highlights the need for the IC to maintain its focus on analytic tradecraft and ensure it is evident in its analytic products and in its workspaces. The IC’s historical record drives home the necessity for continual self-evaluation as well as vigilance to guarantee that the emphasis placed on tradecraft does not disappear or become intermittent or dependent on a single advocate. Analytic tradecraft proficiency – like knowledge of a foreign language – will be lost if not employed, recognized, and rewarded. As noted in the 2005 WMD Report: ‘Our studies, and many observers, point to a decline in analytic rigor within the Intelligence Community. Analysts have suffered from weak leadership, insufficient training, and budget cutbacks that led to the loss of our best, most senior analysts. There is no quick fix for tradecraft problems.’” (p.181)

2.3.7 Organisational and bureaucratic pressure and misguided emphasis

Hoffman, R. et al. (2011) 'Reasoning difficulty in analytical activity', *Theoretical Issues in Ergonomics Science*, 12(3)

“An organisational context can sometimes be disincentivising and demoralising, especially when the analyst is pressured by the need to avoid error and blame.” (p.228)

Rojas, J. T. (2019) 'Masters of Analytical Tradecraft: Certifying the Standards and Analytic Rigor of Intelligence Products'.

“Many experts have pointed to an IC obsessed with the churn of data and overly focused on the tactical and current intelligence needed to impact the tactical battlefield in real time. They have gone as far as to reference this phenomenon as “tyranny”: the tyranny of current intelligence, tyranny of taskings, tyranny of daily operations, and tyranny of the immediate, to name a few.” (p.2)

Lowenthal, M. M. (2013) 'A Disputation on Intelligence Reform and Analysis: My 18 Theses', *International Journal of Intelligence and CounterIntelligence*, 26(1)

“A misguided emphasis is placed on efficiency. Intelligence analysis is an intellectual activity. It cannot be made efficient. The ic should strive for effective intelligence analysis.

All government programs are subject to scrutiny to ensure that taxpayer dollars are being spent wisely and well. Everyone is against waste and in favor of efficiency. But given that intelligence analysis is an intellectual activity, how can it be made more efficient? Should analysts be asked to think faster? Or to handle multiple issues simultaneously? How correct should the Intelligence Community be, and how often, on which issues, for \$80 billion? How do taxpayers know what they are getting for their \$80 billion? Ironically, in the case of intelligence what they get is often the things that do not happen: the foiled attack; the counter-diplomatic move. The IC’s goal should be intelligence that is effective, that meets the needs of policymakers on an ongoing, timely basis—and not a futile search for efficiency.” (p.35)

2.3.8 Communication of reasoning, rigour, and tradecraft

Marchio, J. (2014) 'Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis', *Intelligence and National Security*, 29(2)

“...the emphasis and visibility afforded tradecraft in the IC’s analytic production has fluctuated significantly throughout the community’s existence. Early on many products included source reference citations, explicitly addressed intelligence gaps and analytic assumptions, and prominently highlighted alternative views. Later, however, many of these same tradecraft elements appeared less frequently in finished intelligence products... (p.160)

Infrequent Visibility of Tradecraft in Products

Manifestations of analytic tradecraft in IC products became less frequent and more sporadic after the mid-1960s. In the 1950s CIA Provisional Intelligence Reports and the IC’s National Intelligence Surveys regularly incorporated robust source citations, source summary statements, and annexes addressing intelligence gaps. This was true to a lesser degree with NIEs and other national-level products.

Declassified intelligence products from the late 1960s and 1970s contained few of the tradecraft elements prevalent during the 1950s. Sources, for example, were addressed in less detail or not at all. Similarly, treatment of intelligence gaps and assumptions were noted in only one of the declassified NIEs produced during these two decades. Discussion of analytic methodology, estimative challenges, and alternative analysis fared somewhat better, driven in part by the problems associated with a growing Soviet nuclear force and the Nixon administration’s disdain for the IC’s judgments and mistrust of its methods. For example, National Security Advisor Henry Kissinger’s staff attempted to resolve a dispute over the capabilities of the Soviet SS-9 intercontinental ballistic missile by examining the intelligence evidence at length. Subsequently the intelligence community was required to use a new format for estimates on Soviet military capabilities that included ‘a series of optional analyses and exhaustive displays of the evidence underlying each judgment... (p.169)

Reflecting and perhaps partially explaining the intermittent tradecraft emphasis was the ongoing debate over the merit of including certain tradecraft elements in analytic products. For example, disagreement over whether source citations should be incorporated into products not only surfaced in *Studies in Intelligence* in 1964 but was cited over a decade later in the DCI’s publication *Review of National Intelligence*... (p.170)

Consumers also were confronted with greater amounts of information to digest and less time to do so. ‘Short is better’ has always been the perception of consumer preferences in the IC. In arguing against including source reference citations in intelligence products in 1964, Allan Evans wrote: ‘The first and most important arguments are that our customers won’t read fat papers and “almost certainly” in overwhelming majority don’t want to be bothered with documentation’. James Hanrahan put it even more bluntly: ‘If we have not been asked specifically but feel it desperately important to get something across to the senior policy maker, brevity is the overriding value. Conclusion and judgments are the nub; argumentation can come later’.” (p.176)

2.3.9 Tradecraft advocates

Marchio, J. (2014) 'Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis', *Intelligence and National Security*, 29(2)

“The emergence of tradecraft advocates also helps to explain the periods of emphasis and the seeming lack of attention at other times. During the 1980s and 1990s a small group of senior leaders recognized

the need for rigorous analytic tradecraft and strongly supported initiatives and programs designed to strengthen it. Robert Gates was the most prominent and influential of these advocates. First as the CIA's Deputy Director for Intelligence (DDI) (1982–86), then as Deputy Director of Central Intelligence under William Casey and William Webster, and later as DCI (1991–93), Gates put tradecraft at or near the top of his agenda. As DDI, Gates personally took an intense interest in the quality of CIA's products. Gates or his deputy reviewed all products going to policy-level consumers. In addition, he established CIA's Product Evaluation Staff." (p.177)

2.3.10 Ineffective training

Dhami, M. K. and Careless, K. (2019) 'Intelligence analysts' strategies for solving analytic tasks', *Military Psychology*, 31(2)

"...the intelligence community may wish to rethink the contents of its analytic thinking training. This is because we did not observe many systematic differences in reported strategy use amongst analysts who had completed analytic thinking training and their counterparts who had not. Analytic thinking training typically includes a focus on critical thinking and the use of specific analytic techniques (e.g., Office of the Director of National Intelligence, 2015; UK Ministry of Defence, 2013; US Government, 2009). However, there is no conceptual framework to tie these thinking skills together. There is also an overemphasis on cognitive bias, with few debiasing instructions beyond telling analysts to "be aware of bias!" (see Belton & Dhami, in press)." (p.124)

2.3.11 Objectivity and subjectivity

Irwin, D. and Mandel, D. R. (2019) 'Improving information evaluation for intelligence production', *Intelligence and National Security*, 34(4)

"In our view, there are problems with both the goal itself and with the means the IC has taken to pursue it. The goal of rooting out subjectivity in intelligence production is unrealistic because so many aspects of intelligence production depend on the informed judgements of individuals and on their expert knowledge. If subjectivity were to be denied its place in intelligence analysis, much of that knowledge would need to be excised, but to be replaced by what exactly? Intelligence analysis cannot effectively proceed on the basis of statistical relative frequency data alone. It must draw on analysts' cognitive models of complex geopolitical phenomena, which are inherently subjective. We do not believe intelligence production would be aided by the elimination of subjectivity from its processes, even if it were possible.

The problem with the IC's desire to contain subjectivity, however, has been compounded by its means for achieving that goal. The IC rarely draws systematically and with proper care on scientific theory and evidence concerning effective means for improving expert judgement. Perhaps even more importantly, the IC does not routinely put what it regards as good ideas for intelligence production methods to well-conceived scientific tests. It rarely, if ever, does the equivalent of running a randomly controlled trial to test whether its proposed 'treatment' is more effective than the status quo 'control.' Rather, the IC all too often tasks intelligence personnel with 'thinking up' something better than the status quo, without ever confirming whether these 'good ideas' in fact succeed or fail in achieving their intended effects. Nor is anecdotal evidence likely to be sufficiently revealing since most intelligence products are not systematically tracked for accuracy or other quantifiable facets of skill in expert judgement, in spite of recommendations for such monitoring." (p. 511)

2.3.12 Multiple sources of bias

George, R. Z. (2010) 'Beyond Analytic Tradecraft', *International Journal of Intelligence and CounterIntelligence*, 23(2)

George 2010 provides the following taxonomy of analytic errors.

Source of Bias	Description	Analytic Symptoms
Cognitive	Uses mental "mind-sets," which filter and distort incoming information and cause analysts to interpret it in ways which reinforce existing beliefs about an intelligence subject. Often unconscious and deeply resistant to change.	<ul style="list-style-type: none"> • Holds to a "conventional wisdom" • "Sees what one expects to see" • Judges the status quo as prevailing • Dismisses contrary information • Rests on long-held "assumptions"
Cultural	Interprets phenomena and a target's behavior in terms of one's own culture. Often uses customs, standards of "rationality," and evidence common to oneself in judging the behavior of others. Ethnocentric judgment about the value or characteristics of other ethnic, cultural, or religious group.	<ul style="list-style-type: none"> • Applies "mirror-imaging" to a target • Uses "western" ideas of logic and risk calculation • Attributes specific behavior to groups • Reflects cultural superiority or insensitivity
Organizational	Establishes a shared sense of a "group" identity and mission that shapes how analysts view information and analysis that might challenge their own assessments.	<ul style="list-style-type: none"> • Justifies arguments by past record • Links unit's mission to analysis • Sees unit's successes more than failures • Coordination "waters down" judgments
Political	Tailors analysis to support a preferred policy preference, either because the analyst and/or the policy customer favor it. "Politics" rather than open inquiry drive the intelligence process.	<ul style="list-style-type: none"> • Restricts search for contrary evidence to prevailing views • Permits misuse or selective use of analysis to support policy • Self-censors analysis • Answers intelligence questions that reflect political agenda

Figure 1. Taxonomy of Analytic Errors.

2.4 Technological factors

2.4.1 Poor tools

Hoffman, R. et al. (2011) 'Reasoning difficulty in analytical activity', *Theoretical Issues in Ergonomics Science*, 12(3)

“One reason why new software is not up to par is that the development activity was based on assumptions about the cognitive work of intelligence analysts. Experience shows that the new software tools often do not tie sufficiently (or at all) to the actual work requirements of analysts ... Worse still, tools can actually impede intelligence analysis. In studies conducted by ourselves and our colleagues we have seen many instances where:

- tool use is labour intensive with no obvious value-added, e.g. high cognitive effort in building argument templates for specific problem sets.
- confusion about, gaming of or outright disregard for fusion algorithms (which are intended to support deductive reasoning),
- uncertainty and mistrust concerning the equations that are ‘riding under the hood’, and that magically generate conclusions (in the form of yet more probability numbers),
- lack of standards of judgment for answering and/or scoring questions within the structured arguments,
- disruptions and impediments to collective problem solving,
- a focus on one person-one machine contexts of work leads to underspecification of collaborative processes,
- inappropriate requirements for low-level judgments within the ‘zone of indifference’ – i.e. requiring analysts to make precise, numeric judgments that otherwise could ‘go either way’,
- discomfort with the inflexibility of argument structures or templates.” (p.229)

3 Opportunities to Enhance Analytic Rigour

3.1 Standards

3.1.1 Improve adherence to standards

Marchio, J. (2014) 'Analytic Tradecraft and the Intelligence Community: Enduring Value, Intermittent Emphasis', *Intelligence and National Security*, 29(2)

“What can we learn from the past treatment of analytic tradecraft that can aid the IC in the years ahead? A look at the IC’s past use – or non-use – strengthens the case for the current IC Analytic Standards. The standards are not simply a creation of IRTPA or a knee-jerk reaction to a perceived or real intelligence failure. While not codified earlier as standards, the value of

identifying sources, addressing analytic uncertainty, and defining assumptions were all clearly recognized and practiced early on and throughout much of the IC’s existence. So was the need to identify alternative views and explore the implications for the United States... (p.180)

The record also shows that analytic tradecraft must evolve to provide the greatest value to intelligence analysts and IC consumers. As analysts began to rely on a greater number and variety of technical intelligence collection platforms, providing insight into the nature and quality of the underlying sourcing became more difficult.” (p.181)

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. (2005). Report to the President.

“Improve the rigor and “tradecraft” of analysis.

Our studies, and many observers, point to a decline in analytic rigor within the Intelligence Community. Analysts have suffered from weak leadership, insufficient training, and budget cutbacks that led to the loss of our best, most senior analysts. There is no quick fix for tradecraft problems. However, we recommend several steps: increasing analyst training; ensuring that managers and budget-writers allot time and resources for analysts to actually get trained; standardizing good tradecraft practices through the use of a National Intelligence University; creating structures and practices that increase competitive analysis; increasing managerial training for Intelligence Community supervisors; enabling joint and rotational assignment opportunities; ensuring that finished intelligence products are sufficiently transparent so that an analyst’s reasoning is visible to intelligence customers; and implementing other changes in human resource policies—such as merit-based-pay—so that the best analysts are encouraged to stay in government service.” (p.26)

3.2 Methods

3.2.1 Introduce numeric estimative language

Barnes, A. (2016) 'Making Intelligence Analysis More Intelligent: Using Numeric Probabilities', *Intelligence and National Security*, 31(3)

“The use of numeric probabilities and the adoption of a standardized mapping of verbal probability terms were means to an end: improving the overall quality of the analytic judgments being made. What broader lessons were learned in this regard? The systematic review of the division’s analysis for the

purpose of the forecasting quality study provided useful insights. The review took place over three phases and the results of each phase informed changes in how the division made analytical judgments.

The forecasting quality study highlighted the need for clear, falsifiable judgments. For judgments to be useful to clients, the predicted outcome had to be clearly described. The timeframe for the judgment needed to be clear, either from the overall context of the paper – this could be stated explicitly up front – or specified for an individual judgment. In many cases, analysts were more comfortable making a judgment within a specific – usually shorter – timeframe: ‘the next three to four months’ vs. ‘the next year’. The review process identified instances of vague and poorly-worded judgments. At the time the original assessments were drafted and approved, such judgments were considered sound and authoritative by both analyst and director. However, in hindsight, it became clear that they were not worded in a way that provided meaningful added value for the client. Over time, with greater attention being paid to the judgments made by the division, there was some improvement in their clarity. But this was a constant challenge amidst the daily pressures of producing intelligence analysis on complex situations under tight time constraints. Furthermore, it was found that some analysts often made too many judgments, frequently on issues of marginal importance. This led to subsequent efforts to make fewer but better judgments.” (p.340)

3.2.2 Discuss rigour more in briefings

Zelik, D., Patterson, E. S. and Woods, D. D. (2007) ‘Understanding Rigor in Information Analysis’, in. Proceedings of the Eighth International NDM Conference

“Participatory Exchange Model

The Rigor Metric provides a framework for understanding analytical rigor in information analysis, but there are many possible ways that it might be leveraged to lessen the risk of shallow analysis. One promising direction for using the metric to shape the envisioned future of information analysis focuses on applying it at transition points in the information analysis process. Thus, the Participatory Exchange Model is proposed as an approach for communicating about intelligence analysis outputs—an approach that shapes interactions around facilitating the ability of both analyst and decision maker to understand and discuss the rigor of an analysis process. This alternative approach to the traditional briefing interaction builds on the insights revealed by the study and through the Rigor Metric, emerging from a collaborative effort exploring data overload in intelligence analysis.” ...

There is both empirical and theoretical support for the Participatory Exchange as a promising model for briefing interactions. During the development of the study, a number of professional analysts commented that their best and most effective briefings were, in fact, more similar to conversational dialogs than to formal presentations. This observation parallels the feedback we received from study participants, as the majority of them preferred face-to-face interactions for developing an understanding of analytic process....

Most critically, the Participatory Exchange Model, unlike the traditional model, supports the ability of both analyst and audience to visualize and understand the rigor of a process. Consequently, the interaction plays out more like a conversation —or talking with an audience—than a formal presentation, in which a speaker is talking at an audience.

The implication then, is that the audience does not end a presentation simply with a memory of a few main points, but rather they develop an understanding of the rigor of an analysis in relation to their own goals and interests. Concurrently, briefers develop a better understanding of how their analysis fits within a broader re-planning context. Thus, re-conceptualizing briefing interactions on the success model

of the conversational dialog, rather than the traditional presentation, creates a flexible framework that provides a promising, innovative direction for the future of intelligence briefings.” (pp.5-6)

3.2.3 Better guidance on the evaluation of evidence and sources

Chang, W. et al. (2018) ‘Restructuring structured analytic techniques in intelligence’, *Intelligence and National Security*, 33(3)

“Establish more explicit rules for handling evidence

We propose establishing explicit rules to weight and categorize evidence to promote consistency in the application of SATs and, more importantly, in the assessments they support. We suggest keeping records to identify the role various types of intelligence information played in assessing previous situations. The relationships between evidence (e.g., raw reporting, open source media, background information) and outcomes would help develop base-rates of event occurrence that, over time, could provide ‘outside view’ checks on portfolios of ‘inside view’ case-based assessments. Such information can be used to more properly weight evidence within techniques such as ACH; meanwhile, analysts’ source track records can be used to inform assessments of source credibility when conducting Quality of Information Checks.

Clarifying SAT rules for evaluating evidence can help analysts think critically about how they should interpret new or unfamiliar information, rather than relying on their intuition. Implementing such rules would decrease ambiguity in how new observations are interpreted—which would, in turn, lead to analysts and analytic groups dealing with evidence more consistently.” (p.346)

Irwin, D. and Mandel, D. R. (2019) ‘Improving information evaluation for intelligence production’, *Intelligence and National Security*, 34(4)

“Information accuracy as a probability estimate

For the reasons outlined earlier, we argue that a unitary measure of information accuracy that takes comprehensive account of relevant factors would be more conducive to the evaluation of information than current two-part alphanumeric systems. We further propose that information accuracy be expressed as a numeric probability estimate (e.g., Information A has a .75 probability of being accurate). This method would not only bypass the vagueness and ambiguity inherent in the current scale levels, it would also enable users to grade information with finer discrimination... (p.513)

As noted previously, the position of information evaluation within the intelligence process varies between methods, with implications for how information is rated and the extent to which it shapes analytic judgements. Furthermore, while certain methods promote collaboration between analysts and collectors (e.g., UK JDP 2–00), no guidance is given on how to aggregate multiple perspectives on the same piece of information. For these reasons, we propose that practitioners at each stage of the intelligence process provide individual accuracy assessments, which are subsequently aggregated. Under this system, a collector would first estimate the probability of an item being accurate and provide a brief but explicit outline of his or her reasoning. An analyst receiving the information would follow the same process, without reference to the first rating. To aggregate the ratings, the average probability could be calculated (see Figure 2).”

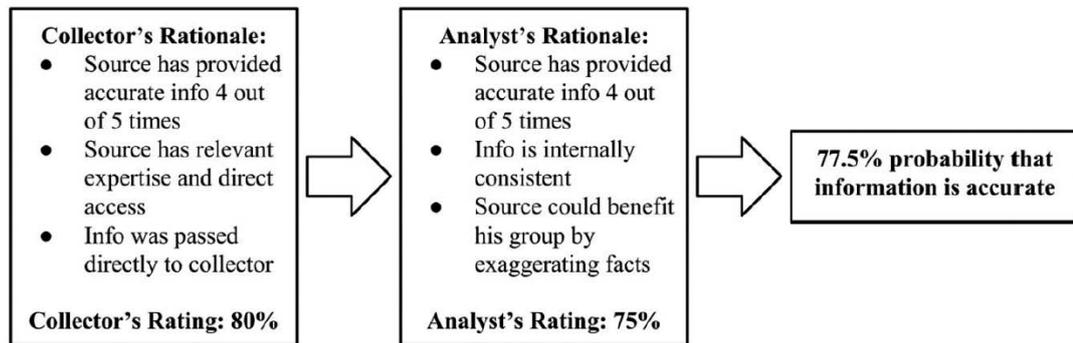


Figure 2. Example of collaborative information evaluation.

(p.515)

3.2.4 Adopt a sensemaking approach to analysis

Hoffman, R. et al. (2011) 'Reasoning difficulty in analytical activity', Theoretical Issues in Ergonomics Science, 12(3)

"If any single factor comes to mind when navigating the morass of difficulty factors, it is complexity. Cognition adapts to complexity, through processes such as problem detection and sensemaking... The sensemaking approach to critical thinking and intelligence analysis leverages our methods of cognitive field research and our understanding of expertise. It can be thought of as complementary to the structured analysis approach, which focuses on helping less experienced analysts overcome reasoning traps and biases... Some of the sources of difficulty might be overcome, some may be inevitable. Shifting from the 'blame game' that is associated with structured analysis to a sensemaking approach might help analysts and organisations cope with difficulty, though no approach, however fresh, could make the inherent difficulty of the cognitive work go away. Changes to organisational cultures (policies, procedures and disincentives) certainly have great potential for eliminating some unnecessary difficulties (e.g. the burdens of reporting). But they could not possibly eliminate the difficulties inherent in problems created by a complex and dynamic world. Changing the thrust of technology development from creating tools to mitigate bias to tools that support perceptual learning and the achievement of expertise would almost certainly lead to improvements in performance and increases in morale. (p.233)

3.2.5 Make analytic reasoning more explicit

MacEachin, D. J. (1994) 'The Tradecraft Of Analysis Challenge And Change In Cia's Directorate Of Intelligence'.

Discussing how to improve intelligence MacEachin 1994 claims that it would help to

"[redefine]...our analytic tradecraft to emphasize "facts" and the "findings" derived from them. This does not mean opinions are no longer valued, but it does require that their credibility be established through intelligence practices that clearly identify what is known, how it is known and with what level of reliability; what is not known that could have importance (sic) consequences; the "drivers" or "linchpins" that are likely to govern the outcomes of dynamic situations; the analytic calculus underlying all conclusion and forecasts; and the uncertainties in any of the components of the analysis and the implications of those

uncertainties for alternative outcomes. Adherence to these principles of analytic tradecraft are to be the standard of professional excellence -- the “professional ethic” -- of the Directorate...(p.2)

Some have suggested that including descriptions of the evidence and sources of it and laying out the analytic calculus is contrary to the need to make the products short and sharply focused. In fact, our experience in implementing these principles has shown that this is not so. A snappy march through the evidence followed by a delineation of the analytic logic for any conclusions makes for a short, punchy paper...(p.9)

Consumers present and past have consistently told us that for them, the value added -- and the credibility --- of the intelligence product is directly dependent on the information conveyed, its reliability, and their understanding of the analytic logic that supports the conclusions. If these are not made explicit and clear, the intelligence product becomes simple an opinion that may be agreed with or swept aside...(p10)

The tools of tradecraft described above... do provide a standard which is inherently resistant to skewing. Conclusions are to be presented as the result of evidence and analysis, not simply as “views.” Disagreements must be focused on the evidence and logic, not the judgment that proceeds from them.” (p10)

3.3 Evaluation

3.3.1 Better evaluation of analytic products and rigour

Chang, W. (2012) ‘Getting It Right: Assessing the Intelligence Community’s Analytic Performance’, *American Intelligence Journal*, 30(2)

“The IC needs to develop a methodology, process, and organization to evaluate the accuracy of its analytic products. This methodology should be scientifically sound, metrics-based, and both quantitative and qualitative. The accuracy of predictive intelligence products should be evaluated in the near, medium, and long term. Without assessing the Community’s performance, we are essentially stuck in “fire-and-forget mode,” unable to capitalize on lessons learned from each mistaken or correct judgment. While the IC occasionally and unsystematically evaluates analytic tradecraft using product evaluation boards, the IC should institutionalize the evaluation of analytic performance.” (p.99)

Rojas, J. T. (2019) ‘Masters of Analytical Tradecraft: Certifying the Standards and Analytical Rigor of Intelligence Products’.

Rojas proposes “establishing unit-level certified Masters of Analytic Tradecraft (MAT) analysts to be trained and entrusted to evaluate and rate the standards and analytical rigor of intelligence products prior to publication... (p.iv)

Unifying the IC efforts under one recognized certification process would be healthy as it relates to establishing a feedback mechanism and evaluating and promoting analytical rigor, thus benefiting intelligence consumers and analysts alike. Having MAT evaluators certify intelligence products and share their findings would be revolutionary. For the first time, decision makers would have an understanding of the quality of analytical rigor that informed the consumed intelligence products that they base their decision making processes on. This change would effectively empower intelligence consumers to judge sufficiency for themselves.

Given the known analysis issues leading to the invasion of Iraq and the level of dependence of the nation’s decision makers on the 2002 NIE, it is assessed that had the NIE been accompanied by a pre-

production evaluation of analytical rigor and adherence utilizing today's tradecraft standards, decision makers would have been informed on the insufficiency of process and thus demanded additional rigor. Instead, decision makers perceived rigor and falsely trusted the analysis, resulting in an over two trillion dollar expense and almost 190,000 lives lost over the course of the Iraq conflict." (p.13)

3.3.2 Measure analytic rigour

Zelik, D. J., Patterson, E. S. and Woods, D. D. (2010) 'Measuring Attributes of Rigor in Analysis', in Macro-cognition Metrics and Scenarios: Design and Evaluation for Real-World Teams.

"In the face of significant production pressures and rapidly proliferating data availability (Woods, Patterson, & Roth, 2002), it is increasingly easy for analysts and decision makers alike to fall into the trap of shallow analysis—believing an analysis to be sufficient in a context where it is not.

In responding to this challenge, the measurement of analytical rigor is advanced as a check against this risk. The model of analytical rigor presented here provides both direction for supporting this critical judgment task and a framework for a macro-cognitive measure of performance that is based on how expert analysts assess rigor. To that end, the concept of analytical rigor in information analysis warrants continued exploration and diverse application as a macro-cognitive measure of analytical sensemaking activity." (p.81)

3.3.3 Measure accuracy of assessments

Chang, W. and Tetlock, P. E. (2016) 'Rethinking the training of intelligence analysts', Intelligence and National Security, 31(6)

"Current efforts to measure analytic quality give almost no weight to accuracy metrics, yet a recent study of analytic forecasting skill in Canada demonstrates the feasibility of measuring accuracy. Declaring that 'face-validity' is enough makes it impossible to figure out how loosely or tightly coupled current metrics of good reasoning process are to correspondence-to-reality metrics. Absent such data, we will never learn how common it is for analysts who can develop great narratives are poor forecasters, or conversely how often the best forecasters prove quite maladroit at developing compelling narrative rationales." (p.915)

Chang, W. (2012) 'Getting It Right: Assessing the Intelligence Community's Analytic Performance', American Intelligence Journal, 30(2)

"For evaluations to work, predictions need to be falsifiable. Three general components need to be present: (1) a clearly predicted event, (2) a probability of this event occurring, and (3) a time horizon for the predicted event to occur. Most assessments, as they are currently written, do have these components. A statement such as "we judge, with high confidence, that overall violence levels in Iraq will decrease over the next three months" can be proven right or wrong

and can be scored. To further facilitate the measurement of accuracy, ODNI also should develop a common judgment vocabulary to be used by the all-source agencies. Expressions of certainty by CIA, DIA, and State/INR are not currently easily compared, although each organization does have an internal sliding ruler placing estimative words on the scale of uncertainty to certainty. The IC's estimative language is already nearly ready to be quantified on a numerical scale. For example, DIA's "What we mean when we say" scale provides seven values for likelihood words between "certainty" and "impossibility..." (p.102)

Another counterargument claims that holding analysts and supervisors accountable for analytic judgments will disincentivize making strong calls. I believe that leadership emphasis on making “strong calls” is misguided and leadership should instead focus its efforts on ensuring that analysts make “clear calls.” Strong calls generally lead to greater discrimination in the occurrence or non-occurrence of an event. However, if there simply is not enough data to support a call, a strong call will become a wrong call. Forcing unwarranted strong calls may incentivize analysts to make only patently obvious predictions. By emphasizing accuracy of calibration, analysts will be incentivized to make difficult, non-obvious, low-probability, high-impact estimates.” (p.104)

3.3.4 Adopt elements of rigour from the social sciences (and other related disciplines)

Landon-Murray, M. and Coulthart, S. (2016) ‘Academic Intelligence Programs in the United States: Exploring the Training and Tradecraft Debate’.

“The article concludes that the delineation between intelligence education and training may not be so stark, largely because of the educational and social science underpinnings of analytic tradecraft and competencies, as well as various issues in IC training and tradecraft. By better connecting professional practice with social science foundations, academic intelligence programs can help create a better transition from education to training.” (p.2)

Marrin, S. (2017) ‘Understanding and improving intelligence analysis by learning from other disciplines’, *Intelligence and National Security*, 32(5)

“As this ‘analytic tradecraft’ was then formalized in the 1990s, it took on the characteristics of social science methodology even more clearly. In 1999, Mary McCarthy, then the Special Assistant to the President and Senior Director for Intelligence Programs at the National Security Council, suggested that academia develop a version of social scientific ‘methodologies for dummies’ for the analytic practitioners in the intelligence community. Specifically, she suggested that managers should train their analysts in social science methodologies not so that they could become methodologists but rather so that they could ‘understand how thinking can be done and how methodologies work,’ with a special request going out to academia to ‘invent some shortcuts’ for the practitioners, to make it easier for them to adapt and adopt social science methodologies for the production of intelligence analysis.

These social scientific ‘methodologies for dummies’ have been developed since then, and given a new name: structured analytic techniques, with Analyses of Competing Hypotheses and its emphasis on falsification bringing social scientific practice directly into intelligence analysis.

Improving the practice of intelligence analysis could entail borrowing practices from social science, with its employment of the scientific method, its ethos of objectivity and independence, and pursuit of rigor. Much of this is already happening, with terminology now associated with intelligence analysis methods including: induction, deduction, independent and dependent variables, hypotheses, and hypothesis testing. Yet even though a number of contemporary analytical manuals in the intelligence community are based on social science practice, there has been little dialogue between the social scientists and intelligence analysts about their respective methodologies. While recent social science practice has a more developed literature on epistemology, formal research methods and grounding analysis in conceptual frameworks, the intelligence field has more formally developed techniques to check and mitigate cognitive bias and link the analytic effort to policy-maker needs. Each has the opportunity to learn something of value from the other. Moreover, as we move into a world in which data seems to be

ever more abundant, the boundaries between intelligence, information and social science research are likely to become more ever more fluid.

Given the similarities between intelligence analysis and social science, to improve the rigor of intelligence analysis we could be asking questions like these of both professions:

- Where do hypotheses come from, and how can they be more effectively generated?
- How do you rigorously distinguish the causal independent variables from those that only correlate?
- How do you identify or control for the presence of extraneous or confounding variables?
- How do you ensure rigor in hypothesis testing, especially when using qualitative methods?
- And what is the current state of the art in terms of analytic best practice in both intelligence analysis and the social sciences?" (p.541)

3.3.5 Evaluate SATs

Chang, W. and Tetlock, P. E. (2016) 'Rethinking the training of intelligence analysts', *Intelligence and National Security*, 31(6)

"If the intelligence community decides to continue utilizing structured techniques, it should, at least, start testing them. One approach would be to pit the community's favored techniques against the best known psychological debiasing tools in long-term comparative validity studies, ideally using accuracy metrics from forecasting tournaments as well as logical-coherence indicators used in most debiasing work. Critics and proponents should agree ex ante on how long long-term comparative studies need to run before defenders of the training status quo would concede the need for change, – or before critics would concede that they overstated the need for change. Agencies should get in the habit of basing training on the results of even-playing-field tests, not untested intuition or institutional inertia." (p.915)

3.4 Training

3.4.1 Better training

Chang, W. and Tetlock, P. E. (2016) 'Rethinking the training of intelligence analysts', *Intelligence and National Security*, 31(6)

"A delicate balancing act: designing a new training system

How should training be redesigned? Acknowledging the psychological truth that cognitive biases pull in opposing directions is a start.... Good judgment requires deftly managing error-balancing and effort-accuracy trade-offs. Balancing must be learned through exercises with objectively correct logical or empirical answers, exercises that provide a more secure epistemological foundation for determining whether training works as intended.

By extension, this means revamped training would hold analysts accountable for both good process and accurate outcomes. It would recognize that analysts can be right for incorrect reasons and vice versa. But predictive accuracy is typically a strong Bayesian signal that at least some of the underlying reasoning was correct. Measurement and accountability together form the core of a self-correcting training system...

Mastering mental balancing

Learning error-balancing starts with improving probabilistic reasoning. Tracking the accuracy of probabilistic estimates enables measurement of over-under errors. A new system should start with the core tenets of probability and review common errors people make. For starters, quantifying uncertainty becomes more natural with practice. Practice at quantifying uncertainty further sensitizes analysts to the boundary between fine-grained and pseudo-precise estimates. Learning the limits of precision across domains deepens our understanding of the appropriate effort to commit to quantifying probabilities (the concept of cognitive triage).” (p.911)

3.5 Recruitment

3.5.1 Recruit differently

Bar-Joseph, U. and McDermott, R. (2008) ‘Change the Analyst and Not the System: A Different Approach to Intelligence Reform’, Foreign Policy Analysis, 4(2)

“...we argue that present criteria for analysts’ selection and promotion, which emphasize high intelligence, strong verbal and written abilities, and good managerial skills, while important, remain suboptimal if the goal is to nurture analysts who can seek and interpret incoming information in creative and novel ways. Hence, more attention should be given to the candidates’ level of openness and tolerance for ambiguity (or need for cognitive closure) in order to overcome unconscious (or unmotivated) biases which obstruct optimal or atypical information processing...(p.128)

Change the analyst, not the system

If repeated attempts to improve and reform the intelligence system have systematically failed, the implication arises that predictable biases in human information processing may limit or restrict the ability of certain individuals within any system to accurately perceive and respond to threat in a creative and timely fashion. Yet not everyone falls prey to such biases. A detailed investigation into major intelligence fiascos shows that while some analysts completely err in estimating the situation, others assess it quite correctly (Bar-Joseph 2005; Bar-Joseph and Kruglanski 2003; Matthias 2003; Murphy 2005). This perspective allows analysts to shine a light on the problem of intelligence analysis through a different spectrum, by asking if any systematic and stable personality differences exist between individuals that may make some better at accurately interpreting intelligence information. In short, we argue that such traits do exist, and can be systematically identified using existing personality psychology tests and inventories. Specifically, personality tests, such as those based on the Five Factor Model (FFM) of personality (Costa and McCrae 2003), might be employed as a means to select, train, and promote analysts who, by virtue of their particular personality traits, remain better equipped to overcome.” (p.135)