The UK assessment failure on Iraqi: Why did it happen and what may we learn from it?

by Sten Arve

Resumé

Denna artikel tar utgångspunkt inom teoribildningen rörande ”Intelligence Failure” och undersöker det brittiska underrättelsearbetet beträffande massförstörelsevapen i Irak 2002/03 för att initiativt söka förklaringar till felbedömningarna och sedan se vilka lärdomar som kan dras. Först undersöks händelseförloppet utifrån ett generellt underrättelseperspektiv med ett särskilt fokus på den regionala dynamiken och biografisk analys. Därefter analyseras detta resultat utifrån tre perspektiv på ”Intelligence Failure”; organisatoriskt, processanknutet, psykologiskt, utgående ifrån forskarna Zegart, Betts och Heuer. Slutligen lyfts lärdomar och slutsatser fram, som bland annat framhäver vikten av historiografiska och andra akademiska metoder inom underrättelseetnästen, betydelsen av att kunna identifiera försök till vilseledning från motståndare eller andra aktörer, samt vikten av att underrättelseetnästen i sitt arbete står fri från politiska påtryckningar om önskvärda bedömningar och resultat. Som helhet visar artikeln att viktiga lärdomar kan dras av tidigare misstag.

This article argues that the UK assessment of the Iraqi WMD (Weapons of Mass Destruction) programme was an intelligence failure due to insufficient relevant collection and validation of sources, combined with cognitive inference among both analysts and policy makers. It was bolstered by deficit consideration to regional dynamics, probably caused by “mirror imaging”. Likewise, there was an insufficient focus on biographical analysis, which is highly relevant to understand authoritarian leadership and its intentions. Furthermore, the failure can best be explained by Heuer’s theory which outlines that psychological factors like perception, analytical strategies and cognitive bias are crucial to prevent intelligence failure.

To demonstrate how the UK assessment was a failure, the article initially highlights historiographic challenges related to learning from failure as well as the nature of intelligence, followed by a literature overview and a case chronology. Subsequently, the article examines the failures of collection and analysis in general and continues with particular attention to regional dynamics and biographical approach. The examinations’ findings are then matched with three “failure theories” in a comparative analysis to see which explains the failure best. Those results are then corroborated with results from the previous examination to generate lessons learned.

Finally, the article’s conclusion and some ideas for further research are presented. General lessons drawn are that scholarly and historiographic methods are vital to ensure valid collection and analysis of intelligence data, which is supported by Heuer’s theories. Likewise, several collection methods should be used to collect intelligence, providing a variety of sources and matching the regional
perspective, while paying attention to analytic tradecraft. Further, the importance of biographical analysis is accentuated. Finally, priority should be given to collection on capability vs intent and enough time must be allocated.

Epistemological disclaimer
When examining intelligence operations and intelligence failure the challenges within historiography must be recognised as well as the inherent problems related to the nature of intelligence, like deception, reciprocal action and secrecy. In historiography, the interpretation of the chronological facts makes up evidence of cause and effect. The strength of the evidence is depending on the strength of the argument behind the interpretation and peer review acceptance. The path forms a narrative that indicate what event was caused by what action. A certain event may stem from multiple causes, which make historiography as challenging as intelligence analysis.

In intelligence, the organisations are constantly involved in a struggle against “outside enemies”, according to Betts. This struggle leads to reciprocal interaction that includes deception, which complicates the search for causation even more. In intelligence, the need for secrecy versus need for sharing, the struggle between timeliness and accuracy, conflicting collection requirements, physical limitations of cognition are inherent problems where you just need to “strike a balance”. Further, as Dahl notes, “by focusing on failure we may lose sight of the successes and best practice and related learning”. So while learning from the WMD failure, explaining causation is intricate and what was working well must also be noted. Furthermore, examination of failure may become “post-mortem”-inquiries focusing on allegations and looking for scapegoats instead of scholarly researching lessons to learn. Consequently, this article adhere to Marrin, who argues that marginal gains are possible, and Hollister Hedley stating; “Though it is impossible to learn once and for all how to prevent the recurrence of something that is inevitable, the hope is that the ratio of success to failure will improve”. This article strive to contribute to that important learning, using historiography and intelligence failure theory while bearing the related challenges in mind.

Literature overview
The literature on the UK assessment on WMD in Iraq show general consensus about why the failure occurred. From a historiographic perspective, this article mainly used secondary sources because classification restricted access to primary sources. However, the Butler inquiry (2004) had access, making the Butler report a “primary” secondary source. The more recent inquiry (2016) by Chilcot (a member of the Butler inquiry) also had access to governmental documents and interviewed 150 sources. Several scholars have given their view on the failure.

Richelson’s Spying on the Bomb (2006) covers the event from a perspective of nuclear proliferation and intelligence whereas Jervis Reports, Politics, and Intelligence Failures: The Case of Iraq (2006) is critical of the inquiries in UK and USA but portrays the Butler one as the “most sophisticated” of them. Davies (2006) A Critical Look at Britain’s Spy Machinery, points at a need for adjustment in the organisation of requirements while Pythian (2006) The Perfect Intelligence Failure looks at the failure within the US Intelligence Community but points at how USA put pressure on the UK government prior to the war.
Betts (2007) *Two faces of Intelligence Failures* also focuses on the US-related failure but makes the point that most western intelligence agencies misinterpreted the Iraqi deception and commends the Butler report for taking a wider view on organisational issues than the US inquiries. Davies and Gustafson (2017) *Weighing the evidence* use the Iraq case HUMINT to illustrate challenges of source validation. That well-known authors positively “peer review” the Butler inquiry and use it as base for further discussion strengthens its credibility as a source. The case’s historiography is not perfect but has a reasonable accuracy.

**Chronology**

The article’s perspective starts with the Iran-Iraq war 1980–88, where Iraq initially attacked Iran and used chemical weapons during the war. Iraq also got assistance from USA while the conflict was ongoing. It ended with a stalemate and lasting grievances between Iraq and Iran. During that war, Israel attacked the Osirak reactor in Iraq, to counter Iraq’s nuclear weapons program. After the war, Iran got assistance from Pakistan and A Q Khan to develop its’ nuclear capabilities. In 1990 Iraq invaded Kuwait due to tensions concerning oil and financial agreements related to the Iran-Iraq war. The international community responded and liberated Kuwait during the Gulf War in 1991 and was startled at how far Iraq had come in pursuit of WMD, in particular nuclear ones. This was revealed through inspections of UNSCOM that was in place 1991–1998. Further, in 1995 the defector Hussein Kamil provided information that the declared Scud missiles had been destroyed. During its reign, UNSCOM reported of continuous Iraqi deception and denial, obstructing the inspections, which in turn led the JIC to be overcautious or worst-case inclined when assessing Iraqi capabilities. UNSCOM was forced to leave Iraq in 1998 and inspections were not re-established until 2002 with the new body, UNMOVIC. The way Iraq forced the inspectors to leave reinforced the mistrust of Iraqi official statements and declarations and how intelligence assessments to evidence into account. In the period 1998–2002, Iraq could reinstitute the strive for WMD weapons without international oversight. The 9/11 terrorist attack in USA changed the JIC (and western) perception of the threat from terrorist organisations, especially related to concerns of them acquiring WMD. Following that, President Bush portrayed Iraq as sponsoring terrorism and pursuing WMD-capability in his “axis of evil”-speech in early 2002. Later same year, Prime minister Blair stated strong support for Bush, “I will be with you, whatever”.

The US administration in 2002–03 increased pressure for a case leading to war. The JIC assessment in early September 2002 contained judgements on Iraqi intentions and capabilities concerning WMD that were balanced, however built on new and not fully validated human intelligence. These judgements were then published without the original caveats in a “Dossier” in late September. In November 2002 UNMOVIC resumed inspection of WMD inside Iraq and inspected 350 sites out of which 44 were new and Iraq proved to be more cooperative than during UNSCOM. UNMOVIC found neither evidence of WMD, nor of full disarmament during its short operation. In the JIC assessments in 2002–2003, there were also some specific ambiguous evidence that played out.

The first was the alleged Iraqi purchase of “Yellowcake” from Niger which later
proved to be a forgery. Another was the Iraqi purchase of 100,000 Aluminium tubes that were seized by embargo in 2001. The tubes were claimed to be evidence of Iraqi breach of sanction and a pursuit of nuclear capability. There was never convincing evidence of nuclear use and the JIC noted this. However, the US National Intelligence Estimate produced in October 2002 viewed the tubes as evidence. The “45-minute claim” referred to intelligence stating that chemical and biological weapons could be deployed within 45 minutes and was never corroborated as evidence. Further, there were claims that Iraq had secret mobile biological weapons laboratories, a claim coming from non-corroborated intelligence. This claim could not be confirmed by findings after the invasion in 2003.

Why the failure occurred – main argument

Conclusions from the chronology and the literature is that the UK assessment of the Iraqi WMD programme was an intelligence failure, mainly due to lack of relevant collection and validation of sources, combined with cognitive inference among both analysts and policy makers. Concerning collection, Butler notes that untried sources gained unproportioned weight due to “scarcity of sources and the urgent requirement for intelligence”. Likewise, Jervis brands it “a case of collection failure in that the evidence collected was scattered, ambiguous, and often misleading”. Chilcot comments that assessments on Iraqi WMD capabilities were not established beyond doubt. Looking at cognitive inference, Butler comments that assessment on biological and chemical weapons had a tendency to inferring the existence of such programmes. The chronology show that negative reporting, like the one from UNMOVIC, did not have impact and that the JIC assessments had bias towards a worst-case scenario. Similarly, Chilcot notes that judgements relied too much “on past behaviour” and there was a lack of alternative hypothesis.

Ambiguous intelligence was generally interpreted as evidence of guilt among analysts as well as policy makers. Turning to policy-makers, the chronology illustrates President Bush branding Iraq as within “axis of evil” and Prime Minister Blair followed suit, leading up to the publication of “the Dossier”. Chilcot notes; “The dossier was designed to make the case and secure Parliamentary and public support for the Government’s position that action was urgently required”. Conclusively, intelligence was sparse and not sufficiently validated. Neither was alternative hypothesis or options used when assessing Iraqi intent and capabilities and the presupposed intent inferred the interpretation of capabilities. On policy level, the presupposition of Iraqi guilt was strong and reinforced by the USA pressuring heavily in favour of war.

Regional dynamics not understood

The collection mistakes and inferential analysis was bolstered by deficit consideration to regional dynamics, probably caused by “mirror imaging”. The dynamics between Iraq, Iran and Israel go back to the 1970’s as Iraq nuclear arms capability “mirrored the trajectory of the nuclear power program in Iran” and Saddam followed the development in Iran closely according to Hegghammer. Further, “Deeply worried about Iraq’s nuclear intentions”, Israel began assassinating Iraqi scientists. Concerning Israel, Betts points out; “there is no evidence that Israel’s destruction of Osirak delayed Iraq’s nuclear
We managed to accelerate it.”

The chronology has illustrated the regional power game between Iraq, Israel and Iran. As a consequence, when Saddam was forced to accept inspections, the use of denial and deception was natural to keep his regional peers in doubt of Iraqi capabilities. This regional power game was not taken into enough consideration when analysing the Iraqi regime’s intentions and capabilities. Butler notes that JIC assessments sometimes “misread the nature of Iraqi governmental and social structures” and that collection on political issues was not highly prioritised.

It seems rather that Iraqi actions were interpreted through a British lens, like in “mirror imaging”. Overall, there are few references to regional dynamics when understanding Iraqi policy.

Biographical approach missing

This article has identified that the UK assessments lacked biographical perspective, which is highly relevant when understanding authoritarian leadership and its intentions. Biographical approach can tell how a leader “applies skills and strategies to the role of the executive”, according to Lambright & Madison. In Iraq, Saddam’s rise to power and subsequently strengthened his grip using the security apparatus, Mukhabarat, promoting an extensive personality cult of himself. Regime protection was a core activity of the Mukhabarat and it included concealment of the WMD program.

According to Koblentz, regime protection is “a crucial factor that have been almost completely ignored” in the case of Iraq. Likewise, Newson & Trebbi heralds the biographical approach when analysing authoritarian elites and their mechanisms. According to Butler, the JIC’s own assessment on Saddam; “Intelligence remains limited and Saddam’s own unpredictability complicates judgements about Iraqi use of these weapons”. In sum, in the case of authoritarian states, biographical intelligence is key to understanding policy and intentions. A more focused collection on policy, combined with a biographical approach might have influenced the JIC to generate more alternative hypotheses.

What theory of “failure” helps us understand?

Three theories; Zegart’s organisational theory, Bett’s process analysis theory and Heuer’s psychological theory can be used to explain why the UK assessment became an intelligence failure. These theories are crucial for examining this case as they are recognised theories used in the failure discourse and provide distinctively different perspectives.

Firstly, Zegart’s perspective relates failure to organisational structures, culture and misleading incentives. Secondly, Betts’ perspective focuses on analysis or policy-maker side, exemplified by the view that failures “have seldom been made by collectors of raw information, occasionally by professionals who produce finished analyses, but most often by the decision maker”. Thirdly, Heuer’s perspective is that human factors cause failure, perception traps like mirror imaging as well as cognitive bias and lack of strategies for analytical judgement. This theory is not specifically a failure theory but is a baseline for intelligence analysis and is similar to failure theories with psychological perspective.

Zegart’s theory

Zegart relates failure to organisational structures, culture and misleading incentives.
Evidence supporting Zegarts theory is sparse. Butler compares the case with four similar WMD-cases and concludes that the “machinery” was working reasonably well and that the case of Iraq was not an organisational related failure.\(^4\) However, Butler follows up with proposals on some adjustments in procedure and staffing.\(^5\) Further, the British machinery is portrayed by Davies as generally successful and having best practices that should be integrated in the US system.\(^6\) Nonetheless, Davies notes malfunctions like “the malaise of requirements” and proposes some reforming, mainly of the validation and requirements function.\(^7\) As for culture, it could be argued that it lacked resilience towards cognitive bias and inference, meaning that Zegarts’ argument has some weight there. Concerning incentives, this article has not found support in the literature that malfunctioning incentives would be part of the failure. In sum, Zegarts’ theory provide some but not comprehensive understanding. This could be taken as a sign that the British machinery is functioning reasonably well.

**Betts’ theory**

Moving on to Betts’ theory, it indicates that we mostly will find roots for failure within decision-making or analysis, rarely in collection. Looking at decision-making, the actual case gives evidence that support Betts’, exemplified with the Dossier that was biased to “make the case” for the policy established from top level. The analysis aspect has a lot of evidence, mainly underlined by the inference, lack of regional perspective and understanding of Iraqi motives and absence of competing hypotheses. However, also collection leaves clear evidence of failure, especially concerning lack of relevant collection and validation. Betts’ theory thus well covers the whole spectrum of failure. There could be two arguments against Betts’ theory. Looking closer at what Betts’ emphasises, it could be argued that his theory does not fit the British system with its collegiate machinery that makes failure a collective responsibility and the demarcation on where failure occurred is blurred.\(^8\) Secondly, concerning Iraq as a clear case of collection failure, it is not a perfect match considering what Betts’ theory emphasises. However, Betts’ theory helps us understand that there were reasons for failure throughout the whole process from collection to decision.

**Heuer’s theory**

Heuer’s theory trace failure to human psychological factors; perception, analytical strategies and cognitive bias. Heuer states, “we tend to perceive what we expect to perceive” and concerning perception of facts in the case of Iraq, there is substantial evidence of “mirror imaging” and absence of different perspectives. The intelligence on “Yellowcake” and the “tubes” are fitting examples. Similarly, Heuer proposes analytical strategies, like competing hypotheses, reasoning by analogy or applying theory to counter the psychological challenges of the mind. The WMD-case provides plenty of evidence pointing at the lack of such strategies. Butlers’ earlier cited remarks on “inferential” and “worst-case inclined” and Chilcot’ comments on lack of alternative hypotheses are such examples. Heuer describes cognitive biases concerning evaluation of evidence, in perception of cause and effect and in estimating probabilities. Concerning evaluation of evidence, there are examples of vividness, like the “45-minute claim” while there was oversensitivity to consistency concerning the sparse intelligence on the WMD program.

Looking at cognitive bias, Heuer states that intelligence analysis often uses narrative
and “dominant concepts or leading ideas” form the story. There is evidence presented earlier in this article that the narrative on Iraq was that Iraq had intent on acquiring WMD and hid their actual capabilities. The collected intelligence was interpreted within this narrative. Heuer’s theory thus seem to provide good understanding on why the intelligence failure occurred. It also tackles the historiographic aspects, narrative, perspective, evidence and causation. Criticism against it could be that it does not directly address organisational shortfalls. However, Heuer’s theory can be applied to all steps in the intelligence process as well as all levels of organisation, which would give beneficial information on areas for reform.

**Theory – best fit**

This comparison demonstrates that the theories should be combined to achieve best comprehensive understanding of the failure, taking several aspects into account, as illustrated in the table below. Zegart’s theory provide some evidence here but might be better suited for locating failure in a more fractured organisational case. Looking at Betts’ theory, it provides to be a relevant explanation on where in the system things went wrong but is not a perfect fit for this case. If time only permits use of one theory, the recommended one is Heuer’s since it seems to best address the challenges of historiography and can be used at all levels of organisation.

<table>
<thead>
<tr>
<th>Theory/ Criteria</th>
<th>ORGANISATION (Zegart)</th>
<th>POLICY/ANALYSIS (Betts)</th>
<th>COGNITION (Heuer)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure</td>
<td>(Collection)</td>
<td>Analytic Strategy</td>
<td></td>
</tr>
<tr>
<td>Culture</td>
<td>Analysis</td>
<td>Mirror imaging</td>
<td></td>
</tr>
<tr>
<td>Incentive</td>
<td>Policy Maker</td>
<td>Cognitive Bias</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: The fit of respective theory to the actual case of WMD failure. The size of the letters illustrates where the core of the theory lies, while the grey background and the bold case show the comparative fit to the failure. Bold and grey background represent the best fit.*

**General lessons to be learned**

Overall, five general lessons learned are identified from this case.

Firstly, looking at this failure, the well working practices of the British machinery must still be noted. The joint effort in assessment still stands as a good example and the collective responsibility seem to mitigate after-action “blame game”. However, tendencies to politicise intelligence, like “the Dossier” must be prevented.

Secondly, this case study confirms the importance of using scholarly methods and a combination of theories to reach comprehensive understanding of intelligence failure. Further, psychological theory suits the historiographic aspects of learning well, like validating evidence and testing narratives. Consequently, the use of psychological theory to craft analytical tools and methods seem highly relevant when improving intelligence collection and analysis.

Thirdly, the value of biographical analysis when assessing authoritarian states is underlined. Likewise, expansion of its use and development of related methodology is recommended.
Fourthly, the case confirms that “all business is local” and regional dynamics must be incorporated in assessments, especially to prevent mirror imaging.

Finally, part of the failure occurred due to faulty assessment of capabilities and overly focus on intentions. This underscores Grabo’s view of the importance of basing assessment on intelligence concerning capabilities. It also underlines the importance of multi-source collection, for corroboration of intelligence.

**Conclusion**

This article examined the UK assessment of the Iraqi WMD programme and demonstrated that it was an intelligence failure, with causes ranging from collection through analysis to policy-making. Further, the lack of perspective on regional dynamics, probably caused by “mirror imaging” was highlighted as one cause for failure. Likewise, the assessment had insufficient focus on biographical analysis, utterly relevant in the case of an authoritarian-led state. Three theories of failure were applied to the case and the comparison showed that Heuer’s psychological one explains the failure best, though comprehensive understanding requires a combination of theories.

All this illustrates the multi-causal nature of intelligence failure and that scholarly methods should be used to analyse them. Several lessons learned have been generated, they all point to the need for tradecraft, integrity and creativity when working with intelligence. The case also illustrates the impact of deception, exemplified with the forged Yellowcake intelligence and its impact on assessment and decision-making. Further research to the origins of that intelligence could give valuable insights concerning deceptive intelligence. Last, the extent to which Britain was bound by US directed foreign policy and how that influenced intelligence also merit further research.

The author is Lieutenant colonel in the Swedish Air Force serving at the National Defence College.
Notes

1. This article was written by the author during the “Intelligence History” course within the Master Studies in Intelligence (MAISS) at Brunel University, 2018/19.

2. In UK, “assessment is a function of Government, not solely an intelligence ditto”, (Davies (2006) “A Critical Look at Britain’s Spy Machinery”). In this article, the focus is on the intelligence failure even if other aspects are touched upon, hence theories of intelligence failure are used to explain the case.


6. Ibid., p. 15.


12. Ibid, pp. 65-68.


34. Ibid, p. 108.
47. Heuer’s theory is similar to the theory used by Bar-Joseph, Uri and Kruglanski, Arie W.: “Intelligence Failure and Need For Closure: On the Psychology of the Yom Kippur Surprise”, *Political Psychology*, vol. 24, no. 1, 2003, pp. 78-79.
49. Ibid., pp. 142-145.