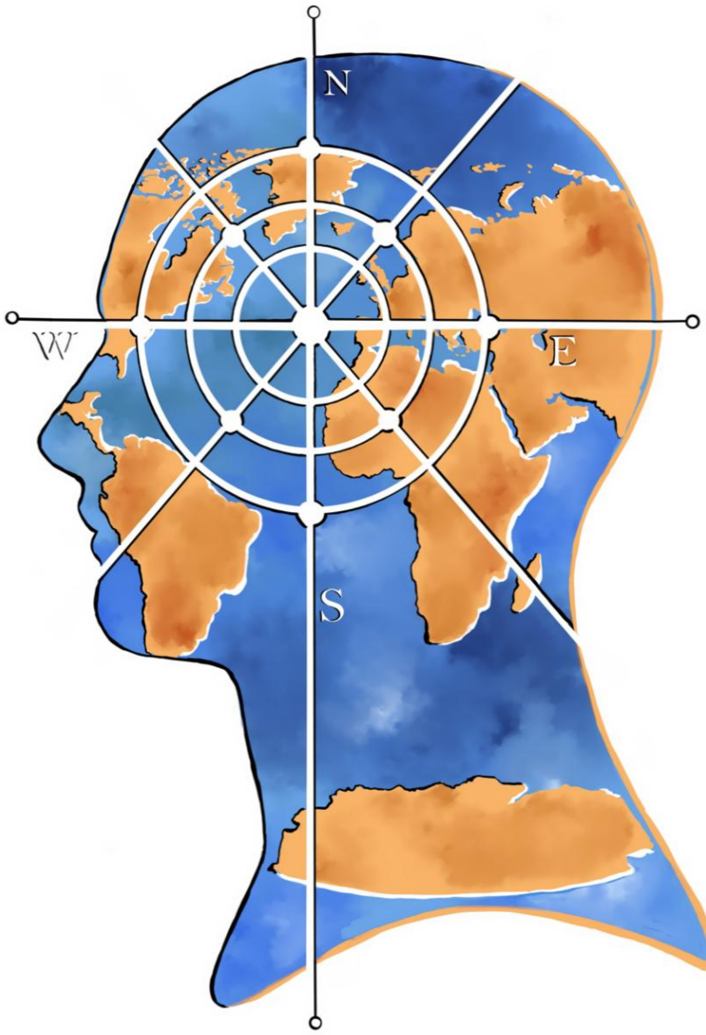
Volume 8 Number 2 December 2025 ISSN 2585-383X

JOURNAL of EUROPEAN and AMERICAN INTELLIGENCE STUDIES

AN INTERNATIONAL PEER-REVIEWED JOURNAL



Research Institute for European and American Studies – RIEAS
School of Law and Government, Dublin City University
International Centre for Policing & Security, University of South Wales

JOURNAL *of* EUROPEAN *and* AMERICAN INTELLIGENCE STUDIES

AN INTERNATIONAL PEER-REVIEWED JOURNAL

Volume 8 Number 2 December 2025 ISSN 2585-383X

JOURNAL *of* EUROPEAN *and* AMERICAN INTELLIGENCE STUDIES

An international peer-reviewed journal

Sponsored by the **Research Institute** *for* **European** *and* **American Studies – RIEAS**

Under the editorial direction of the

School of Law and Government, Dublin City University

and the **International Centre for Policing & Security, University of South Wales**

Subscriptions

Individual rate: *100 Euros | 100 US Dollars | 85 UK Pounds*
Institutional rate: *200 Euros | 200 US Dollars | 170 UK Pounds*

Mission and Scope

The *Journal of European and American Intelligence Studies* (*JEAIS*, formerly the *Journal of Mediterranean and Balkan Intelligence —JMBI*) is published by the Research Institute for European and American Studies (RIEAS) under the editorial direction of the Department of Security and Intelligence Studies at Coastal Carolina University. It is an international academic-led scholarly publication that focuses on the field of intelligence and related areas of study and practice, such as terrorism and counterterrorism, domestic and international security, geopolitics, and international relations. The journal's rationale is driven by the global nature of security challenges, where we are called more than ever to communicate and work collaboratively to solve our common problems. Thus, the *JEAIS* aspires to promote an international dialogue between diverse perspectives and experiences, based on original research on the broader European and American practice and study of intelligence. The *JEAIS* is an all-inclusive academic platform that allows accomplished and emerging scholars and practitioners from both the public and private sectors to share their knowledge, ideas and approach to intelligence studies. By crafting each journal issue through a rigorous and highly selective screening process of potential contributors, and an exhaustive review process, the *JEAIS* adheres to its mission, which is three-fold: (a) to provide an equal opportunity for academics and practitioners of intelligence to discuss and challenge established and emerging ideas; (b) to address existent knowledge gaps by advancing new knowledge; and (c) to shape the evolution of intelligence scholarship beyond traditional communities of research.

Editorial Offices

John M. Nomikos, Founding/Honorary Editor: PO Box 427, 17402 Alimos, Greece.
E-mail: secretary@rieas.gr | Tel: +30-210-991-1214 | RIEAS: www.rieas.gr.

Jonathan Smith, Managing Editor: Department of Intelligence and Security Studies, Coastal Carolina University, PO Box 261954, Conway, SC 29528-6054, United States
E-mail: jonsmith@coastal.edu | Tel: +1- 843-349-6573 | CCU: www.coastal.edu.

JOURNAL *of* EUROPEAN *and* AMERICAN INTELLIGENCE STUDIES

An international peer-reviewed journal

Sponsored by the **Research Institute** *for* **European** *and* **American Studies – RIEAS**

Under the editorial direction of the

School of Law and Government, Dublin City University
and the **International Centre for Policing & Security, University of South Wales**

Contents

Editor's Note

Joseph Fitsanakis

Professor, Department of Intelligence and Security Studies, Coastal Carolina University

We are pleased to welcome readers to Volume 8, Issue 2, of the *Journal of European and American Intelligence Studies* (*JEAIS*). This issue brings together a diverse and methodologically rich set of contributions that collectively interrogate some of the most enduring and contested problems in intelligence studies: analytic rigor and judgment under uncertainty, the communication of probability and confidence, the structural conditions shaping contemporary information environments, and the strategic implications of information operations for democratic governance. Across empirical, conceptual, and critical traditions, the articles in this volume reflect the field's continued maturation and its willingness to engage both the internal mechanics of intelligence work and the broader ecosystems in which intelligence operates.

The issue opens with Gideon Manger and Sanne van der Weide's empirical examination of the relationship between analytical rigor and predictive accuracy in intelligence assessments. Drawing on an original dataset of assessments produced during analyst qualification training within the Netherlands Armed Forces, the authors directly address a question that has long preoccupied both scholars and practitioners: whether adherence to established tradecraft standards measurably improves forecasting outcomes. Their findings offer important nuance. While rigor is positively associated with successful predictions, its relationship with the precision of probabilistic judgments proves weaker than expected. Particularly noteworthy is their analysis of "50–50" assessments, which emerge as both methodologically less rigorous and substantively less useful for intelligence consumers. The article makes a valuable contribution by empirically grounding debates about analytic standards, probability expression, and evaluative frameworks—while also raising important questions about the cross-cultural transferability of analytic rating scales.

Jeremy Levin's article continues the focus on probability and judgment but approaches the problem from a conceptual and methodological standpoint. Levin challenges the uncritical application of quantified probability to qualitative analytic judgments, particularly in contexts characterized by limited data and narrow historical baselines. He proposes a distinction between communicated probability and analytic certainty, arguing that the latter more accurately captures the logic underpinning many intelligence judgments. By introducing argument mapping as a tool for calibrating analytic certainty, Levin offers a framework designed to enhance transparency, replicability, and collaborative reasoning. This contribution speaks directly to ongoing debates about how intelligence organizations should reason, communicate uncertainty, and maintain rigor when statistical approaches are insufficient or misleading.

Shifting from analytic cognition to structural power, Elena Botts' "Opaque Architectures" offers a critical examination of the convergence between media consolidation, cultural funding, and intelligence cooptation in contemporary information environments. Drawing on cases from the Euro-American and Russian contexts, Botts argues that state-affiliated financing mechanisms increasingly function as instruments of epistemic enclosure rather than mere support for cultural production. The article advances the concept of an "epistemic cartel" to describe a durable infrastructure of perception management in which transparency is redefined through state-sanctioned visibility. This theoretically dense and provocative contribution extends intelligence studies into dialogue with media theory, political economy, and critical security studies, underscoring the field's relevance to broader questions of democratic accountability and knowledge production.

Alan Cunningham's article returns the focus to contemporary strategic competition by examining Russian information operations and their impact on American foreign policy discourse. Emphasizing the role of domestic intermediaries in amplifying disinformation, Cunningham situates political security as a multidimensional challenge encompassing both human and national security concerns. The article highlights the permeability of democratic systems to sustained influence campaigns and argues for a more systematic integration of political security into policy planning. In doing so, it contributes to a growing body of literature that treats information operations not as episodic disruptions, but as enduring features of modern conflict.

The issue concludes with Adam Hanzel's review of Simon Ball's *Death to Order: A History of Modern Assassination*. Hanzel situates Ball's work as a rare and comprehensive treatment of assassination as a transnational political practice, emphasizing its analytical value for scholars and practitioners across multiple disciplines. The review complements the issue's broader themes by reminding readers of the historical continuities that underpin contemporary security practices.

Taken together, the contributions in Volume 8, Issue 2, reflect the intellectual breadth of intelligence studies today. They demonstrate the field's capacity to integrate empirical evaluation, conceptual innovation, and critical analysis, while remaining attentive to the practical and ethical stakes of intelligence work. We hope readers find this issue both challenging and illuminating, and that it stimulates further research and debate across the many domains in which intelligence intersects with policy, society, and power.

Re-envisioning and Calibrating Certainty, Probability, and Confidence in Qualitative Analytic Judgments

Jeremy Levin

Owner, Questimation

Abstract

One of the primary tasks of intelligence analysis is generating analytic judgments to reduce the uncertainty around key unknowns. For this we use probability, generally defined as the likelihood of an event or situation occurring. This is fine when addressing events, situations, and behaviours for which we have the history and data to quantify probability. However, the more we narrow our analysis and the smaller our historical data set, the less we can rely on quantified probability and the more we rely on qualitative logic and reasoning to make predictions. This article contends these qualitative judgments are better approached by identifying our analytic certainty rather than attempting to determine their probability. Further, this article contends that if we separate our concept of probability into communicated probability, how we communicate to our clients and consumers, and analytic certainty that we calibrate and assign to judgments, we can then use argument maps to more objectively calibrate our judgments in a process that is more transparent, replicable, and rigorous, enabling better discussion and collaboration on our qualitative judgments.

Intelligence analysis begins where information ends.

One of the core jobs of intelligence analysis is exploring the unknowns beyond the information available on the actions, events, behaviours, and situations important to our clients and consumers, then delivering analytic insight to these clients and consumers based on our analytic and subject matter expertise. By the nature of the job, it is impossible to perform analysis with absolute certainty. Instead, analysts use probabilities to make and express their judgments, interpretations, forecasts, and predictions.

The current probability terms generally used in the US[1] and UK[2] intelligence services to estimate[3] and communicate probability range from "almost certainly not" to "almost certainly" along a percentile range scale, hereafter called *communicated probability*. Each nation and service uses slightly different terms and allocates percentile ranges slightly differently, but they are all roughly similar to *Table 1*.

1%-5%	6%-20%	21%-40%	41%-59%	60%-79%	80%-94%	95%-99%
Remote Chance or Almost Certainly Not	Very Unlikely or Very Improbable	Unlikely or Improbable	Roughly Even Chance; Possible	Likely or Probable	Very Likely or Very Probable	Almost Certain

Table 1: *Sample words of estimative probability*

At first glance this makes sense, and matches our mental picture and model of probability. However, we use more than one type of probability in intelligence analysis: quantitative and qualitative. So let us start by taking a closer look at probability.

Quantitative probability looks for patterns in events, situations, and behaviours, and uses numbers and mathematics to determine the probability of those events and situations occurring. For example, on a standard six-sided die, the probability of rolling a '1' is ~16.67% - or, 1 in

[1] Office of the Director of National Intelligence. *Intelligence Community Directive 203: Analytic Standards*. Washington, D.C.: ODNI, (January 2, 2015): https://www.dni.gov/files/documents/ICD/ICD_203.pdf.

[2] Intelligence Analysis. 2025. "Explaining Uncertainty in UK Intelligence Assessment." GOV.UK. (March 24, 2025): https://www.gov.uk/government/publications/explaining-uncertainty-in-uk-intelligence-assessment/explaining-uncertainty-in-uk-intelligence-assessment.

[3] In this article, the terms "estimate probability" and "calibrate probability" are both used, and require definition. For the purposes of this article, "estimate probability" is used as assigning probability to an assessment, while "calibrate probability" refers to identifying the appropriate probability to assign to the assessment.

6. The probability of rolling a '1' or '2' on the same die is ~33.33% - or, 2 in 6. And so forth. So, if you rolled the die 600 times, you should get a '1' approximately 100 times.

We apply this to human events or situations regularly, in everything from shopping habits to eating habits to spending habits to voting habits. When you gather large amounts of data on large amounts of people doing a large number of things, you can very effectively use this data to determine probabilities and predict behaviour. If you set up a camera to watch a new store display near the entrance of the local branch of Brand X store, and notice that in the first week 45% of the people entering the store stopped to look at the display, and of those 32% purchased an item from the display, you could draw some insightful conclusions on how a similar display in another branch of the store might impact purchasing. You might not be able to predict what a single person would do or not do, but you could relatively accurately predict the behaviour of a large group of people entering the store over time.

Quantitative probability becomes more problematic when we try to apply it to individual decision-making and behaviour. For example, if Jeremy has been invited to 17 parties over the last 5 years, and has attended 4 of them, what is the probability of Jeremy attending another party he was just invited to?

Standard quantitative probability would place the likelihood at ~23.53%; a 4 in 17 chance.

But what if we add more data to this? For example, of the 17 parties Jeremy was invited to, 5 invitations came from his wife's friends, 5 from his friends and 7 from his colleagues. Of those, he attended 3 of his wife's friends' parties, 1 of his friends' parties, and none of his colleagues'.

Now the probabilities might be different. Maybe there is a 60% chance he will attend a party hosted by his wife's friends, a 20% chance he will attend a party hosted by one of his friends, and no chance he will attend one of his colleague's parties.

Let us add more data. Of the 17 parties he was invited to, Jeremy cooked dinner for his family on 10 of the days the parties were held. And, of those 10 days, 8 of them overlapped with parties hosted by Jeremy's colleagues, and two of them overlapped with parties hosted by his friends. How does that change the probability calculations?

And let us add more data. Of the 17 parties Jeremy was invited to, it was raining on 12 of the days the parties were held, overlapping with 6 of his colleagues' parties, 4 of his friends' parties, and 2 of his wife's friends' parties. Also, of the 17 party days, Jeremy spent more than an hour reading a book on 14 of the days, overlapping with all but 3 of his colleagues' parties. Also of the 17 party days, Jeremy had insomnia the night before 13 of them; all 5 of the parties hosted by his wife's friends, 4 of the parties hosted by his friends, and 4 of the parties hosted by his colleagues.

How does this data change the probability of Jeremy attending this party? What data is relevant to Jeremy attending parties, and what data is not? What variables or factors interact with each other to influence Jeremy's party attendance, and how? Do we know how common or unique the variables are on the non-party days to see if they differ on the party days in a way that impacts behaviour? Can we say that if Jeremy has insomnia the night before a party he's only ~23.53% likely to attend the party? Can we say that Jeremy having insomnia makes it ~76.47% likely he will not go to a party (13 out of 17 times)?

And, what if one of Jeremy's colleagues invites him to go on a weekend road trip? This has never happened before, so we have no data on Jeremy accepting, declining, or following through on something like this. Can we use any of the probabilities from his party attendance to predict what he will do with this invitation?

These are common problems in intelligence analysis.

If you know what you are looking for and have the ability to collect it, quantitative analysis can give great insights into events and behaviour. But, it requires a lot of time, a lot of data, and very clear, controlled variables.

Otherwise, what you will probably rely on is qualitative probability, generally referring to the qualities or nature of what you are analysing. This is where subject matter expertise and abductive logic become extremely useful.

In our above examples, an analyst performing quantitative analysis would not need to know the identities of the individual shoppers entering the store branch, but would need to know the area's demographics and how demographic differences impact shopping behaviour in general. They need to have expertise in the methods of analysis, not necessarily the subjects of analysis. And, once they have their data, they use inductive logic in their analyses.

For more specific problems – like Jeremy attending a party or a weekend road trip – we need to look at the nature of the situation and Jeremy's particular qualities to understand how the factors and variables in the situation impact Jeremy's decision-making and behaviour. We would use abductive logic to look at all the factors at play and generate an assessment that gives the best explanation or prediction that takes into account everything we see.

This is where knowing Jeremy's habits and preferences helps. Suppose you know he likes getting together with people in small groups, but won't attend a party if he thinks there will be more than a handful of people there. He likes to read, and loves to cook, and he maintains a few good friends but likes to separate his work from his private life. His wife has more casual friends than he does, but does not like going places by herself. Jeremy and his wife both love traveling, especially by car where they can see the landscape and stop at places that look interesting so they can explore. Jeremy regularly takes a long time to get to sleep, and it doesn't seem to impact his life much.

None of this focuses on his historic patterns of behaviour, but focuses on his personal characteristics (and some of his wife's characteristics). Knowing these, would you be better able to assess what Jeremy will do? Do you better understand what information you are missing to make a fairly accurate assessment on him attending this particular party?

Applying qualitative analysis and using the United States' Intelligence Community's (USIC's) standard method of expressing certainty, we might say 'Jeremy is unlikely to attend this party because he generally does not go to parties. But, we do not yet know who invited him or how many people are expected to be there, so we assess this with low confidence. He probably will go on the road trip with his colleague because he loves to go on road trips. But, we assess this with only moderate confidence because he likes to keep his work and private life separate.'

This is how intelligence analysts have performed and communicated their analyses for years, and it forms the foundation of the US and UK intelligence communities' 'Probability Yardsticks,' (*Table 1*) placing probability terms against percentile ranges of probability to help their clients understand the likelihoods the analysts attribute to their analytic judgments and conclusions.

But, will Jeremy 'probably' go on the road trip, or does the analyst think Jeremy will go on the road trip but has some remaining doubts that decrease their certainty in this? Is Jeremy actually unlikely to go to the party, or does the analyst believe Jeremy will not go to the party but is not certain of this because they don't have enough information?

If you break it apart and look at it closely, Jeremy will do what Jeremy will do. There is no probability in this. The 'probability' referred to is actually the analyst's certainty in the judgment they are making. When an analyst says "Jeremy will probably do this," what they mean is "I am fairly certain Jeremy will do this." When an analyst says "Jeremy is unlikely to do this," what they mean is "I am fairly certain Jeremy will not do this."

It is a subtle difference, but an important one for the tradecraft of intelligence analysis and how we both calibrate and express analytic certainty.

I believe analysis often gets lost in the attempt to calibrate the probability of external events, situations, and behaviours, when instead it can be better used to develop and evaluate the logic and reasoning underpinning a judgment's analytic certainty.

To do this, though, I think we should first revisit our concept of qualitative probability, and separate our *communicated probability* from our *analytic certainty*.

Table 1 (above) depicts the current Words of Estimative Probability (WEPs) used by the US, UK, and North Atlantic Treaty Organization (NATO) intelligence communities. But, from a perspective of *analytic certainty* rather than probability, *Figure 1* is closer to what analysts actually mean when they use the WEPs above.

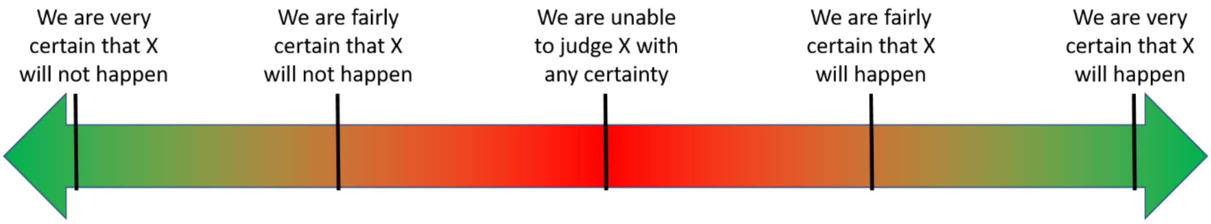


Figure 1: *What analysts mean when using Words of Estimative Probability*

As you can see, the area of greatest uncertainty is the centre, where the analyst is unable to make a judgment with any certainty, and the analyst becomes more certain in their judgment as they move toward the ends of the scale. This certainty will be based on the underpinning structure of our analysis, including the breadth, depth, and strength of the information base used to generate the judgment, the potential for deception or misidentification, the logic and reasoning used to arrive at the judgment, the number and criticality of assumptions, the novelty or familiarity of the event or situation, the plausibility of alternatives, etc. This will overlap with the current criteria the USIC uses to calibrate its analytic confidence; however, it can be better used to calibrate our certainty.

Additionally, analysts could use information and reasoning to justify and support their analytic certainty. The analytic community generally does this already when supporting and justifying their judgments on what actors *will* do, or what behaviours or events *will* occur. We are less rigorous when assessing what actors will *not* do, or what behaviours or events will *not* occur. For example, we could not have any degree of certainty that Event X will *not* occur simply due to an absence of information suggesting it *will*; absence of evidence is not always evidence of absence. We *could* have some certainty that Event X will not happen if we had evidence stating the actors involved had considered and discarded the idea of inciting Event X, or if we could reason that the repercussions of Event X would deter the relevant actors from inciting it. This evidence and reasoning would not necessarily indicate what *will* happen, but it would enable us to judge what *will not* happen – leaving all other options, actions, and events as possibilities.

Similarly, we could not judge that "Actor X is unlikely to take Action Y" simply because there is no evidence to suggest Actor X *will* take this action; we would need evidence or reasoning to specifically indicate Actor X *will not* take Action Y—giving us a degree of certainty in the judgment we are making. We would not need evidence to indicate Actor X will take a different action instead, but the presence of this evidence could lead to a different judgment.

So, let us revisit our above example. Using different language, we could communicate the above judgment as 'we are fairly certain Jeremy will not go to the party because he generally avoids parties and large gatherings, although he does go to small parties or those his wife wants to attend. We would be better able to analyse this if we had more information on who is hosting the party and how many people are expected, but we were unable to acquire it given the technical limitations and time constraints on this analysis, so we only have low confidence in this judgment.'

This brings us back to analytic confidence. Analytic confidence is currently based on the underpinning structure of analysis we recommend using for our certainty. However, *analytic confidence* could be better used to evaluate and communicate the thoroughness and rigour we have applied to our analysis, such as the hypothetical sample depicted in *Figure 2*. This is currently lacking in the USIC, which I believe is a critical shortcoming to its analysis.

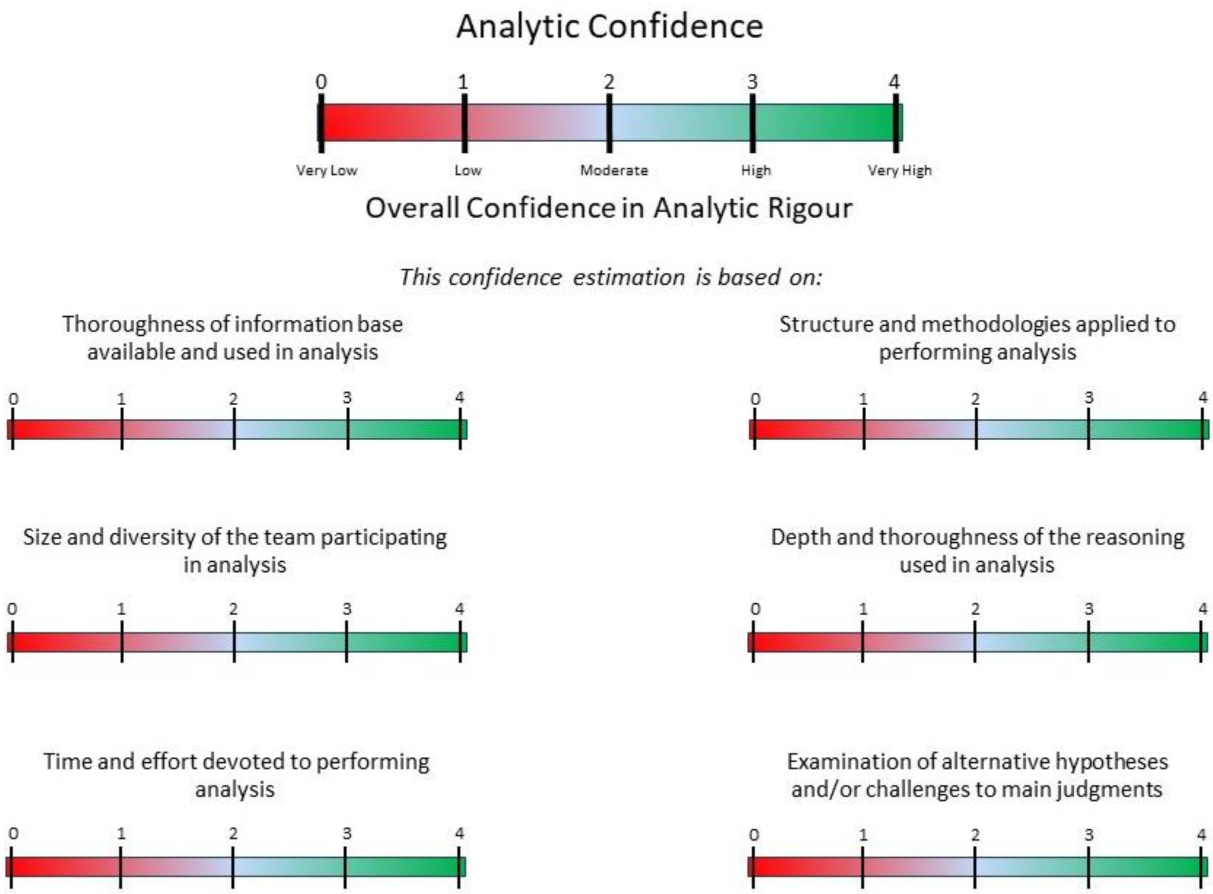


Figure 2: *Hypothetical analytic confidence measurements*

Ideally, intelligence analysis should be based in a holistic perspective of the drivers, factors, and variables that influence the outcome, event, situation, or behaviour being analysed. It should be performed with a full information base, examining multiple hypotheses, and include reporting data that both supports and contradicts each hypothesis. Lastly, it should be collaborative with input to the analysis from multiple subject-matter experts. If our analysis meets all these criteria, we could have extremely high confidence in the analysis we perform.

Most analyses do not meet this standard, though. Intelligence-producing organisations are under incredible pressure to deliver insight to clients who are making decisions on compressed timelines. Our adversaries and competitors are actively attempting to deny us information, deceive us, and overload us with useless information, thereby, increasing the difficulty in gathering a complete and reliable information base. It is virtually impossible to bring holistic subject-matter expertise to bear on every analytic problem being analysed. Because of these, and many more challenges, analytic organisations must balance thoroughness and rigour against expediency and achievability.

However, every analytic sacrifice comes at a cost, and one of the costs is confidence in our analysis. We cannot have the same confidence in the analysis performed quickly, by a single analyst, using only the information that comes readily to their mind, and based only on their intuitive logic and reasoning, as we would in the analysis performed by a team of diverse analysts over several days, using a collaboratively generated information base, and performed using one or more structured analytic techniques. The judgments coming from both might be the same, but we would have greater confidence in the judgment generated by the team than we would in that from the single analyst.

Similarly, we would have greater confidence in a judgment that explored the underpinning reasoning in-depth, including the potential objections and shortfalls to the reasoning, than we would in a judgment that only explored the most prominent supporting reasons and information. Or, we would have greater confidence in analysis that explored and weighed a variety of alternatives to determine the most likely judgment than we would in analysis that only explored a single judgment to identify the information that supported it.

Our analytic confidence should, therefore, be based on the process used to generate our judgments, not the information and support underpinning them. We, as intelligence professionals, should honestly calibrate the effort, thoroughness, and rigour we are able to apply to our analytic problems, given the time, information, and systemic challenges we face, and express it honestly to our clients. It is our duty, as analytic professionals, to honestly inform our clients and consumers when we are forced to take analytic shortcuts to meet deadlines or use a partial information base due to technical or time constraints. Being honest with ourselves about the confidence we have in our own analysis will probably help our clients have more confidence in us. Criteria such as the hypothetical example in *Tables 2-7* would help us maintain both rigour and transparency to accomplish this.

Table 2	**Thoroughness of information base available and used in analysis**
Value	**Criteria**
0	Very little information available or used to generate analytic judgments, or all information available for analysis comes from a very small number of sources, or information used for analysis was primarily based on the performing analyst's recollection of reporting and developments.
1	Information used to generate the analytic judgments came from a small number of sources that all generally support the analytic judgments, or information used to generate the analytic judgments came primarily from a single intelligence collection method.
2	Information from many sources of information and/or multiple intelligence collection methods were used to generate the analytic judgments, that all generally support the analytic judgments.
3	Information from many sources of information and/or multiple intelligence collection methods were used to generate the analytic judgments, including information that both supports and contradicts the analytic judgments.
4	Many pieces of information from many sources from multiple intelligence methods were used to generate analytic judgments, addressing multiple variables and perspectives on the analytic problem or questions, including information that both supports and contradicts the analytic judgments.

Table 3	Structure and methodologies applied to performing analysis
Value	**Criteria**
0	Analytic judgments were generated by a single analyst without using structured analytic techniques or methodology
1	Analytic judgments were generated using a structured analytic technique performed independently by the analyst, or analytic judgments were generated by a team of analysts using intuitive logic and reasoning in analytic discussion internal to the analytic team
2	Analytic judgments were generated by a team of analysts using intuitive logic and reasoning in analytic discussion involving multiple analytic teams from separate divisions or organizations, or analytic judgments were generated using structured analytic technique(s) with the assistance of a senior analyst or intelligence officer from within the same division to help ensure methodology and thoroughness
3	Analytic judgments were generated using structured analytic technique(s) with assistance from a senior analyst or intelligence officer from a separate division or organization to help ensure methodology and thoroughness
4	Analytic judgments were generated using structured analytic technique(s) under the guidance of a professional facilitator to help ensure applicability, methodology, and thoroughness

Table 4	Size and diversity of the team participating in analysis
Value	**Criteria**
0	Analytic judgments were generated by a single analyst, or from a small number of analysts on the same team
1	Analytic judgments were generated by a group of analysts collaborating among several teams in the same analytic division or section
2	Analytic judgments were generated by a group of analysts collaborating among several teams in different divisions or sections of the same organization, or analytic judgments were generated by a group of analysts collaborating among several teams between two or three organizations that have similar analytic focuses and purposes
3	Analytic judgments were generated by a group of analysts collaborating among several teams between two or three organizations with varying focuses and purposes, or analytic judgments were generated by soliciting and synthesizing analytic input from multiple teams from several organizations with varying focuses and purposes
4	Analytic judgments were generated by a group of analysts collaborating among several teams between several organizations with varying focuses and purposes, or analytic judgments were generated using structured, facilitated analytic input from multiple teams from several organizations with varying focuses and purposes

Table 5	Depth and thoroughness of the reasoning used in analysis
Value	**Criteria**
0	Analytic judgments generated using intuitive logic and reasoning
1	Analytic judgments generated by a group of analysts using analytic discussion to develop logic and reasoning, or analytic judgments generated by a single analyst using an outline or argument map to externalize their logic and reasoning
2	Analytic judgments generated by multiple analysts using an argument map to externalize their logic and reasoning
3	Criteria from #2 above, and the logic and reasoning is well-developed, including supporting reasoning, objections to the reasoning, and rebuttals to the objections
4	Criteria from #3 above, and the logic and reasoning is thoroughly developed from multiple perspectives, including less-likely or alternative scenarios or hypotheses

Table 6	Time and effort devoted to performing analysis
Value	**Criteria**
0	Due to the nature of the analytic problem there was little time available or necessary to perform analysis, or due to the intelligence client's decision-making requirements there was little time available to perform analysis
1	Analysts had to curtail their research, analysis, and/or production in order to meet production or decision-making requirements
2	Analysts had to make time and effort trade-offs between research, analysis, and production to meet production and/or decision-making deadlines
3	Production and decision-making deadlines allowed ample time for analysts to perform research and analysis
4	There were no production deadlines or decision-making pressures on the analytic problem, and the analysts were able to thoroughly research, analyze, and collaborate when generating their analytic judgments

Table 7	Examination of alternative hypotheses and/or challenges to main judgments
Value	**Criteria**
0	Little-to-no examination or development was performed on challenges or alternative hypotheses to the main analytic judgments, or the analytic judgments were determined to have few credible challenges or alternatives
1	Challenges and/or alternative hypotheses with clear, explicit information support or historical precedence were examined, or challenges and/or alternative hypotheses that were logically obvious were examined
2	Analytic judgments were generated by an examination of two or more competing hypotheses or interpretations of the available information, or analysts identified and developed one or more challenges or alternatives to the main analytic judgments as part of the analysis and production process
3	Analytic judgments were generated using structured analysis to identify, develop, and explore several competing hypotheses, or analysts identified, developed, and explored multiple challenges or alternatives to the main analytic judgments as part of the analysis and production process
4	Main and alternative analytic judgments were thoroughly identified, developed, explored, and compared using structured analytic techniques designed or intended to aid this effort

However, the question remains: from where does our certainty come? How do we calibrate and assign certainty to our analytic judgments?

I propose the calibration to be based on measurable movement away from analytic uncertainty and toward certainty.

Referring back to *Figure 1* above, an analyst would never be able to generate judgments or assessments under 50% *analytic certainty*, since any judgment under 50% is movement toward certainty *against* the event or situation occurring. Therefore, we propose approaching certainty calibration on a "50% to 99%" scale, with 50% being the most uncertain and 99% being the most certain. (Analysis would not provide probability estimates for 100% certainty either way.) (*Figure 2*)

Figure 3: *Certainty calibration on a 50% to 99% scale*

For example, let's assume an analyst believes it is very unlikely the adversary will attack – in the 6%-20% probability range on the scale from *Figure 1*. Breaking this apart, this 6%-20% chance the adversary will attack actually means the analyst is 80%-94% certain the adversary will do something else – *anything* else – but not attack. Or, if an analyst assesses that Actor X is unlikely to take a particular action, in the 21%-40% probability range, the *actual* meaning is that the analyst is 60%-79% certain the actor will take any other action, but not *this* action.

In this model, we would convert low probability – the 1% "almost certainly not" through the 49% "roughly even chance" – to positive certainty. For example, an assessment of "Actor X will almost certainly not perform action Y" – in the 1%-5% portion of *Figure 1*'s scale – is converted to a 95%-99% certainty the actor will perform any action *other* than Y, but not Y, or 95%-99% certainty that a different actor might perform action Y, *but not* Actor X. Or, an analyst could judge "Event X is unlikely to happen" (21%-40%), which we would convert to 60%-79% certainty that anything *other* than the event will happen, *but not* Event X.

So, to be clearer on what we are actually assessing, our analysis should be designed to move us away from the greatest uncertainty toward greater certainty, and our calibration of this certainty should come from where on *Figure 2* our judgments fall.[4]

However, the challenge remains: how do we calibrate *analytic certainty* to place our judgments on this scale?[5]

The leading methods of calibrating the *probability* of qualitative judgments are Bayesian probability, the Delphi Method, and expert opinion. Bayesian probability, essentially assigning numeric value to the variables then putting those numbers into a formula to generate a probability percentage, is one of the more rigorous methods to calibrate probability. However, it is one of the most complicated, confusing, and labour-intensive methods to do so. So, very few analysts calibrate using Bayesian probability.[6] [7] [8] In fact, from my 30 years of experience in intelligence, few analysts are able to do so; most of us believed Bayesian probability is best performed by and best left to computers.

The Delphi method is much more achievable. In this method, a group of analysts is tasked to calibrate the probability of a judgement to reach a consensus through several rounds of questionnaires and anonymous feedback. The idea behind this method is that the aggregated anonymous expertise would gravitate toward a more accurate estimation. This method is much

[4] Note: this is for calibrating certainty purposes only. For analytic communication to decision-makers, the current probability terms and matrices establish mutual understanding and should continue to be used for analytic production.

[5] Calibrating the certainty of qualitative assessments will never be exact, as these calculations depend on unknown variables for which we make subjective calculations.

[6] K.J. Wheaton, "Teaching Bayesian Statistics to Intelligence Analysts," Journal of Strategic Security 1, no. 1 (November 2008): n.p., https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=1035&context=jss. Digital Commons+1

[7] "Top 5 Intelligence Analysis Methods," Sources & Methods (blog), (December 4, 2008): https://sourcesandmethods.blogspot.com/2008/12/top-5-intelligence-analysis-methods.html. sourcesandmethods.blogspot.com

[8] Karvetski, Christopher W., David R. Mandel, and Daniel Irwin. "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation." *Journal of Behavioral Decision Making* 33, no. 5 (2020): 658–671. https://doi.org/10.1002/bdm.2170.

more approachable than Bayesian probability, but it takes time – and time is one of the scarcest resources in intelligence analysis. The intelligence production space between information and decision is growing narrower by the day. This method also carries two significant risks. The first risk is groupthink, where members of the group gravitate toward agreement rather than considering outlying opinions. Second, it risks the analytic conclusions becoming "lowest common denominator" judgments. This means judgments the group can agree with can become too weak or ambiguous to be of any significant use to a decision-maker.

This leaves expert opinion as the most common method to calibrate qualitative probability. It is generally much faster than the Delphi method since it is usually a single expert or a small group of experts involved, and much simpler than the Bayesian method since it is based on expert opinion, not dependent on complicated mathematics. However, there are several problems with expert opinion. First, it is rarely, if ever replicable. Every analyst will have a different opinion, based on what each considers to be the key variables and information. Additionally, there is little way for an expert to justify why they assign a given probability. Seldom is there something an analyst can point to and say "*This* is why I believe the judgment to be 'very likely' rather than 'likely'." In addition to confounding a transparent calibration of probability, this can preclude constructive dialogue on probability estimates altogether; if analysts do not agree on a probability estimate and do not use the same information to underpin the calibration or identify specific criteria, methods, and weighting they used to generate each of their probability estimates, they cannot easily work together toward a truer probability estimate. Finally, expert opinion rarely, if ever, has an audit trail. There is no record of how or why an analyst estimated the probability of an event to be 'likely' rather than 'very likely,' even if such a record were possible.

This leaves us in an analytic conundrum. We are tasked with rigour, transparency, and accountability in our analysis, with no practical way to achieve this in the way we calibrate the probability of non-quantifiable judgments and analytic conclusions.

What we need is a way to apply rigour to calibrating the probability – or, in our model, the *certainty* – of our qualitative judgments, in a way that is transparent, replicable, auditable, accountable, and most of all, approachable.

Using Argument Maps to Calibrate Certainty

Using argument maps for this may be a solution. I have long used argument maps as a production tool for intelligence analysis. Before using a traditional outline to structure an analytic argument for production, I recommend that my students build argument maps to fully detail analytic arguments and reasoning, objections, assumptions, alternatives, and logic. By inputting references to their source reporting and evidence, these argument maps can effectively detail the strengths and weaknesses of their support base. With just another short step we can use a well-crafted argument map to provide a more transparent, more rigorous, and more accountable certainty calibration for qualitative judgments.

Schrag et al. had a similar approach in 2016 with their FUSION model of computerised probabilistic argument mapping,[9] but their model involved extracting argument maps from text-based analytic products, then inputting them into a computer for algorithmic Bayesian analysis and probability estimation.

For our purposes, though, let's examine how an analyst, without Bayesian computer assistance, can do this.

I will not go into detail on how to create an argument map; this is a skill I believe every analyst should have, and there are several resources openly available on the internet to learn this skill. We are starting from the assumption that the analyst can use an argument map to structure their analytic conclusions, the logic and reasons supporting these conclusions, the objections to their conclusions, and the rebuttals to those objections. The better and more complete an analyst maps out their analytic arguments, the more complete their logic, reasoning, and evidence will be, helping ensure better and more accurate certainty to their calibrations.

While a fully-detailed argument map is the ideal, it is not a requirement; this step can be performed in whatever time and depth the analyst or analytic team can devote to it. Decision-makers must often make their decisions under extreme time pressure, which means that analysts are under equal time pressure to deliver relevant insights that their clients can use for these decisions. Analysts can quickly draft an argument map based on their subject-matter expertise and the information they have readily available, and use it in this process. It would not be as accurate as a fully developed argument map, nor would the analyst be able to justify the same analytic confidence in such a quickly-formed map, but it would externalise the analyst's thinking and apply better rigour to their logic and reasoning, giving a better foundation to their certainty calibrations. These maps could then be used later for more detailed analysis to improve both accuracy and confidence.

To get started, the analyst would use a strength scale for each of the reasoning, objection, and rebuttal nodes. We choose to use a 1-10 scale, 1 being the weakest reason for the node's claim or reason, and 10 being the strongest, but analysts could use any scale they wish as long as the scale and criteria for judging the strength of each node is consistent between all nodes to establish a consistent strength calibration throughout the map.

The strength of each node would be calibrated using a clear set of criteria applied to the supporting reason or logic in each node. Again, this should remain consistent throughout the map, and should include clear guidance on calibrating the node's strength using these criteria to help facilitate objective consideration of each node. For example, analysts could calibrate each node's strength by evaluating:

[9] Robert Schrag, Joan McIntyre, Melonie Richey, Kathryn Laskey, Edward Wright, Robert Kerr, Robert Johnson, Bryan Ware, and Robert Hoffman, "Probabilistic Argument Maps for Intelligence Analysis: Completed Capabilities," Computational Models of Natural Argument, (2016): https://ceur-ws.org/Vol-1876/paper07.pdf

- Number of and corroboration among sources of information underpinning the node
- Source/sourcing base reliability (evaluating the reliability of the *provider* of information)
- Information reliability (e.g. verifiability, consistency with previously known/believed information)
- Potential for deception
- Novelty or familiarity of the information's background and environment
- Soundness of the logic
- The criticality of the assumptions underpinning the claim

For example, *Table 8*[10] below is a sample of how we might transparently calibrate the strength of each node's claim.

Table 8: *Calibration of the strength of each node's claim*

Strength	Number and corroboration among underpinning sources of information	Reliability of the provider(s) of information	Reliability of the information or evidence	Potential for deception or mis-identification	Novelty or familiarity of the information's background and environment	Soundness of the logic underpinning this claim	Criticality of assumptions underpinning the claim	Total (Method 1: Add the total strength values below and divide by 10 to get the total strength of the claim or information)
1	Only one or two sources provided this information	Provider(s) of information are usually unreliable or provide information probably intended to deceive	The information may be correct but is contradicted by a large body of evidence or would be a clear, radical reversal from historical precedent	Information or claim is almost certainly deception or is regularly mis-identified	The background, environment, or context is very new and/or is highly volatile or dynamic	The information base is missing key information that makes the logic critically dependent on correct interpretation and inference	The claim is an assumption that is very plausibly incorrect	
2	Very few sources available with this information, but they generally corroborate each other	Provider(s) are generally unreliable or have a clear bias and a history of providing information with mistakes, omissions, or characteristics probably intended to influence	The information may be correct, but is contradicted by several sources or would be a clear change from historical precedent	Information or claim is probably deception or has been mis-identified many times in the past	The background, environment, or context is very new and/or still changing rapidly	There are many gaps in the information base that make the logic very dependent on correct interpretation and inference	The claim is an assumption that has a realistic possibility of being incorrect	

[10] Note: this is an example for discussion, and not to be considered a final, definitive, or authoritative set of criteria to calibrate the strength of a claim. This is just one of many ways to calibrate the strength of a claim.

3	Few sources of information and significant disagreement among sources	Provider(s) have a clear bias and history of providing information with mistakes that are probably intentional	The information is corroborated by few sources but contradicted by most other sources, or contradicts most interpretations of historical precedent	Information or claim is plausibly deception or has been mis-identified several times in the past	The background, environment, or context is new and/or still changing and developing unpredictably	There are many gaps in the information base and the logic depends significantly on correct interpretation and inference	The claim is an assumption, but is unlikely to be incorrect	
4	Many sources available and there is significant disagreement among them	Provider(s) have clear bias and history of providing information with mistakes that are probably unintentional	The information is corroborated by several sources but contradicted by many other sources, or is based on a misinterpret-tation of historical precedent	Information or claim may be deception or has occasionally been mis-identified in the past	The background, environment, or context is generally unfamiliar and/or still changing and developing	There are multiple gaps in the information base and the logic often depends on correct interpretation and inference	The claim is a strong assumption with little potential to be incorrect	
5	Few sources available and the more reliable sources generally corroborate each other, but there is significant disagreement in the body of information	Provider(s) have history of providing information with known bias but mistakes are probably unintentional	The information is corroborated by many sources but there is significant disagreement among all sources, or based on a clearly biased interpretation of historical precedent	Information or claim is unlikely to be deception or has generally been identified correctly in the past	The background, environment, or context is generally unfamiliar but seems consistent with historical examples from other situations	There are multiple gaps in the information base and the logic sometimes depends on correct interpretation and inference	The claim is a very strong assumption with almost no chance of being incorrect	
6	Many sources available and some are reliable; sources generally corroborate each other, but there is significant disagreement in the body of information	Provider(s) have history of providing information with mistakes or bias that is probably unintentional	The information is corroborated by many sources but there are still many sources that disagree, or is based on a biased interpretation of historical precedent	Information or claim is probably not deception or is often identified correctly	The background, environment, or context is generally unfamiliar but is demonstrably consistent with historical examples	There are some gaps in the information base and the logic sometimes depends on correct interpretation and inference	The claim depends on one or more assumptions that could plausibly be incorrect	

7	Few sources available and most corroborate each other, but there is some disagreement	Provider(s) have history of providing generally reliable information, with some mistakes or bias	The information is corroborated by many sources but there are still a few sources that disagree, or may be based on a biased interpretation of historical precedent	Information or claim is very likely not deception or is usually identified correctly	The background, environment, or context is generally familiar and seems consistent with historical examples	There are few gaps in the information base and the logic is generally sound	The claim depends on one or more assumptions that are unlikely to be incorrect	
8	Many sources available and most corroborate this, but there is some disagreement	Provider(s) have history of providing usually reliable information, with a few mistakes or bias	The information is corroborated by most sources or is based on generally accepted interpretation of historical precedent, but there are still a small number of sources that disagree	Information or claim is almost certainly not deception or is normally identified correctly	The background, environment, or context is generally familiar and demonstrably consistent with historical examples	There are few gaps in the information base and the logic is sound, with few logic gaps or uncertain inferences	The claim depends on one or more assumptions that are unlikely to be incorrect	
9	Many sources available and all or nearly all corroborate each other	Provider(s) have history of providing reliable information	The information is corroborated by nearly all sources or is based on objective historical precedent	Information or claim comes from trusted, non-deceptive sources or is almost always identified correctly	The background, environment, or context is generally understood and consistent with historical examples	There are almost no gaps in the information base and the logic is sound, with very few logic gaps	The claim depends on one or more assumptions that are very unlikely to be incorrect	
10	Confirmed through testing, observation, or historical occurrence	Provider(s) have history of providing objective, confirmed information	The information is tested and verified	There is almost no chance the information is deceptive or mis-identified	The background, environment, or context is well understood and very consistent with historical examples	There are almost no gaps in the information base and the logic is sound, well-structured, and supported	The claim does not depend on assumptions, or depends on assumptions that are extremely unlikely to be incorrect	
TOTAL								**Claim Strength:**
Method 2: Add the total strength values above and divide by 7 to get the total strength of the claim or information								

Objection nodes would capture contradictory information, contradictory interpretations of data, challenges to the underpinning assumptions, and plausible alternatives to the supporting reason or conclusion. However, a critique of the reasoning node's information base or strength would *not* be a valid objection, since that weakness should be captured in the reasoning node itself.

Similar to objection nodes countering reasoning nodes, rebuttal nodes counter objections using the same process and criteria.

From here we would generate separate *Supporting Reasoning* and *Objection* percentages. The calibration would start with the most fully-detailed argument map feasible, identifying the logical reasoning, objections, and rebuttals applicable to the conclusion, with as much information and evidence, assumptions, and logical inferences and interpretations explicitly identified as possible. Analysts would then use explicit, objective calibration criteria to assign a strength value to each node; for example, the analyst could use the 1-10 scale identified above, with 1 as extremely weak and 10 as nearly irrefutable.[11]

This will quite possibly be the most difficult task in this process. Analysts would need to evaluate and calibrate each node's strength *completely independently* from all other nodes. Every node would need to be evaluated and calibrated according to its own merits.

Note: This would still be a subjective calibration of strength. Currently, I believe it is impossible to objectively calibrate such strength completely. However, using clear criteria in a transparent process would enable greater discussion and dialogue on the merits and critiques of each node, as well as enabling replicability in the process of calibrating strength and certainty.

Once the strength of each node in the argument map is established, we can calibrate movement from the greatest uncertainty (50%) toward the greatest certainty (99%).

For this, I use a formula to calculate the reasoning both for and against an analytic judgment, which calculates the difference between a judgment's supporting reasoning and the objections to that reasoning, and adding it to 50%, our starting point of greatest uncertainty, to give us a certainty percentage.

Written out, the formula would be: $\frac{1}{2}(S\% - O\%) + 50\% = C\%$, where:

- S = Supporting Reasoning for the analytic conclusion (using S to differentiate it from R, used later)
- O= Objection to the reasoning or conclusion
- C = Certainty

Supporting Reasoning percentage = S Actual / S Maximum. For example, let's assume there are 10 reasoning nodes underpinning the analytic conclusion, and we are using a 1-10 scale of strength. The analyst has already calibrated each node's strength independently; we now add

[11] The collaboration and rigour employed to generate the argument map forming the basis for probability calibration would likely be a key factor in estimating analytic confidence. For example, an argument map quickly generated by a single analyst or team might justify very low analytic confidence, while an argument map developed over several days or weeks in collaboration with multiple teams approaching the analysis from multiple fields of expertise might justify very high analytic confidence.

these together to determine the actual strength – S Actual. For example, we will assign an arbitrary 64 to be our S Actual. For our 10 reasoning nodes, the maximum strength would be 100. So, our Supporting Reasoning percentage would be calculated as 64 / 100 = 0.64, or 64%.

Objection percentage = [O Actual - R Actual] / O Maximum. Using the same process as above, the analyst would determine the actual strength of objections and the actual strength of rebuttals (R), as well as the maximum strength of objections. For the Objection percentage, though, we are interested in the strength of objections *after* they are rebutted, so we subtract R Actual from O Actual before dividing by O Maximum. For example, if 7 of our 10 simulated reasoning nodes have objections, 5 of which are rebutted, we could simulate an O Actual of 54, an R Actual of 42, and our O Maximum would be 70. So, our calculation is (54 – 42) / 70 = 0.1714, or approximately 17%.

For our certainty calculation, C, we are attempting to find our movement from uncertainty toward certainty, so our starting point is 50% - the highest level of uncertainty. So, we need to halve our strength percentage to identify the movement from 50%, and add it to that 50%. Applying this to our example, ½ (64%-17%) + 50% = C%, or 23.5% + 50% = 73.5%, which becomes our *analytic certainty*.

Argument Mapping Examples

Let's use several sample argument maps to put this into practice and help our conceptualisation. For consistency and comparison, we will use the same map but re-assign the strength of the nodes. These are *Figures 4, 5, 6*, and 7; *Figure 4* depicts an argument of average strength, *Figure 5* depicts a stronger argument, *Figure 6* depicts a weaker argument, and *Figure 7* depicts an extraordinarily weak argument in which the objections are stronger than the reasoning.

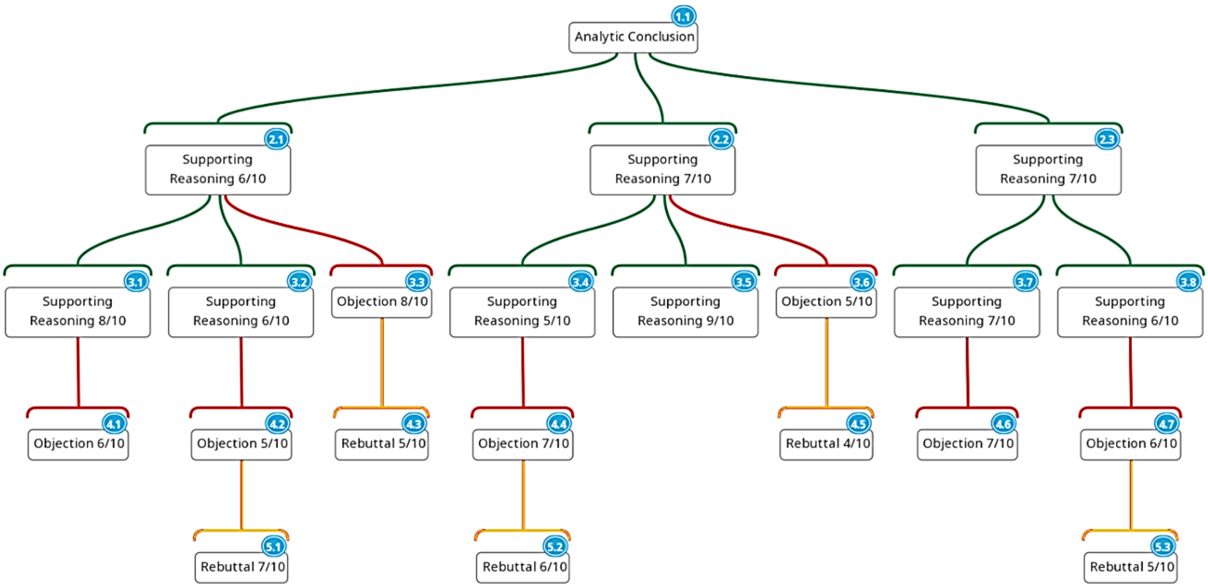


Figure 4: *Supporting Reasoning: 57/90 ≈ 0.633 ≈ 63%. Objection: 17/70 ≈ 0.242 ≈ 24%. Certainty: ½ (63% - 24%) + 50% ≈ 70%*

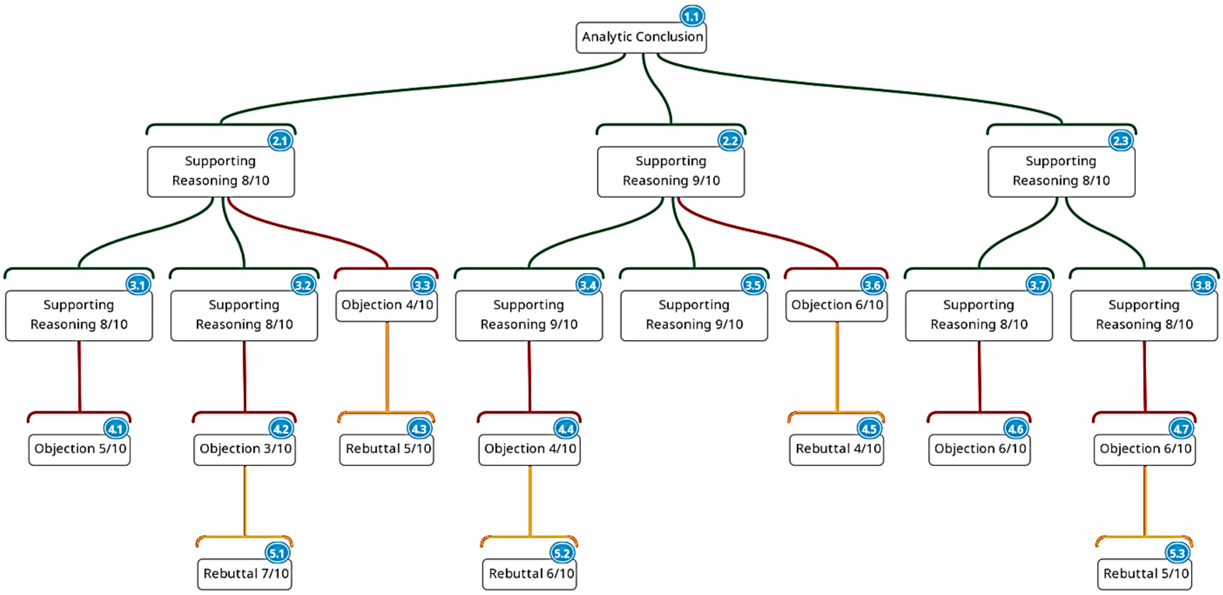


Figure 5: *Supporting Reasoning: 75/90 ≈ 0.833 ≈ 83%. Objection: 7/70 ≈ 0.1 ≈ 10%. Certainty: ½ (83% - 10%) + 50% ≈ 87%*

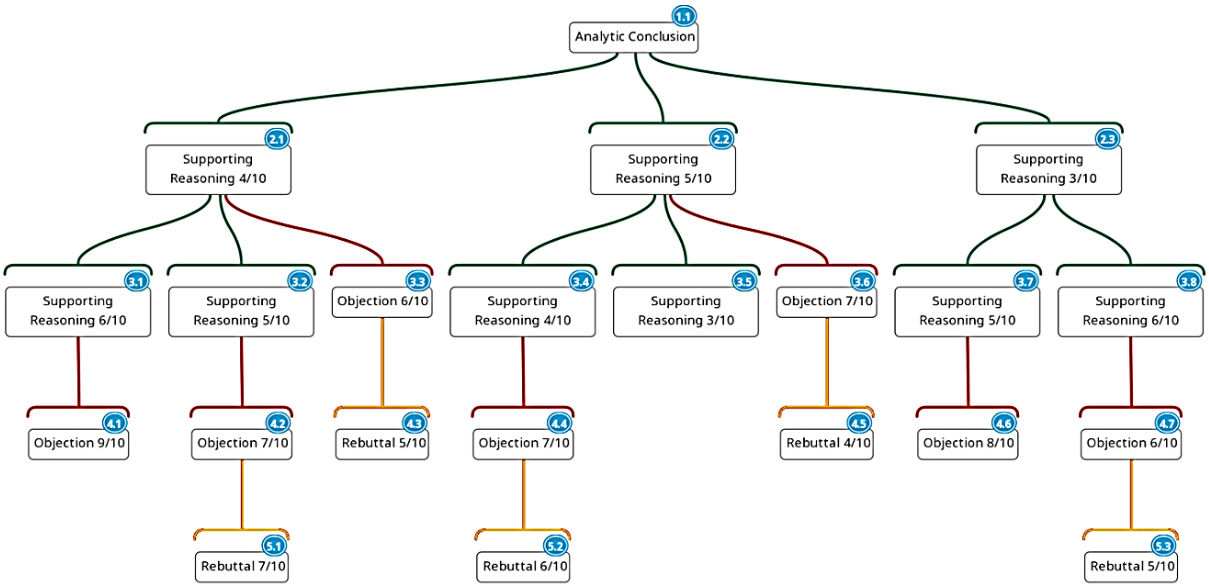


Figure 6: *Supporting Reasoning: 41/90 ≈ 0.455 ≈ 46%. Objection: 23/70 ≈ 0.328 ≈ 33%. Certainty: ½ (46% - 33%) + 50% ≈ 57%*

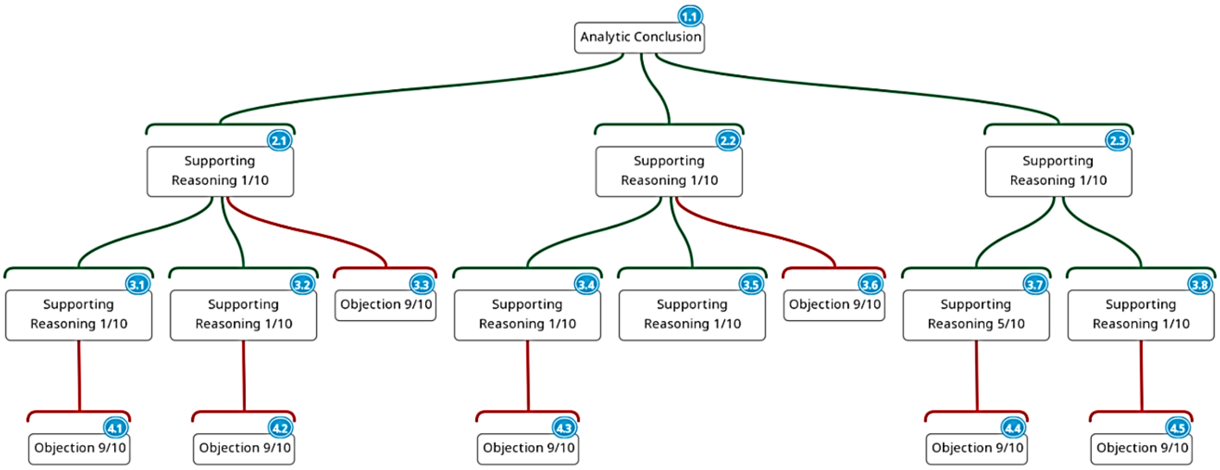


Figure 7: *Supporting Reasoning: 9/90 ≈ 0.1 ≈ 10%. Objection: 63/70 ≈ 0.9 ≈ 90%. Certainty: ½ (10% - 90%) + 50% ≈ 10%*

I have rounded off the percentages deliberately; since weighting the nodes is still subjective, the certainty percentages cannot be exact nor quantifiably justified, so we still want to put them in certainty *ranges* for our final estimates to accommodate the subjectivity. However, they get us closer to a "true" certainty calibration than expert opinion alone. Also, we have deliberately made these maps simple to increase their understandability; more realistic maps should be consistent with these findings, but would be far more difficult to conceptualise in an example.

Each of these argument maps has 9 reasoning nodes and 7 objection nodes, giving us the 90 and 70 maximum strengths for each. From there we calculated the actual strength of each argument, and calibrated each argument's resulting certainty. These samples are contrived, but they illustrate how the process would work.

As we can see from these samples, certainty in this model directly correlates to the strength of supporting information and argumentation for the analytic conclusion. Stronger reasoning and weaker objection result in greater certainty, and weaker reasoning and stronger objection result in lower certainty. And, in cases where the strength of objections is stronger than the strength of reasoning, such as *Figure 7*, the sub-50% certainty suggests to us that our analytic conclusion is probably incorrect and should be revisited. (Note: this does not mean its opposite is correct; just that the judgment being made is probably incorrect.)

There could be cases in which one branch of an analytic argument has significant objections that *mathematically* would undermine the entire argument's strength. For example, in this argument map (*Figure 8*), the middle branch of this argument weakens the entire argument, when it may only be this line of reasoning that is faulty.

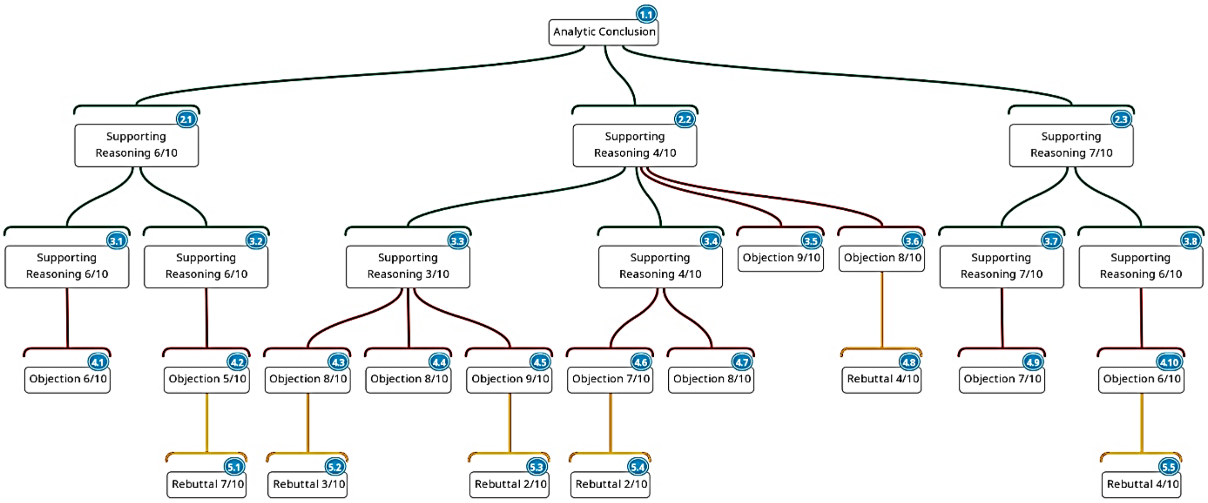


Figure 8: *Supporting Reasoning: 49/90 ≈ 0.544 ≈ 54%. Objection: 59/110 ≈ 0.536 ≈ 54% Certainty: ½ (54% - 54%) + 50% ≈ 50%.*
Middle Branch Only – Supporting Reasoning: 11/30 ≈ 0.366 ≈ 37%. Objection: 46/70 ≈ 0.657 ≈ 66%. Certainty: ½ (37% - 66%) + 50% ≈ 35.5%.
Left and Right Branches Only – Supporting Reasoning: 38/60 ≈ 0.633 ≈ 63%. Objection: 13/40 ≈ 0.325 ≈ 33%. Certainty: ½ (63% - 33%) + 50% ≈ 65%.

The middle line of reasoning is weak, likely faulty, and in this example it undermines and weakens the entire argument. The temptation would probably be to remove this line of reasoning from the argument and proceed with only the left and right branches to assign a ~65% certainty to the main conclusion. However, the middle branch probably represents an important alternative to the main judgment that may be ignored by simply removing this line of reasoning.

Instead, as with the weak argument from *Figure 3*, in this case the analyst probably should revisit their analysis to include a deeper exploration of the objections or alternatives brought out in this line of reasoning, to help ensure analytic rigour and integrity.

To demonstrate, let's look at an argument in a map developing the claim, 'the use of structured analytic techniques improves analytic results.'[12] (*Figure 9*)

[12] This argument is for demonstration only. It is not a complete argument, but has the basics we would expect to find in an argument. The weighting is also approximate, and for demonstration purposes only. It does not reflect a rigorous examination of the argument or the evidence underpinning it.

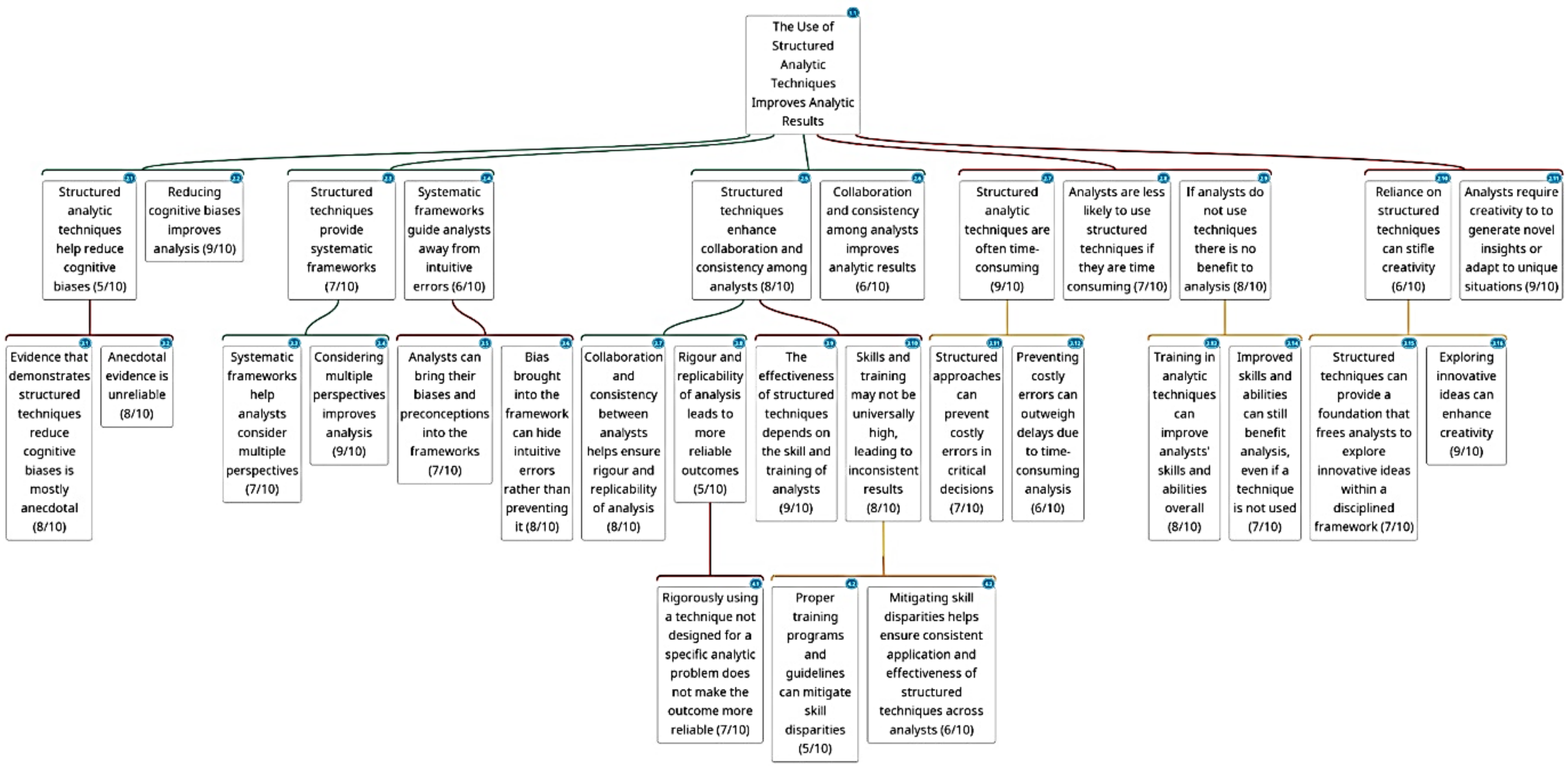


Figure 9: *Supporting Reasoning: 68/100 ≈ 68%. Objection: 39/120 ≈ 32.5% Certainty: ½ (68% - 32.5%) + 50% ≈ 67.75%*

In this argument map, there are several reasons supporting the claim, several objections to it, and several rebuttals to the objections.

Communicated Probability

Once analysts determine the *analytic certainty* of their judgments, they can revert back to their organisation's accepted *communicated probability* terminology to communicate the results to their clients and consumers with greater specificity, confidence, and utility that better aids in decision-making.

For example, let's return to our simulated "very unlikely the adversary will attack" – in the 6%-20% probability range on the *communicated probability* scale. To address the *analytic certainty* for this, the analyst will have explored the information base, the drivers for and against conflict, etc and come to the analytic conclusion, with a high degree of certainty, that the adversary would be unwilling or unable to attack. There is no indication what the adversary *would* do at this point – that would be the result of separate analysis – but attacking is something the adversary will very likely try to avoid. The *analytic certainty* of this judgment is therefore "very certain," in the 80%-94% range.

But in our simulation, the client has asked, "Will the adversary attack?" To answer the question we can turn the judgment back around on the probability scale from *Table 1*, placing it back in the 6%-20% range, so the *communicated probability* would be, "It is very unlikely that the adversary will attack."

This process is still subjective at its core, since it will be up to the analysts to develop a complete argument map and assign the strengths to each node in the map. But, as we can see from the above examples, by relating analytic certainty to an argument's reasoning, objections, and rebuttals, we can enable analysts to identify specific reasoning to justify their certainty calibration and increase their calibration's transparency and replicability.

Additionally, calibrating certainty using this method would enable greater examination and discussion of the analytic argument, and give teams or managers greater ability to audit and add to the analysis and certainty calculations. This would likely result in more complete analytic arguments, more accurate calculations and calibrations, and more confidence in our assessments. Also, new information can be incorporated into the argument map as analysts receive it, enabling them to regularly update and confidently identify increasing, decreasing, or consistent certainty of their assessments.

Even more, explicitly identifying the information and assumptions they rely on to calibrate certainty – both supporting and contradicting their analytic conclusions – enables analysts to work with their collection managers and assets to gain information that would give them greater certainty. This could also help them identify indicators of significant changes to the *communicated probability* of their assessments, which clients and decision-makers can then use for planning and operations, as appropriate.

Conclusion

Calibrating probability in intelligence analysis remains a persistent challenge, with traditional methods often lacking transparency and rigor. By distinguishing between *communicated probability* (what is shared with clients) and *analytic certainty* (rigorously calibrated and assigned based on evidence and reasoning) analysts can improve the clarity and reliability of their assessments.

The use of argument maps offers a structured, transparent way to assign probabilities to qualitative judgments. This method allows analysts to systematically evaluate supporting reasoning, objections, and rebuttals, applying clear criteria to each node. While calibration remains partially subjective, this approach increases accountability, facilitates discussion, and enables regular updates as new information emerges.

Ultimately, this process empowers analysts to justify their assessments more clearly, supports more robust dialogue within teams, and provides decision-makers with more reliable and actionable intelligence.

Bibliography

Karvetski, Christopher W., David R. Mandel, and Daniel Irwin. "Improving Probability Judgment in Intelligence Analysis: From Structured Analysis to Statistical Aggregation." Risk Analysis 40, no. 5 (2020): 1040-1057. https://doi.org/10.1111/risa.13443

Office of the Director of National Intelligence. Intelligence Community Directive 203: Analytic Standards. Washington, D.C.: ODNI, January 2, 2015. https://www.dni.gov/files/documents/ICD/ICD_203.pdf

R. Schrag, Joan McIntyre, Melonie Richey, Kathryn Laskey, Edward Wright, Robert Kerr, Robert Johnson, Bryan Ware, and Robert Hoffman. "Probabilistic Argument Maps for Intelligence Analysis: Completed Capabilities." In Proceedings of the 16th Workshop on Computational Models of Natural Argument (Vol. 1876), edited by Floris Bex, Floriana Grasso & Nancy Green, 34-39. New York, USA: IJCAI Co-located, 2016. https://ceur-ws.org/Vol-1876/paper07.pdf

Wheaton, Kristan J., Jennifer Lee, and Hemangini Deshmukh. "Teaching Bayesian Statistics to Intelligence Analysts: Lessons Learned." Journal of Strategic Security 2, no. 1 (2010): 39-58. https://doi.org/10.5038/1944-0472.2.1.3

. Accessed October 20, 2025. https://digitalcommons.usf.edu/jss/vol2/iss1/3

Intelligence Analysis. "Explaining Uncertainty in UK Intelligence Assessment." GOV.UK. March 24, 2025. https://www.gov.uk/government/publications/explaining-uncertainty-in-uk-intelligence-assessment/explaining-uncertainty-in-uk-intelligence-assessment

"Top 5 Intelligence Analysis Methods." Sources & Methods (blog), December 4, 2008. Accessed October 20, 2025. https://sourcesandmethods.blogspot.com/2008/12/top-5-intelligence-analysis-methods.html.

JOURNAL *of* EUROPEAN *and* AMERICAN INTELLIGENCE STUDIES

An international peer-reviewed journal

Sponsored by the **Research Institute** *for* **European** *and* **American Studies – RIEAS**

Under the editorial direction of the

School of Law and Government, Dublin City University

and the **International Centre for Policing & Security, University of South Wales**

CALL FOR PAPERS

The *Journal of European and American Intelligence Studies* (*JEAIS*) is seeking papers focusing on the field of intelligence and related areas of study and practice, such as terrorism and counter-terrorism, domestic and international security, geopolitics, and international relations. The papers should contain or examine original research on the broader European and American practice and study of intelligence, but also highlight intelligence themes from other regions of the world, to include Africa and Asia, as well as Oceania. Submissions will undergo rigorous and highly selective screening, as well as an exhaustive review process. Particular attention will paid by the editors to papers that discuss and challenge established or emerging ideas, address existent knowledge gaps by advancing new knowledge on intelligence-related topics, and examine intelligence scholarship beyond traditional communities of research.

Relevant Topics Include:

international intelligence cooperation ▪ intelligence operations ▪ intelligence analysis ▪ inter-agency cooperation ▪ practitioner-policymaker relations ▪ military intelligence ▪ politicization of intelligence ▪ counterintelligence ▪ intelligence and sub-state violence ▪ intelligence, race, and gender ▪ intelligence and privacy ▪ intelligence and democracy ▪ intelligence ethics

Abstracts of up to 250 words may be submitted to:
secretary@rieas.gr and *christian.kaunert@southwales.ac.uk*

Deadline for Abstract Submissions: Monday February 16, 2026
Deadline for Paper Submissions: Monday May 4, 2026

Instructions for Authors

- Submitted manuscripts must be maximum 8,000 words, excluding a 250-word abstract (required) and any footnotes, as well as references. Manuscripts that exceed the word limit will be automatically rejected and returned to their authors.

- Manuscripts will be accepted for submission and evaluation with the understanding that their content is unpublished, original work by their authors, and have not been submitted for publication elsewhere.

- All accepted manuscripts and artwork become the property of the publisher, which is the Research Institute for European and American Studies (RIEAS).

- The entirety of manuscripts, including title page, abstracts, tables, legends, and references, should be typewritten and submitted in a Word-type file. No portable document format (PDF) documents will be accepted.

- Submissions should be 1.5-spaced and use Times New Roman size 12 as their standard font.

- All margins should be at least one inch in length, and all pages should be numbered consecutively throughout the manuscript.

- Titles must be as brief and clear as possible. On the title page, please include full names of authors, their academic and/or other professional affiliations, their contact information (including email accounts) and their complete mailing address for correspondence.

- All references should be numbered consecutively and listed as footnotes at the end of every page. In the text, references should be cited by a superior character of the corresponding number.

- For further information on writing style, consult *The Chicago Manual of Style*, 17th edition.