

Weapons as Systems

Components, Infrastructure and Human Organisation

Zawwar Sami

Independent Researcher, Systems Engineer, and Cybersecurity Specialist

ORCID: 0009-0004-5819-3017

W Series · Paper W1 · Version 1.0 · 14 October 2026

Abstract

A military capability cannot be inferred from the performance of its visible equipment alone. I develop this argument through a specific problem: identical individual availability figures can conceal different amounts of usable system service. A finite construction holds three required functions at 75 percent availability each while their joint availability varies from 25 to 75 percent. The result uses elementary intersection bounds; its significance here is that component-level summaries leave the service claim underdetermined when their temporal relationship is unreported. I distinguish component performance, integrated performance, availability, sustained service, and demonstrated outcome, then develop a claim-level boundary and evidence profile to connect each statement with appropriate support. A filled F-35 assessment separates measured readiness, programme objectives, and service requirements. Two Patriot investigations distinguish an identified failure from uncertainty about wider outcomes, while the Berlin Airlift illustrates successful sustained provision. Together, the analytical construction and documentary cases locate the central assessment problem in the relationships that convert resources into service over time. The contribution is an argument about what capability evidence must preserve and a practical format for doing so, not a new probability theorem or an empirically validated forecasting model. Physical effectiveness and ethical permission remain separate judgments.

Keywords: military capability; systems engineering; sustainment; infrastructure; human organisation; evidence; technological change

1. The object is only the beginning

A weapon is easy to photograph. The conditions that make it usable are harder to put in the same frame.

An aircraft has measurable physical properties. It also needs a serviceable configuration, trained people, maintenance arrangements, supplies, information, and decisions that can be carried out together. Some of those conditions sit inside the aircraft. Others belong to organisations that do not manufacture it, operate it, or even share the same immediate priorities. Looking closely at the machine is necessary. Stopping there leaves the larger claim unfinished.

I begin this series with that unfinished claim. When we say that a country, a military organisation, or a platform possesses a capability, what exactly have we established? We might mean that a component passed a test. We might mean that an integrated system completed an exercise. We might mean that sufficient equipment can be made available when required, or that an organisation can continue providing a service after the first successful demonstration. We might mean that an intended outcome actually occurred. These statements are related, but they are not interchangeable.

My concern is the movement between them. A limited result can become a much larger public claim without anyone identifying the additional assumptions. A successful demonstration becomes evidence of dependable service. Ownership becomes readiness. Technical possibility becomes an expectation of strategic effect. Each transition can be justified, but each needs an argument of its own.

The proposition that weapons belong to larger systems is well established. Systems engineering distinguishes the verification of requirements from validation in intended use. Research on military change examines organisational adoption as well as technical invention. Systems safety asks how interactions and control arrangements produce outcomes that cannot be understood through isolated component failures. Military capability frameworks already include personnel, training, logistics, and infrastructure.[1–7] The question is how to preserve these distinctions when a technical result becomes a wider capability claim.

My starting hypothesis is more specific than the claim that context matters: individual readiness figures can remain unchanged while the service available from their combination changes. What is missing is the relationship between the conditions of service, including when they coincide. A precise assessment therefore needs to preserve more than an inventory of parts and their separate scores. The unit of analysis is a specific capability claim. Its task, configuration, conditions, duration, dependencies, and evidence determine how far the analysis must reach. This makes it possible to ask both a material question—what arrangement can produce the service?—and an evidential question—what observations justify believing that it can?

These questions must remain distinct. A system may work when an outside observer cannot verify it. A well-documented system may still perform poorly. The purpose of better evidence is to improve judgment about capability. A change in the underlying relationship between resources can change service; a change in its description changes what we can justify believing about it.

This is the foundation of the W series. Later papers can examine physical limits, materials, energy, mobility, and different military platforms. Here I establish the terms on which those developments should be evaluated. The machine matters. So do the conditions under which its properties become something a human organisation can actually use.

2. Scope, sources, and the kind of argument being made

This paper combines conceptual analysis, an exact finite construction, and selected documentary cases. The construction isolates a relationship by holding individual availability constant; its numbers are synthetic and do not estimate any military platform. The documentary cases examine how capability claims are supported in public records. It combines research on systems and military technological change with engineering guidance and public oversight reports. The sources were selected for their relevance to the distinction between an artifact and an organisation's sustained ability to use it. They do not constitute a systematic review of all military technology research.

The cases serve different purposes. The F-35 sustainment audit concerns fleet readiness over time. The Dhahran investigation concerns a particular failure in an operating system. A separate assessment of Patriot in the Gulf War concerns the limits of evidence about outcomes. The Berlin Airlift provides a successful case of sustained military support. These

cases distinguish service, outcomes, and the evidence for them; they are deliberately contrasting illustrations, not a representative sample or a test of predictive accuracy.

I use “capability” to mean an ability to perform a specified task under stated conditions. “Sustained capability” adds a time horizon and the arrangements needed to continue performing it. Neither expression implies that the task is desirable, lawful, or strategically wise. A physically effective system can be used unjustifiably. A justified protective intention can also fail in practice. Keeping these judgments separate is necessary to evaluate both honestly.

I previously examined the distinction between potential action and verified effect in artificial agency; here I extend the question to material systems and sustained organisational service.[12]

There are further limits. Public records reveal some arrangements and conceal others. An audit can establish that an observed problem exists without identifying every cause or every possible remedy. Historical examples cannot settle how all future technologies will behave. The method proposed here is intended to improve the specificity of an assessment. Its usefulness still needs to be tested against alternatives.

3. What existing work already gives us

3.1 Systems engineering and systems safety

NASA's systems engineering guidance distinguishes checking a product against requirements from establishing that it serves intended use in its intended environment.[1] This distinction matters whenever a component test is asked to support a broader statement about an operational system. A requirement can be satisfied while leaving unanswered whether the requirement itself captures the circumstances of use.

Leveson's systems approach to accidents supplies a complementary warning: harmful outcomes can arise from interactions and inadequate control even when individual components behave as specified.[2] Component reliability therefore cannot exhaust a system-level assessment. For the present argument, the implication is limited but important. A list of successful parts does not by itself establish that their interaction is adequate for the claim being made.

Maier's systems-of-systems work further directs attention to the architecture of collaboration and to the consequences of misclassifying the system under study.[3] I use that general insight rather than assuming that every weapon arrangement belongs to one fixed technical category. A platform, a fleet, and a collection of supporting organisations can require different analytical boundaries.

None of this makes technical detail secondary. Material properties constrain what the larger arrangement can accomplish. Organisation cannot remove an energy requirement by coordinating more carefully. It can, however, affect whether sufficient energy is supplied, whether the equipment remains serviceable, and whether the available performance is appropriate to the task. Physical constraints and organisational arrangements act together.

3.2 Military adoption and the distribution of capability

Biddle's analysis of the Gulf War challenges explanations that attribute outcomes to either technological superiority or skill alone.[4] The relevant lesson here is their interaction. It is inadequate to treat human organisation as a residual factor added after the “real” technical explanation has been completed.

Horowitz examines how financial and organisational demands shape military adoption.[5] Ding and Dafoe place greater emphasis on the broad diffusion, industrial development, and complementarities associated with general-purpose technologies.[6] These approaches do not provide identical accounts of technological advantage. Read together, they expose a question that equipment inventories leave open: which institutions and supporting activities must develop before a technical possibility becomes usable at the required scale?

An invention can therefore be widely known while its military application remains uneven. The bottleneck may lie in acquiring equipment, reorganising work, developing relevant skills, or maintaining supporting industrial capacity. Which explanation matters is an empirical question. Calling all of them “the system” without distinguishing their roles would conceal rather than resolve it.

3.3 Existing frameworks and the contribution of this paper

The Ministry of Defence's through-life guidance names eight defence lines of development: training, equipment, people, infrastructure, doctrine, organisation, information, and logistics.[7] This already places equipment inside a wider capability arrangement. Assurance cases likewise connect claims, arguments, evidence, and objections, with revision as knowledge and circumstances change.[8]

The substantive argument is that a capability assessment must preserve the relationships on which service depends; separate measures can be accurate yet insufficient for the joint claim. Section 5 establishes this point with a controlled mathematical construction. The **claim-level boundary and evidence profile** then makes the relevant relationships, scope, and evidence explicit in a practical assessment. The profile combines five distinct types of claim with a record of scope, dependencies, and conditions for reconsideration. Its particular use is to examine a transition: does evidence supporting one statement also support the broader statement now being made? Table 1 locates this use alongside existing work.

Table 1. Existing foundations and the use proposed in W1.

Foundation	Established function	Use in this paper
Verification and validation [1]	Relate a product to requirements and intended use.	Identify whether a capability statement exceeds the configuration or conditions examined.
Systems safety [2]	Examine interactions and control beyond isolated component reliability.	Include dependencies that materially affect the specified service.
Defence lines of development [7]	Organise the contributing dimensions of capability.	Link relevant dimensions to one task, time horizon, and evidential question.
Reliability models [16]	Specify the logic and dependence assumptions relating parts to system performance.	Expose the information lost when only separate availability summaries are retained.
Assurance cases [8]	Connect claims, rationale, evidence, and objections; support revision.	Present the evidence for a capability statement in a compact, explicitly scoped profile.

The construction and profile address the same problem at different levels. The construction shows why separate availability summaries cannot identify service in the stated model. The profile asks whether an actual assessment contains the relationship evidence needed for its own claim, including shifts from a component to a fleet or from a demonstration to continuing service. A well-developed assurance case can already express these distinctions. W1 offers a focused assessment format, not an expressive power unavailable to existing methods. Its comparative value depends on whether it helps readers detect consequential shifts with less ambiguity; Section 10 describes how that should be tested.

4. Five claims that should not be collapsed

Consider the statement that a system “works.” It can refer to at least five different claims.

Component performance. A part exhibits a specified property under test conditions. The evidence belongs first to that part and those conditions, rather than to every configuration in which it might later be used.

Integrated performance. A defined combination performs a task. Interfaces, software versions, supporting equipment, and operator arrangements now matter. Evidence attaches to that configuration; the consequences of changes require assessment.

Availability. A usable system can be provided when needed. This introduces repair, scheduling, personnel, supplies, and competing demands. Technical performance does not establish how often that performance is available.

Sustained service. An organisation can continue providing the specified capability over a stated period. Replenishment, maintenance capacity, replacement skills, industrial support, and institutional continuity enter the claim. One successful event leaves much of this question open.

Demonstrated outcome. Adequate evidence establishes that the intended result occurred. A recorded command or completed procedure may establish an action without establishing its effect. Conversely, incomplete observation does not show that no effect occurred.

These are different questions, not maturity levels. A hypothetical transport aircraft makes the distinction concrete. It can complete a documented delivery on Monday, be unavailable for scheduled maintenance on Tuesday, and return to service on Wednesday. Monday's outcome remains established during Tuesday's unavailability. Neither observation alone answers whether a fleet can provide a daily service for six months. Conversely, an unused reserve aircraft may be available without having demonstrated the outcome of the mission for which it is held. There is no single upward sequence through these states.

The distinctions also prevent a misleading model of dependency. Some functions have indispensable prerequisites; others have substitute routes, reserve capacity, or partial modes of service. A common dependency can disable arrangements that looked independent on an equipment list. Which relationship holds is a matter for the actual configuration. Multiplying loose ratings for technology, training, and logistics would create an appearance of precision without a defensible measurement model. The profile therefore records dependencies and uncertainty rather than producing a universal capability score.

5. Identical readiness figures, different service

5.1 A finite construction

Suppose a generic service requires three functions to be available together: equipment, qualified personnel, and supporting infrastructure. Divide the observation window into four equal periods. Each function is either available throughout a period or unavailable throughout it. For this deliberately closed model, the service is feasible if and only if all three are available. Assume compatible functions, sufficient capacity, constant demand, and no buffer, substitute, or carry-over between periods. These assumptions define the example; they are not claims about every military system.

Now hold each function's individual availability at three periods out of four. Table 2 gives two arrangements with identical individual totals. A 1 means available and a 0 unavailable; each four-digit sequence reads from period 1 to period 4.

Table 2. Equal individual availability does not determine joint service.

Function or result	Arrangement A	Arrangement B
Equipment	1110	1110
Qualified personnel	1110	1101
Supporting infrastructure	1110	1011
Individual availability, each function	75%	75%
All three available together	1110	1000
Feasible service periods	3 of 4: 75%	1 of 4: 25%

Nothing in the individual percentages distinguishes the arrangements. Nevertheless, their feasible service differs by a factor of three. Any score calculated solely from those three percentages must assign the same value to both arrangements and therefore cannot recover their different service levels. Changing the weights in that score cannot supply the missing relationship.

This is the precise assessment problem: even perfectly measured individual availability does not identify joint availability when the pattern of overlap is unknown. In an open real-world system, jointly available prerequisites may still be insufficient because additional conditions matter. The closed construction already establishes the insufficiency of the separate percentages under unusually favourable assumptions.

5.2 The bound and its assumptions

Let a , b , and c denote the fractions of the observation window for which the three functions are individually available. Let s denote the fraction for which all three are available together. Without assuming independence, an elementary intersection bound gives:

Bound: $\max(0, a + b + c - 2) \leq s \leq \min(a, b, c)$.

The upper bound follows because the overlap cannot exceed the availability of any required function. For the lower bound, the fraction of periods with at least one unavailable function is no greater than the sum of their unavailable fractions, $(1 - a) + (1 - b) + (1 - c)$. Subtracting this sum from one, and respecting a lower limit of zero, gives the expression above. This is a counting argument, not a new probability theorem.

With $a = b = c = 0.75$, the permissible interval is 0.25 to 0.75. Both ends are attained in Table 2. There are exactly $4 \times 4 \times 4 = 64$ arrangements of the three unavailable periods. Exhaustive enumeration gives 24 arrangements with one jointly available period, 36 with two, and four with three. These are counts of synthetic arrangements, not probabilities assigned to real operations. They provide a reproducible check of the construction and its attainable endpoints.

Multiplying the three fractions instead gives 0.421875. That is a model-based probability, not an observed fraction in either four-period record. It requires mutual independence of availability at a randomly selected period; the three fractions alone do not establish it. NIST's series-reliability guidance likewise makes independence an explicit assumption, though its non-repairable, first-failure model is distinct from this finite availability construction.[16] Neither model licenses multiplying arbitrary readiness measures. In particular, the F-35 MC and FMC metrics discussed below are nested categories, not independent functions to insert into this example.

5.3 What the construction changes

The decisive step is the move from separate readiness to jointly usable service. More precise component percentages cannot remove the ambiguity if the relationship remains unobserved. The analyst needs joint records, a justified dependence model, or direct evidence of the service itself. This is why adding another descriptive category to a checklist is not necessarily enough.

Joint availability still leaves a further question about continuity. Over eight equal periods, 11110000 and 10101010 each provide 50 percent availability. The first contains four consecutive available periods; the second never contains more than one. A task requiring two consecutive periods can fit inside the first record but not the second. The time structure required by the task must therefore survive the aggregation used to describe capability.

The construction does not recommend synchronising real maintenance or staffing interruptions. Safety, fluctuating demand, reserve arrangements, and other tasks may require a different pattern. It establishes a diagnostic result: apparent equality at the level of parts need not be equality at the level of service. Whether changing that relationship is feasible or beneficial is a separate empirical question.

Human organisation is implicated in this relationship without being reduced to a residual factor. It can affect how functions coincide, how interruptions are absorbed, and whether the periods of available service match the periods of need. The claim to investigate is consequently stronger than “people matter”: under some conditions, an organisation can alter usable service without changing the separate availability totals. The finite model proves that possibility; it does not estimate its size in a real institution.

6. Choosing a boundary that is large enough

“Everything is connected” is a poor stopping point for research. If a system includes every social and material condition that has ever helped produce it, analysis becomes impossible to complete. The boundary has to follow the claim.

For a material tested in a laboratory, the sample, environment, and measurement procedure may be sufficient to answer the immediate question. A claim about maintaining a fleet incorporating that material for several years also raises manufacturing consistency, repair, replacement supply, and skills. The boundary expands because the proposition has changed.

Duration and scale matter for the same reason. A brief operation can draw on stored resources; continued service may depend on replacing them. One available platform is different from simultaneous service by many platforms. A report that omits these

dimensions leaves the reader to supply them, often more generously than the evidence warrants.

I use three questions to decide whether a dependency belongs inside an assessment. Could a plausible change in it materially alter the claimed service within the stated period? Is there evidence that it is adequately provided, or that a substitute exists? Would excluding it hide an assumption that a reader could reasonably mistake for an established fact?

These questions provide a stopping rule, though not a mechanical algorithm. Dependencies with no plausible bearing on the specified service can remain outside the assessment. Unexamined but consequential conditions should be recorded as assumptions. The analyst can then explain why a boundary is adequate instead of claiming to have modelled an entire institution.

Human organisation enters through specific responsibilities. Who trains and retains the necessary people? Who resolves a problem that crosses a contractor and a public agency? Who approves a change, and who receives evidence of its effect? Assigning responsibility in a profile is not proof that the responsibility is being fulfilled. It identifies where the claim needs organisational evidence.

Ownership is similarly distinct from control. An organisation may own a platform while relying on external parts, specialist labour, information, and support arrangements. Dependence is not automatically a defect. Its importance depends on the task, alternatives, and time horizon. Making that dependence visible prevents equipment ownership from being silently treated as assured service.

7. A compact profile for capability claims

The following profile can be used when an assessment begins and when its conclusions are communicated. Supporting records may be extensive, but the proposition and its evidential boundary should remain readable.

Table 3. The six fields of a capability-claim profile.

Field	What the assessment should state
Task and claimed result	What is to be accomplished, and which type of claim is being made?
Boundary and configuration	Which equipment, software, people, interfaces, and organisations does the statement cover?
Conditions, scale, and duration	Under what circumstances, at what scale, and over what period should it hold? What overlap or continuity of functions does the task require?
Dependencies and responsibility	Which inputs and support functions are necessary, what substitutes exist, and who provides them?
Evidence and uncertainty	What was observed, including relevant relationships between functions? Which results are joint observations, separate measures, assumptions, or unknowns?
Change and reconsideration	Which changes require review of the claim or its evidence, and who should perform it?

To use the profile, first preserve the original wording of the claim and classify what it asserts. Then populate the fields from traceable sources, separating observations from plans and assumptions. Finally, write the strongest conclusion the evidence supports. If that conclusion is narrower than the original statement, identify the missing support explicitly. The output is a qualified judgment, not simply a completed form.

Evidence must be recorded at its actual level. A design document establishes intended arrangements; inspection can establish that resources were present; a demonstration establishes performance under observed conditions. Operating records can support a broader service claim, subject to measurement quality and coverage. Each can be useful without being sufficient for every question.

The profile should retain both supporting and disconfirming observations. Success in one period and difficulty in another may reveal a condition of performance rather than justify a single verdict about the platform. Likewise, an unknown field should remain an unknown. Empty evidence cannot be repaired by confident wording.

Reconsideration is about the relevance of evidence after change. A test remains a historical fact, but a changed configuration, staffing arrangement, supply system, or environment may weaken the inference from that fact to present service. Review should be proportionate: identify the relationship affected, retain evidence that still applies, and seek new evidence where the relationship has changed. It is unnecessary to repeat every test after every adjustment.

For a retrospective public analysis, the analyst can propose review triggers but cannot assign duties to an institution. An internal assessment should name the responsible role and review interval. Keeping that difference visible avoids presenting an analyst's recommendation as an implemented control.

8. Worked assessment and contrasting cases

8.1 F-35: from a readiness objective to a qualified assessment

The June 2026 GAO sustainment report provides both a measured baseline and a future objective. Across fiscal years 2021–2025, the reported mission-capable rate fell from 67 to 44 percent and the full-mission-capable rate from 38 to 25 percent. GAO describes the Joint Program Office's Global Support Solution Reset as aiming for 80 percent mission capable and 65 percent fully mission capable by 2030.[9, Highlights; Figures 4–6; “JPO Developed an Updated Sustainment Strategy and Funding Needs”]

The decreases are 23 and 13 percentage points. These rates measure the percentage of time aircraft in the possession of F-35 units can perform at least one tasked mission, or all tasked missions, respectively. They are not percentages of successful combat engagements or necessarily the fraction of the total inventory ready on a particular day. Full mission capability is nested within mission capability, so multiplying the two rates would be inappropriate.

The claim to assess: the programme aims to achieve the stated readiness levels by 2030. This accurately describes an objective. An unstructured summary might strengthen it to “the new strategy will restore the fleet's required capability.” That second sentence is an illustrative overstatement, not a quotation attributed to GAO or the programme. It changes a goal into an assured outcome and assumes that the goal satisfies every user's requirement.

Table 4 applies the profile to that transition. The first five fields draw on GAO's public findings; the final field states my proposed review conditions, not an existing programme commitment.

Table 4. Filled F-35 profile: assessment as of the June 2026 report.

Field	Filled assessment
Task and claimed result	Increase fleet readiness to the stated 2030 objectives. This is a prospective availability and sustainment claim, not a demonstrated combat outcome.
Boundary and configuration	For the baseline used here: U.S. F-35 variants in unit possession, with their maintenance and supply arrangements. Do not silently extend the U.S. aggregate to every partner fleet or individual configuration.
Conditions, scale, and duration	Compare fiscal-year 2021–2025 measurements with a 2030 objective. Preserve the definitions of MC and FMC, the changing fleet composition, and the distinction between an annual rate and availability at one time.
Dependencies and responsibility	JPO, the military services, and contractors share implementation. GAO identifies parts supply, repair capacity, maintenance practices, and staffing among relevant dependencies. Funding and contracted activity require evidence of resulting service.
Evidence and uncertainty	GAO supports the historical rates and the existence of the plan. It identifies risks and missing formal mitigation plans. The 2030 result is unobserved, and the programme's targets do not fully match all service needs.
Change and reconsideration	Reassess against later readiness records if parts capacity, maintenance arrangements, software/configuration mix, fleet size, or metric definitions change. In programme use, a designated owner should record the comparison and unresolved assumptions.

The revised conclusion: the programme has a documented plan to improve readiness, but the June 2026 evidence does not demonstrate achievement of the 2030 objectives. Even achievement would need to be compared with service-specific requirements before claiming that the fleet's required capability had been restored.

The last distinction is substantive. GAO records an Air Force objective of 80 percent FMC, compared with the programme's 65 percent FMC goal; Navy officials regarded the proposed goals as meeting their sustainment requirements.[9, “JPO Lacks Formal Risk Mitigation Plans for GSS Reset”] Thus “meeting the programme target” and “meeting the user's requirement” are different claims even if both use the word readiness. A list containing equipment, people, and logistics would not, by its completion alone, settle that mismatch.

The profile also separates a rate from the number of aircraft available. A growing fleet can have a lower availability rate without a proportionate decline in its absolute available capacity. Determining that quantity requires consistent population and time data. I therefore make no numerical estimate of available aircraft from the headline percentages alone.

What has the exercise added? It has made four otherwise compressible distinctions visible: objective versus observation; unit-possession denominator versus total inventory; annual rate versus point-in-time provision; and programme goal versus service requirement. GAO's investigation already contains the information. The profile's role is to preserve it in the capability judgment, not to claim a discovery the audit missed. No experiment here

establishes that the format outperforms a careful unstructured reader.

8.2 Dhahran: an operating system changes with time

GAO's investigation of the Patriot failure at Dhahran on 25 February 1991 identified a software problem whose effect increased with continuous operating time. The failure occurred after more than 100 hours of operation. Twenty-eight American soldiers were killed in the attack; modified software arrived the following day.[10]

For this paper, the relevant lesson is the relationship between a configuration, its operating conditions, and the process through which a correction reaches it. Time in operation was part of the conditions under which the system behaved. The availability of a technical correction somewhere in the organisation was not equivalent to its presence in the affected deployed configuration.

A component-centred account could identify the software defect and stop there. A capability assessment also asks which operating assumptions the evidence covered, how information about changed conditions moved, and whether the deployed arrangement had actually changed. Those questions do not replace the technical explanation. They explain why the technical explanation belongs to an operating system rather than to an isolated piece of code.

The case does not prove that longer operation always reduces performance or that every software update requires the same response. It shows why duration and configuration cannot be omitted when they are causally relevant. A historical test result remains informative only within an argument about how its conditions relate to the conditions of use.

8.3 Patriot outcomes: uncertainty is not a measured failure rate

A separate GAO assessment of Patriot's Gulf War performance concluded that the available data could not conclusively establish how well it had performed. The report examined limitations in the evidence used to assess outcomes.[11] That conclusion concerns the warrant for a performance claim. It should not be converted into a precise rate of success or failure.

This distinction is central to the paper. An event can occur without being adequately observed. A recorded system action can occur without establishing the intended effect. Where the evidence is limited, confidence should be limited; the analyst should not fill the gap with either technological enthusiasm or technological dismissal.

The assessment also makes observation part of the system boundary when a claim depends on demonstrated outcomes. Records, measurement arrangements, and their continuity affect what can later be established. They do not necessarily determine the physical outcome itself. A better evidence system can improve accountability and learning while leaving underlying performance unchanged. Conversely, a physically capable system can remain difficult to assess publicly.

This case must be kept separate from Dhahran. One investigation addresses a specific failure with an identified mechanism. The other addresses the adequacy of evidence for a broader performance judgment. Collapsing them would reproduce the very error the paper seeks to avoid: replacing a precise claim with a larger one that the evidence does not support.

8.4 Berlin: a successful service with identifiable support

A framework built only around failure can become a device for finding fault after the event. The Berlin Airlift supplies a contrasting observation. Air Mobility Command records more than 2.3 million tons of cargo delivered between 26 June 1948 and 30 September 1949.[14] This is evidence of repeated material delivery over an extended period, rather than an inference drawn from an aircraft's rated payload.

The National Museum of the U.S. Air Force describes maintenance support during the peak months for 324 aircraft, 1,600 motor vehicles, and 500 items of ground power equipment, with supporting work continuing in the United States, England, and Germany.[15] These details make the enabling system visible: sustaining the delivery service required maintaining more than the aircraft that carried the cargo.

Table 5. Applying the same profile to successful sustained delivery.

Field	Berlin Airlift assessment
Task and claimed result	Deliver supplies to Berlin repeatedly. The documented claim is sustained transport service with recorded cargo delivery.
Boundary and configuration	Allied air transport and its supporting ground and maintenance arrangements; not aircraft payload specifications alone.
Conditions, scale, and duration	The historical operation from June 1948 to September 1949. The total belongs to this period and setting.
Dependencies and responsibility	Air and ground personnel, serviceable aircraft, vehicles, power equipment, and depot support. The maintenance record identifies concrete contributors rather than an unspecified organisational factor.
Evidence and uncertainty	Official historical accounts document aggregate delivery and supporting activity. These sources do not establish that every daily requirement was met or quantify each dependency's causal contribution.
Change and reconsideration	Reuse of the result for another operation would require fresh evidence about transport access, support arrangements, demand, and conditions. This is an analytical transfer condition, not a historical forecast.

The warranted conclusion is positive: the organisation provided an extended delivery service under the recorded historical conditions. The profile does not turn success into uncertainty simply because more information could always be collected. It distinguishes what the successful record establishes from what would require a new argument.

This case concerns military support rather than a weapon's destructive effect. That is an intentional boundary: it tests whether the same questions can describe an enabling service succeeding, while making no claim to validate combat effectiveness. It also does not prove that this profile would have predicted the success in advance. Its role is to prevent a method for assessing capability from becoming a vocabulary reserved for failure.

9. Limits of the method

The finite construction proves a limited non-identification result under explicit assumptions. The documentary worked example illustrates the profile's use, not its comparative validation. Neither result establishes that the profile will improve an organisation's decisions in practice. A capable engineer, auditor, or assurance practitioner can identify the same distinctions without this table. The proposal should earn its place by making important assumptions easier to inspect when a conclusion moves between technical teams, institutions, and public debate. Completing six fields is not evidence that the underlying system is ready.

Public evidence imposes another limit. Some relevant records are restricted, missing, or difficult to interpret. An outside analyst can state what the available record establishes without pretending to know everything that exists. “Not established by this evidence” is often the correct conclusion; “cannot work” requires a different argument.

Boundary choices also involve judgment. An analyst may omit a dependency that later proves consequential or include so many remote conditions that the assessment becomes unusable. The stopping questions in Section 6 make that choice explicit, but they cannot remove the need for subject expertise. A profile should identify disagreement about a boundary rather than manufacture consensus.

Retrospective explanation presents a further risk. After an outcome is known, it is easy to list dependencies that make the result seem inevitable. The cases here were selected to illustrate different evidential situations, including success; that balance does not remove selection bias or establish a general causal law. Where possible, claims, conditions, and proposed explanations should be recorded before the outcome, with alternatives preserved.

Finally, attention to organisation must not dissolve physical limits into social explanation. Institutions can improve the provision and use of material resources. They cannot make an inadequate energy supply sufficient by describing it differently. The value of a systems account is to show where a constraint bears on the service and what evidence supports the proposed response.

10. How this proposal could be evaluated

The next research step has two parts: investigate the relationship identified by the construction, and test whether the profile improves judgments that matter. For the first part, obtain timestamped records of required functions and delivered service over a common observation window. Compare what separate availability summaries can establish with what joint and continuity records add. The key question is whether missing overlap explains an evidential or service gap after accounting for demand, capacity, substitutions, and other relevant conditions. If it does not, that would limit the proposed mechanism's explanatory importance in that setting.

For the second part, a suitable study could present assessors with public capability claims and the same source material, then compare an unstructured assessment with one using the profile. The comparison should examine whether assessors distinguish the five types of claim, identify material unsupported assumptions, and express uncertainty consistently.

Cases should be selected before assessing the method and should include apparent successes, failures, and genuinely indeterminate outcomes. Otherwise a collection of visible failures would make any dependency framework look more informative than it is. Assessors should not be rewarded merely for producing longer lists of concerns. The question is whether the concerns are relevant and supported.

A second stage could examine whether different readers identify similar boundaries and revision triggers. Disagreement would be informative. It might indicate ambiguous definitions, insufficient source material, or a substantive dispute about the task itself. A useful framework should help expose these differences rather than conceal them behind a common score.

A stronger prospective evaluation would record a claim before a documented change and examine whether the identified triggers help determine which conclusions require review afterward. Such work would need access to reliable configuration and service records. Public retrospective cases alone cannot substitute for it.

No comparative or prospective field study is reported here. The computed result is the exact synthetic construction, and the historical analysis is documentary. The further hypothesis is that preserving the relevant relationships will improve assessment and sometimes reveal opportunities to improve service. It remains open to empirical testing.

11. Human responsibility and the purpose of capability

A more accurate account of military capability does not supply a reason to use it. I want to retain that distinction throughout the W series.

Human responsibility appears in this paper in two forms. It is an empirical part of the system: people establish goals, organise work, maintain equipment, interpret information, and decide what happens next. It is also a moral relationship: people remain responsible for the purposes and consequences of the arrangements they create. An institution cannot answer the second question merely by showing that the first is technically effective.

The ICRC's position on autonomous weapon systems expresses concerns about human judgment, control, and the consequences of delegating force.[13] I cite it as an institutional ethical and policy position, not as proof that its recommendations already constitute binding rules everywhere. The broader point for W1 is that an assessment of capability needs to remain open to an assessment of purpose.

My interest in defence begins with the protection of human life. Understanding weapons seriously requires understanding the damage they can produce, the limits of claims made about them, and the institutions that accept responsibility for their use. It also leaves open a question that runs beyond this first paper: how much protection can be achieved without making killing the measure of technological progress?

That question does not justify overstating the performance of protective systems. A humane intention increases the importance of honest assessment. If people depend on a system for protection, the gap between a persuasive description and dependable service has consequences for them. The same discipline that limits inflated claims about force should apply to claims about defence.

12. Conclusion

Weapons have physical properties. Military capability belongs to a larger relationship between those properties, an operating configuration, supporting institutions, conditions of use, and time. An assessment becomes unreliable when evidence about one part of that relationship is allowed to stand for the whole.

The finite construction establishes why accurate individual availability figures can leave service underdetermined: the same component totals support different joint patterns and different continuity. I have distinguished component performance, integrated performance, availability, sustained service, and demonstrated outcomes so that this gap can be located rather than hidden inside a general claim that a system works. The proposed profile makes the meaning of a claim explicit, identifies the dependencies it assumes, and records when its evidence should be reconsidered. The F-35 exercise shows how these distinctions change the wording of a real readiness assessment. The contrasting historical cases show that the profile can describe success as well as evidential limits; its comparative usefulness remains open to evaluation.

The central discipline is to preserve the relationships that make the claimed service possible. A capability assessment should state what the evidence establishes about those relationships, not merely how impressive its separate parts appear. Where the claim extends further, show the argument, the assumptions, and the uncertainty. This does not diminish technology. It locates technology in the world where people must build, maintain, govern, and live with it.

References

- [1] NASA. (2016). NASA Systems Engineering Handbook, NASA/SP-2016-6105 Rev 2. Sections 5.3-5.4; accompanying product-realization guidance. <https://www.nasa.gov/reference/5-0-product-realization/>
- [2] Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, 42(4), 237-270. [https://doi.org/10.1016/S0925-7535\(03\)00047-X](https://doi.org/10.1016/S0925-7535(03)00047-X)
- [3] Maier, M. W. (1998). Architecting principles for systems-of-systems. *Systems Engineering*, 1(4), 267-284. [https://doi.org/10.1002/\(SICI\)1520-6858\(1998\)1:4%3C267::AID-SYS3%3E3.0.CO;2-D](https://doi.org/10.1002/(SICI)1520-6858(1998)1:4%3C267::AID-SYS3%3E3.0.CO;2-D)
- [4] Biddle, S. (1996). Victory misunderstood: What the Gulf War tells us about the future of conflict. *International Security*, 21(2), 139-179. <https://doi.org/10.1162/isec.21.2.139>
- [5] Horowitz, M. C. (2018). Artificial intelligence, international competition, and the balance of power. *Texas National Security Review*, 1(3), 36-57. <https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/>
- [6] Ding, J., & Dafoe, A. (2023). Engines of power: Electricity, AI, and general-purpose, military transformations. *European Journal of International Security*, 8(3), 377-394. <https://doi.org/10.1017/eis.2023.1>
- [7] Ministry of Defence. (n.d.). Plan the through-life approach. Digital MOD, Secure by Design guidance. Accessed 11 October 2026. <https://www.digital.mod.uk/policy-rules-standards-and-guidance/secure-by-design/plan-through-life-approach/plan-the-through-life-approach>
- [8] NASA Office of Safety and Mission Assurance. (2020, September 22). New tool for developing safety assurance cases. <https://sma.nasa.gov/news/articles/newsitem/2020/09/22/new-tool-for-developing-safety-assurance-cases>
- [9] U.S. Government Accountability Office. (2026, June 11). F-35 sustainment: Actions needed to ensure updated strategy improves persistent readiness challenges. GAO-26-108113. <https://files.gao.gov/reports/GAO-26-108113/index.html>

- [10] U.S. General Accounting Office. (1992, February 4). Patriot missile defense: Software problem led to system failure at Dhahran, Saudi Arabia. IMTEC-92-26. <https://www.gao.gov/assets/imtec-92-26.pdf>
- [11] U.S. General Accounting Office. (1992, September 22). Operation Desert Storm: Data does not exist to conclusively say how well Patriot performed. NSIAD-92-340. <https://www.gao.gov/assets/nsiad-92-340.pdf>
- [12] Sami, Z. (2026). From knowing to doing: Tool access, execution, and evidence in artificial agency. A8, version 1.8. Independent research preprint. <https://zenodo.org/records/23287948>
- [13] International Committee of the Red Cross. (2021, May 12). ICRC position on autonomous weapon systems. <https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems>
- [14] Air Mobility Command. (2014, September 30). The Berlin Airlift celebrates 65 years. U.S. Air Force. <https://www.amc.af.mil/News/Article-Display/Article/786451/the-berlin-airlift-celebrates-65-years>
- [15] National Museum of the United States Air Force. (n.d.). Maintaining the Berlin Airlift. Accessed 11 October 2026. <https://www.nationalmuseum.af.mil/Visit/Museum-Exhibits/Fact-Sheets/Display/Article/197652/AFmuseum/maintaining-the-berlin-airlift/>
- [16] National Institute of Standards and Technology / SEMATECH. (n.d.). e-Handbook of Statistical Methods, Section 8.1.8.2, "Series model." Accessed 11 October 2026. <https://www.itl.nist.gov/div898/handbook/apr/section1/apr182.htm>