

Coping with Counter-Evidence in Complex Argument Structures

By Richard Maguire B.Eng, MSc, MIMechE, MSaRS; SE Validation Limited
and Dr. Dirk Brade; LFK-Lenkflugkörpersysteme GmbH.

1. INTRODUCTION

Aerospace engineers are more frequently using argument structures to demonstrate the confidence in predictions about system behaviour. This type of tool is becoming more widespread as scientists attempt to overcome the problems in analysing increasingly complex and inter-related systems. When a predictive argument structure is constructed, it represents an ideal with explicit targets for evidence to satisfy the argument. The evidence targets can be key to convincing an authority about the airworthiness, integrity or functional capability of a system, and even a system-of-systems. As such, the evidence may take the specific form of targets of accreditation, or targets of verification and validation.

Unfortunately, the ideal rarely works out, evidence is not available due to resource (time, money and people) constraints, or the details actually provide counter-evidence. This is evidence that opposes the predicted outcome. Every system has counter-evidence, even if it is just minor performance failures during testing and evaluation. This counter evidence then ruins the argument structure, which has to be hastily re-drawn and worked out. This can be expensive and it does not enhance the probative force of the original argument.

Methodologies of coping with counter evidence are currently not fully mature. More effort in understanding this evidence and its implications is required to provide engineers with decision support information and a starting point when they come across the phrase “Well, nobody expected that to happen.”

2. THE REQUIREMENT FOR THE CONSIDERATION OF EVIDENCE

The increase in complexity in modern avionics makes it difficult to understand the predicted and modelled behaviour of the system in its normal operations and in its failure modes. Evidence must be gained to provide insight into the relationships between complex non-linear operational variables and the demonstration of performance from the system of interest. Several UK and European standards explicitly call for evidence to be collected, collated and presented as part of safety cases, airworthiness certification and modelling validity and verification.

Within the MoD standard 00:56 [MoD 2004], clause 11.3.1 states that the Contractor shall provide compelling evidence that safety requirements have been met, and that the quantity and quality of the evidence shall be commensurate with the potential risks posed by the system and the complexity of the system. Guidance in section 9.5.6 of part two of the standard indicates that counter evidence has the potential to undermine a previously accepted argument, and that the validity of the whole safety case may be called into question. The process of searching for and acting upon counter-evidence are important parts of a robust safety management system, and they should be explicitly documented in the safety case.

Items of evidence also constitute the foundations for the Verification and Validation of Simulation Models and Simulation Results (M&S V&V). According to the "Region Europe Verification, Validation, and Acceptance Activity" (REVVA) process [REVVA 2004] (which is a candidate for short term standardization), a simulation model can be considered as valid, if within the experimental frame of interest, the simulated system behaviour is (under the given accuracy requirements) indistinguishable

from the behaviour of the system of interest. A simulation model can be considered to be correct, if its representation in its various forms and the transformation between these forms complies with given requirements, formalisms, rules, and regulations. Whether a simulation is valid and correct or not is predominantly a consequence of the rigour and care taken during model development. Verification and validation are conducted to confirm or deny that a Simulation Model or Simulation Results are fit for purpose. As with every assurance measure, the outcome of this confirmation or denial process needs to be backed up with appropriate evidence. This evidence shall positively demonstrate that the model meets all formulated requirements on validity and correctness, or – as counter-evidence - show that certain requirements are not met.

3. FAILURE TO BACK-UP THE NULL-HYPOTHESIS

In many engineering assessments the proof that a complex system does what it should, and only what it should, is often a search for different evidence types. Usually, due to system or model complexity it cannot be unmistakably proven that the system is safe or the model is valid. It is usually impossible to provide a complete positive proof for any non-trivial system, it is difficult for people to analyse full system behaviour under normal and failure conditions and to confidently predict system performance. One rather creates the null hypothesis that the system is safe or the model is valid. Subsequently one subjects the system or model to systematic, rigorous analysis, testing, and verification, to figure out whether the null hypothesis can be maintained or whether it must be rejected as wrong. Results from analysis, testing, and verification, that indicate that the null hypothesis is wrong, are called counter-evidence. Counter-evidence arises in both areas as evidence of incorrect behaviour or performance. However, finding just one point of failure in a system does not necessarily prove it to be totally incorrect. A single failing or even several failings may still be perfectly tolerable, but only if the evidence and argument for this position can be presented, i.e., only if a compelling argument can be made that despite of the deficiency the safety or validity requirements are still met.

4. THE PROBLEMS AND BENEFITS THAT COUNTER EVIDENCE CAN BRING

Counter-evidence has the potential to have a much stronger effect than positive evidence, as even the identification of that single fault or a single failure during a system's operational life or pre-qualification testing, may be sufficient to invalidate the claim for safety, airworthiness or fitness for purpose. Items of counter-evidence can arise from:

- In-service incidents and events
- Insecurities in software development techniques
- Fault detection analysis
- Proof of concept trials
- Dynamic and static testing
- Flight trials and tests

The discovery of counter-evidence can have severe impacts on a project. Major aircraft programmes have become more complex to manage with resultant schedule and cost implications if significant counter-evidence starts to be discovered. In such cases it may be necessary to obtain additional evidence, carry out remedial actions or even take the system out of service altogether. Further more, if any remedial design action is taken, re-certification and re-qualification must be done to ensure that one fix has not introduced additional failures.

A rigorous search for counter-evidence followed by objective analysis and correction may be viewed as classic positive evidence of a robust management and evaluation process. This is equally true for safety arguments and more generic fitness for purpose arguments. This evidence must be recorded for it to have a positive effect on the overall probative force of the argument being made. Also the absence of

documented counter-evidence is rarely indicative of the absence of counter-evidence, it is often indicative of an absence of documentation resulting from an inadequate management and evaluation system.

Counter-evidence is always evidence of a boundary of one sort or another – be it exceeding the operational envelope or limiting the operational capability. When counter-evidence has been uncovered the boundary has become explicit so that something can be done about it. This is of great and often underestimated value, even if it is unpleasant in the short term for the project – better to find it early and before the in-service event.

5. THE GOAL STRUCTURING NOTATION A RESUMÉ

A specific graphical display system has been put forward to enable the presentation of an argument case organised in these various styles, it is known as ‘Goal Structuring Notation’ or GSN [Kelly 2003]. The principle elements of GSN are shown in Figures 1 and 2 below. These elements may then be joined together using links that show how the whole structure combines together to form a cohesive argument and justification for the demonstration of the primary goal. The links can be given directional arrows to indicate whether the structure is a top-down style or a bottom-up style.

The notation is infinitely flexible and can be used to represent any complex argument, not just safety based. It is particularly suited to the types of standards and regulations that are goal-based rather than prescriptive in nature, so any industry could make use of them. The flexibility means that it is highly likely that completely different goal structures could be drawn to satisfy the same complex problem – and both would be correct, just from a different point of view [Maguire 2006].

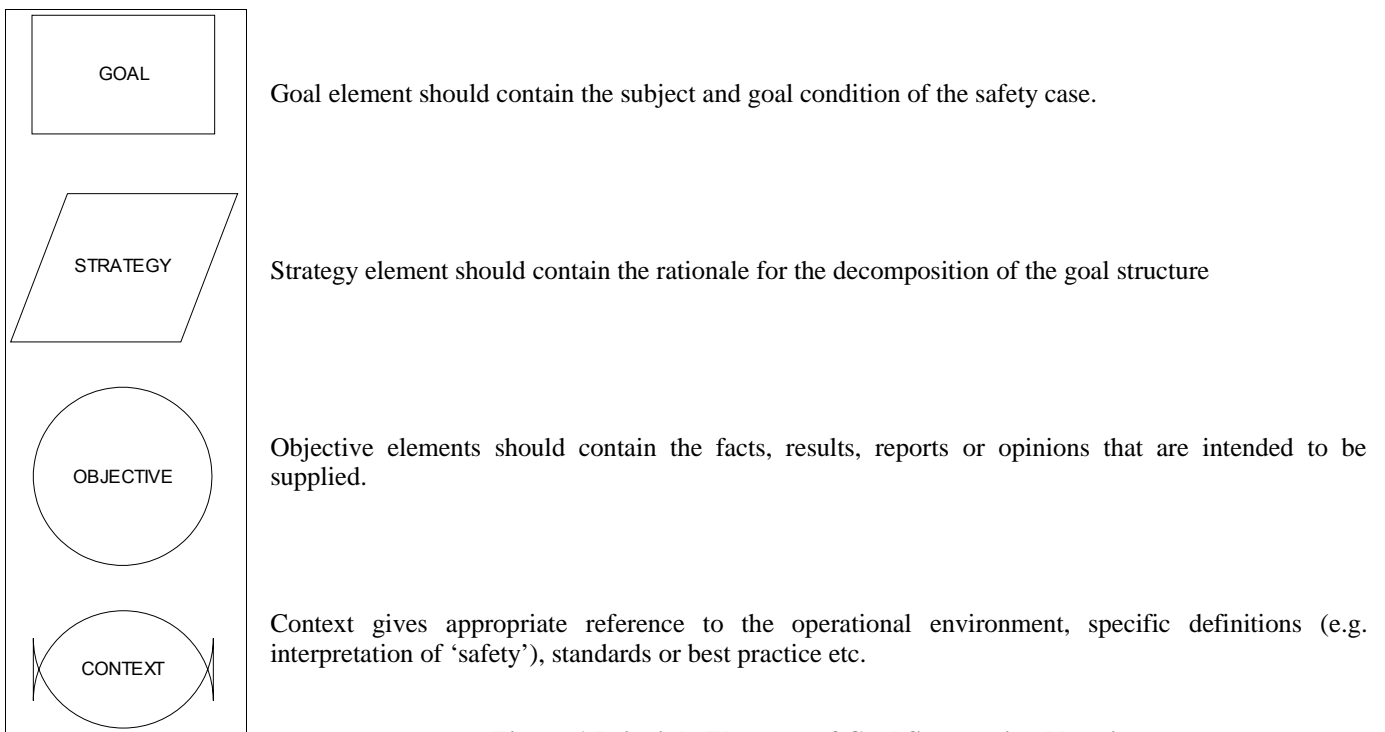
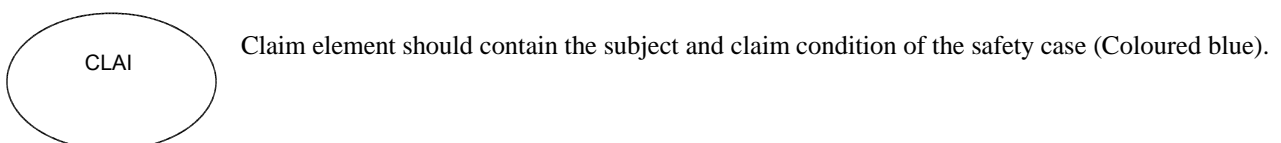


Figure 1 Principle Elements of Goal Structuring Notation



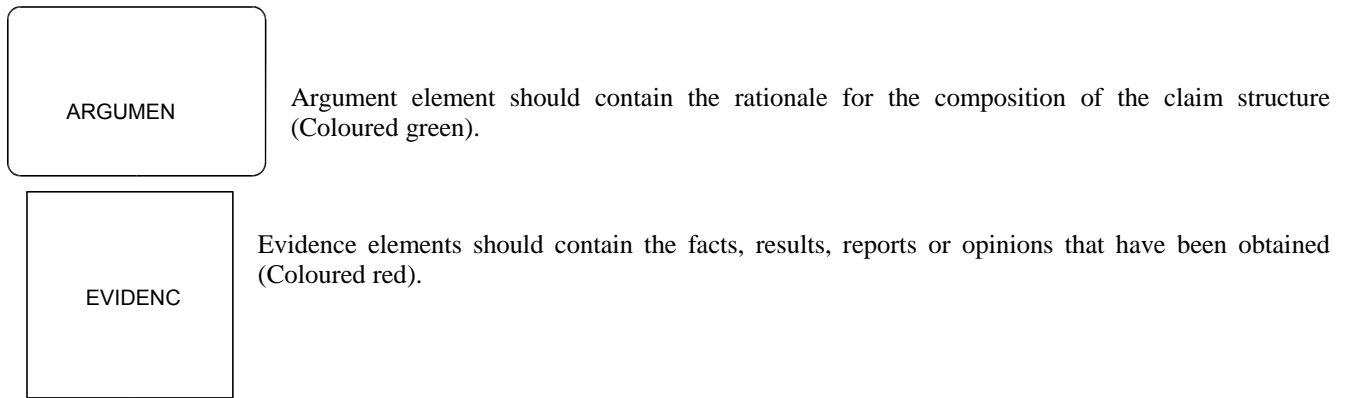


Figure 2 Principle Elements of Claim Arguing Notation

6. METHODS OF COPING WITH COUNTER-EVIDENCE

The UK has evolved some of its standards in this area to be more goal based. This is where the result or evidence requirements are specified, but the methods and tools are not. The method and tool selection is left to the discretion of the project engineers and managers, based on the anticipated level of risk, the required level of confidence and the abilities, preference and availability of support and staff [Maguire 2006]. As stated earlier, methods in this area are still not fully mature, but having acknowledged that, there are a number of options still open to the engineer.

In a well-managed program, during which a safety-critical system is developed, requirements concerning both test criteria and test implementation are contractually defined. If the safety-critical system passes all safety tests, it is considered to be safe. If it does not pass a test, then the reasons for failure need to be revealed. Depending on the severity of the failure and the failure analysis results, the system may still be considered as safe. However, usually re-testing is required after system modification.

A similar approach is taken for the validation of simulations. According to the REVVA methodology [REVVA 2004], the ‘Target of Acceptance’ documents the hierarchical top-down refinement and decomposition of the user's intended top-level use statement into a complete set of examination sub-objectives. This is done in the form of a directed acyclic method until both the precise experimentation conditions and the goal parameters have been fully identified. When the verification, analysis and test methods specified are executed, the individual results are captured as items of evidence. A positive result increases the belief that a Validity Criterion is met, a negative result decreases it. (For a quantitative discussion of the measurement of convincing force and probative force in the context of the REVVA methodology, please refer to [Brade 2004]) If the outcome of a validation activity is unexpected, the reasons for the apparently invalid model behaviour need to be found. It may be the case that the unexpected result does – after analysis and explanation – not decrease the perceived validity of the simulation results. However, if the V&V strategy was well defined, it is more likely that the simulation needs to be adjusted to meet the validity criteria.

Considering the commonly used ‘V’ approach of problem decomposition down the slope (including the definition of test cases or validation criteria) and solution re-composition up the slope (including the assembly of test or validation results) in complex system engineering, counter evidence most dramatically affect the up-slope of the ‘V’. Whatever type of argument you are trying to promote – safety, validity or fitness, the establishing of comparisons across the ‘V’ can be near impossible if counter-evidence removes the evidence you were anticipating when the goals, strategies and target solutions was decomposed. The original decomposition is usually the result of systematic analysis of the dependencies of the desired system capabilities and their foundation. The development of a safety or validity related

goal structure down the 'V' is of as high importance as the development of good (traditional) test criteria and should always be treated as such.

In a less well managed program under resource pressure in the case of significant counter-evidence the decomposition regime may be rapidly re-done to fit available resources – sometimes to fit the evidence that has arisen. The rapid re-think of the decomposition, often without the cross-system consultation that was done the first time around, can lead to mistakes and confusion, loss of configuration control and ultimately more problems than were indicated from the counter-evidence! The re-composition back up the 'V' is then attempted from a poorly understood set of solution evidence, such that claims and arguments are loose and do not offer the best probative force for decision support.

As the well managed project progresses, any required changes to the decomposition structure can be kept in hand using strict configuration control. The Goal-Strategy-Objective notation of Goal Structuring Notation (GSN) can be continually updated to reflect the changing position of evidence – positive or counter – in as much iteration as required.

The two formats commonly used in GSN can be used together to assist in managing the occurrence of counter-evidence and to maintain traceability through one or more adaptations of the decomposition structure. The original decomposition of goals using strategies to obtain solution targets can be retained. The re-composition of evidence using arguments to substantiate claims can be fed from the solutions obtained, be they positive or counter items. Now, as the re-composition path through the argument structure no longer replicate the decomposition path through the strategies any more, combining the two formats gives a powerful tool to demonstrate and re-present to the customer the different argument structures to substantiate the claims for simulation validity or system safety. A graphical example of this dual format use is given in Figure 3 below. The use of two similar, but not identical notations clearly highlights the originally planned decomposition vs. the opportune composition given the evidence at hand.

RELATED WORK

The on-going REVVA Generic Process research contains an explicit phase called "Assess Items of Evidence" during which the probative force is determined. The probative force of positive evidence may be classified "confirmation" (weakest), "strong confirmation", "near prove" or "prove" (strongest). The probative force of an item of counter-evidence could be classified "denial" (weakest), "strong denial", "near disprove" or "disprove" (strongest). Then in the next step of the REVVA Generic Process is "Assess Assembled Body of Evidence" according to the planned strategy, it is evaluated to which degree both the confirming and denying items of evidence impact the belief that a validity criterion is met or not. An approach has been developed such that this subjective evaluation process can be put on a formal, quantitative basis, using Propositional Logic, Possibility Theory [Dubois 1988], and Monte Carlo Simulation. Numerous other approaches are imaginable, based on e.g., Bayesian Network or Belief Theory. Again, research in this area is still underway [See <http://www.vva.foi.se/reports.html>].

A further alternative method is the deliberate use of a variant-rich architecture where many-to-many relationships are all mapped so that if one area fails, a valid configuration can still be maintained. There are architecture tools to assist in this method where variation points within the system (the tested points for evidence) are explicitly captured along with relationships and restrictions between them. In this way, when counter-evidence arises in any particular area disrupting a predicted argument, there is already another valid argument in place. This methodology takes much more effort in the problem domain to identify the many-to-many relationships at each required evidence point. Research in this area is still underway developing tools in Matlab® and Simulink® [See www.software-acumen.com].

6. EXAMPLE OF USE OF GOAL STRUCTURING NOTATION

This example, shown in Figure 1 below seeks to demonstrate the way in which the path through decomposition and re-composition can be different, and where if one item of evidence is lacking, other items of evidence may still support the re-composition argument. It is only an extract from a generic example, the identifiable details have been removed, and as such it has not been subject to project review and release processes.

7. SUMMARY

Counter-evidence is endemic within complex systems and aerospace is no exception. Test and evaluation is dedicated to provide positive evidence support claims about safety, fitness, validity and airworthiness. However, it is potentially disastrous to the project not to have a methodology for coping with counter-evidence. Without a handling mechanism in place, there is actually little value in doing the test and evaluation in the first place. The paper presents an approach how to use two related, but different notations for the decomposition (planning) and re-composition (evaluation) of argument structures, to reveal differences between the planned, contractually agreed argumentation and the finally chosen argumentation. Based on this open presentation of differences their impact can be assessed more easily.

REFERENCES

Dubois, D. and H. Prade. 1988. Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York and London.

Kelly 2003: "A Systematic Approach to Safety Case Management", University of York, York, UK, 2003.

Maguire 2006: "Safety Cases and Safety Reports", Ashgate Publishing, 2006.

MoD 2004: "Safety Management Requirements for Defence Systems Part 2", Interim Defence Standard 00:56, Issue 3. Ministry of Defence, December 2004.

REVVA 2004: PROSPEC. 2004. THALES JP11.20 Report JP1120-WE5200-D5201-PROSPEC-V1.3. Accessible at <http://www.vva.foi.se/reports.html>

Brade, D. 2004. Quantitative Uncertainty Metrics in M&S VV&A. Foundations '04: A Workshop for VV&A in the 21st Century, Tempe, AZ.

Figure 3 : Example use of GSN methods to cope with counter evidence.

