11th ICCRTS
COALITION COMMAND AND CONTROL IN THE NETWORK ERA

A Framework to Model and Measure System Effectiveness

TOPICS
Network-Centric Metrics

Neill Smith
Mission Software Pty Ltd
52 Dempster Avenue,
Balwyn North, VIC 3104
AUSTRALIA
Phone: +61 3 9857 4716
Fax: +61 3 9857 8549
Email: nsmith@acm.org

Thea Clark
Defence Science and Technology Organisation
506 Lorimer St,
Fishermans Bend, VIC 3207
AUSTRALIA
Phone: +61 3 9626 7059
Fax: +61 3 96267084
Email: Thea.Clark@defence.gov.au

Point of Contact: Neill Smith[1]
DSTO Melbourne
506 Lorimer Street, Fishermans Bend
VIC 3207 AUSTRALIA
Phone: +61 3 9626 7460
Fax: +61 3 9626 7084
Email: Neill.Smith@dsto.defence.gov.au

1 This work was funded under contract to DSAD, DSTO Australia

A Framework to Model and Measure System Effectiveness

Neill Smith
Mission Software Pty Ltd
52 Dempster Avenue,
Balwyn North, VIC 3104
AUSTRALIA
Phone: +61 3 9857 4716
Fax: +61 3 9857 8549
Email: nsmith@acm.org

Thea Clark
Defence Science and Technology Organisation
506 Lorimer St,
Fishermans Bend, VIC 3207
AUSTRALIA
Phone: +61 3 9626 7059
Fax: +61 3 9626 7084
Email: Thea.Clark@defence.gov.au

Abstract:

This paper presents further development and extension of the ideas presented at the 2004 CCRTS. The challenges and issues involved in measuring effectiveness are further explored and a method is proposed to guide its measurement based on approaches currently applied within the discipline of Decision Science. The proliferation of Networked systems has increased the importance of being able to model networked systems and to incorporate both quantitative and qualitative measures.

Effectiveness is defined emphasising that it is a problem domain measure which needs to support the comparison of systems. A simple thought experiment clarifies and illustrates various issues associated with aggregating measures of performance (MoP) and comparing MoEs. This experiment highlights the difficulty in creating MoEs from MoPs and prompts a mathematical characterisation of MoE which allows Decision Science techniques to be applied.

Value Focussed Thinking (VFT) provides a disciplined approach to decomposing a system and Bayesian Network (BN) Influence Diagrams provide a modelling paradigm allowing the effectiveness relationships between system components to be modelled and quantified. The combination of these two techniques creates a framework to support the rigorous combination measurement of effectiveness.

The paper concludes with a tabulation of the types of systems to which this approach can be applied.

Introduction

Earlier investigations by the authors [1, 10] and others [2,3,4,5,6,7,8,9] revealed various issues which are still to be resolved when attempting to define and measure system effectiveness. Many of these issues are becoming increasingly important as systems are networked together. The major ones are:

- the need to deal with a system in its broader system context (as part of a system of systems);
- the failure to predict the effectiveness of disruptive technology;
- the increasingly important need to combine both quantitative and qualitative measures (particularly in Networked systems) and
- the relationship between performance measures and effectiveness measures and how to aggregate these measures.

In addition, it is becoming just as important to be able to deal with uncertainty. This uncertainty can occur in the effectiveness measures, the interactions between system components and even in the contributions of system components to overall effectiveness.

A simple thought experiment suggests that many approaches to mapping performance measures to effectiveness measures are inadequate, especially as systems become more networked and complex in behaviour. To overcome the shortcomings of traditional approaches to measuring effectiveness it is proposed that it is critical to measure effectiveness in the problem domain and an approach from Decision Science is used to produce a clear distinction between the problem and solution domain. The problem domain objectives are used to create a Bayesian Network model of the interactions between elements in such a way that the effectiveness of the elements can be combined to indicate overall effectiveness.

MoE definition

Various definitions have been proposed, beginning in the 1950's and progressing through MORS and NATO definitions in the 1980's [6]. These definitions are largely hierarchical and have yet to resolve how to aggregate and propagate performance and effectiveness measures through the hierarchies. These definitions tended to focus on measurement and effectiveness criteria. Sproles (2002) [2, 3] refocussed the discussion of effectiveness back to the more general question of **"Does this meet my need?"** and hence defined Measures of Effectiveness (MoE) as

"standards against which the capability of a solution to meet the needs of a problem may be judged. The standards are specific properties that any potential solution must exhibit to some extent. MoEs are independent of any solution and do not specify performance or criteria".

Needs can be satisfied by various solutions. The solutions may be unique or may share aspects of other solutions. Each solution may (and usually will) have different performance measures.

Sproles distinguishes between Measures of Performance (MoP) and MoE by declaring that MoP measures the internal characteristics of a solution while MoE measure external parameters that are independent of the solution – a measurement of how well the problem has been solved.

The primary focus of the framework proposed here is to compare systems and to produce a rank ordering of effectiveness, as suggested by Dockery's (1986) MoE definition [8]

"A measure of effectiveness is any mutually agreeable parameter of the problem which induces a rank ordering on the perceived set of goals".

The goal is not to derive absolute measures as they do not support the making of comparisons between disparate systems whose measures may be based on totally different characteristics and produce values with different ranges and scales.

The two aspects of these definitions of MoE were emphasised in the definition of MoE by Smith and Clark (2004) [1]

"A measure of the ability of a system to meet its specified needs (or requirements) from a particular viewpoint(s). This measure may be quantitative or qualitative and it allows comparable systems to be ranked. These effectiveness measures are defined in the problem-space. Implicit in the meeting of problem requirements is that threshold values must be exceeded".

In common with Sproles [3], it is accepted that effectiveness is a measure associated with the problem domain (what are we trying to achieve) and that performance measures are associated with the solution domain (how are we solving the problem).

A Simple Thought Experiment

To investigate the challenges of measuring effectiveness a simple thought experiment [10] was developed and various measurement ideas were tested. This experiment had the important characteristics that: (1) everything was measurable, (2) effectiveness (in the problem domain) was easily specified and (3) various measurement regimes could be tested. The experiment consisted of various simple computer programs which all produce identical outputs (except for a deliberately incorrect one). For the purposes of this experiment a computer program can be viewed as a sequence of activities which use and manipulate resources (computing variables, memory, power etc). This is directly analogous to any system which sequences activities and manipulates resources.

Within the context of computer programming, an effective program is one which produces the correct outcome, based on various constraints. Given correctness, it is generally accepted that a program which is faster and uses less resources (whatever this may mean) is more effective. Five programs were developed, for brevity only three will be discussed (MoE3 is deliberating incorrect). The programs (Figure 1) have the following form:

```
Program
   MoE3(in,out);
Var X,Y,Z : real;
        K: integer;
Begin
  X :=0; Y:= 0;Z :=0;
  While (X<101) do
       Begin
        Z := – Y;
        Y := Y + X;
        X :=  X + 1;
        Out(Z);
       End; {of while }
End.
```

```
Program
   MoE4(in,out);
Var X,Y,Z : real;
        K: integer;
Begin
 X :=0; Y:= 0;Z := 0;
 While (X<101) do
   Begin
    Z := – Y;
    X :=  X + 1;
    Y := Y + X;
    Out(Z);
  End; {of while }
 End.
```

```
Program
   MoE5(in,out);
Var X,Y,Z : real;
        K: integer;
Begin
 X :=0; Y:= 0;
 For K:=0 to 100 do
 begin
   Z := – Y;
   Y := Y + K;
   Out(Z);
 End; {of for K}
End.
```

Figure 1 Simple programs

In terms of correctness, the variable Z in Out(Z) defines a correct and hence an effective program. In common with most approaches to measuring effectiveness, the effectiveness of these programs can be characterised by simple measures based on variable usage (that is, memory reference to the variable) and be formulated as an aggregate of these base level measures (MoPs). The MoEs chosen (without loss of generality) were average variable usage and weighted average variable usage (biased towards X or Z). The variables of interest (within the programs) are K X, Y, Z and the measures chosen are (for a single loop of the code):

(usage(K) + usage(X) + usage(Y) + usage(Z)) /4	***average***
($^2/_3$ usage(K)+ $^2/_3$ usage(X) + $^2/_3$ usage(Y) + 2 usage(Z)) /4	***weighted Z***
($^2/_3$ usage(K)+ 2 usage(X) + $^2/_3$ usage(Y) + $^2/_3$ usage(Z)) /4	***weighted X***

These measures have been chosen to be consistent with Utility Theory approaches (that is, Utility is a weighted sum of utility with weights constrained to be between 0 and 1 and summing to 1). Usage is a simple count of the use of the variable. This formulation of effectiveness parallels the approach taken by many authors [11, 12]. The variable usage and MoEs are shown in the Figure 2 where it can be seen that MoE3 and MoE4 have identical values even though MoE3 produces an incorrect output and hence is not effective.

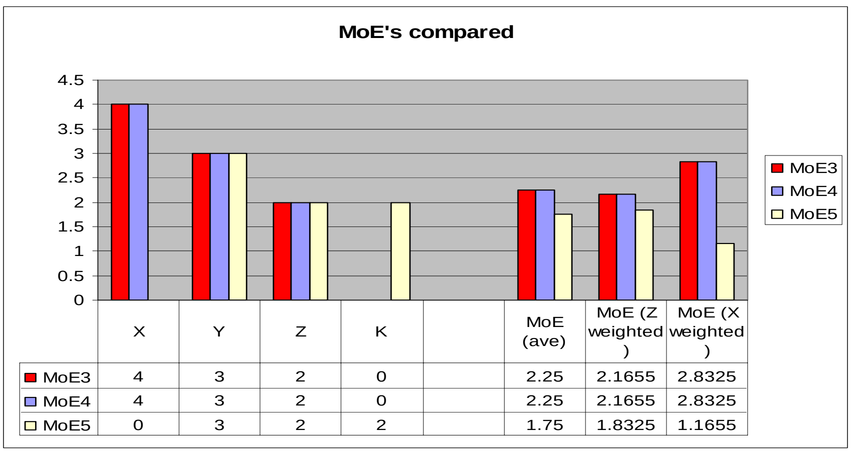


Figure 2 MoE comparisons for thought experiment

This simple experiment highlights a critical aspect of comparing effectiveness between systems, namely that at the low level (performance) two systems can have identical measures and hence produce identical aggregate effectiveness measures even though one system is actually not effective. As an aside, based on any measure used, MoE5 uses fewer resources and hence is comparatively more effective.

Although it is obvious that MoE3 is not effective because the lines of code have been switched (X needs to be incremented before Y is updated). This comparison is not so obvious or easy when complex, networked, or future (vaguely defined) systems are compared. Due to complexity or uncertainty we cannot compare systems based on outcomes (which maybe unknown) but we continue to measure effectiveness on the assumption that internal measures aggregate to indicate effectiveness. This simple experiment provides a counter-example to the validity of this approach and suggests that further thought needs to be devoted to deciding how to measure effectiveness.

Even within the constrained circumstances of this simple experiment various observations can be made:

- Choosing appropriate measures is non-trivial: Even with clearly identified and measurable elements there is no obvious candidate for which internal attributes to use or at what level of granularity.
- Even though everything is measurable, the choice of aggregation method is challenging: is any particular form superior or doesn't it matter? At the performance level no aggregated measure will distinguish between MoE3 and MoE4. MoE5 appears to be best but could some alternative choice of measure have changed this outcome?
- Identical performance measures can be aggregated to give identical effectiveness measures even though one system is not effective. Measures in the solution domain can provide misleading indications of effectiveness in the problem domain. Just creating a measure from MoP's does not guarantee an adequate MoE.
- Given the failure to distinguish effectiveness based on internal measures, maybe more holistic measures (like structural complexity, or runtime) which are in the problem domain should be considered: MoP5 is best, based on any measure. This strongly suggests that, based purely on internal performance measures, it is most effective solution and it would be expected that holistic measures would confirm this view.
- Given compliance, comparative effectiveness makes sense. In other words, given the achievement of requirements, other attributes can enable the ranking of systems.

Measuring Effectiveness

The simple thought experiment suggests that, for ranking of comparable systems, effectiveness measures need to capture the problem domain requirements and map the solution back to this problem domain. This mapping will not always be direct, as shown above. MoE needs to compare and aggregate effectiveness measures in the problem domain and these measures may be influenced and quantified by measuring attributes in the solution domain. Measurement Theory [13, 14] highlights the challenges and constraints on the validity of this mapping process.

MoE Properties
To achieve the aims of the definition of effectiveness (Smith and Clark, above), effectiveness measures should have the following properties:

1. The measure needs to **increase as effectiveness increases** (not all weighted sums will do this),
2. The measure needs to be **bounded above by an ideal system and bounded below by zero for non-compliance,**
3. To manage **complexity and allow for system decomposition**, any measure needs to represent and support system decomposition and aggregation (for equivalent systems aggregate measures must be equivalent regardless of level of decomposition).
4. To facilitate comparisons between systems (which may have different internal characteristics and differing primary purposes) it is necessary to normalise the final effectiveness scores. The range [0,1] is chosen (with 0 denoting an ineffective system and 1 denoting a perfectly effective system)
5. Ideally the measures should be **ratio scales** [13, 14] which means that they have a natural zero point and numbers which are multiples of each other directly indicate their value. (For example, a system with an effectiveness measure of 0.8 is twice as effectiveness as a system with a measure of 0.4). Ratio scales directly support the achievement of properties 1 to 4.

These properties should be used to choose amongst alternative approaches to defining MoEs. Two approaches from Decision Science meet these mathematical requirements and are considered as candidates for measuring effectiveness:

1. Multi-attribute Utility Theory (MUAT) [11, 12, 15] and Value Focussed Thinking (VFT) [15, 16].
2. The probabilistic modelling technique of Bayesian Network and Influence Diagrams (BN) [10, 17, 18].

Both these approaches deal with measures between 0 and 1 with MUAT measuring utility and BN using probability.

MAUT and VFT

Multi-attribute Utility Theory (MAUT) is used to formulate and evaluate utility functions and Value Focussed Thinking (VFT) is used to determine the fundamental objective and value hierarchy used to derive the utility function. Values are the attributes that the fundamental objectives are measured against to confirm their attainment. VFT also creates a means-end network indicating how the fundamental objective will be attained. The value hierarchy is derived from the problem domain whilst the means-end network of a system is determined in the solution domain.

To measure effectiveness the fundamental objectives are critical as they provide the values against which effectiveness will be assessed; VFT dictates that fundamental objectives should have these properties:

- **Essential**: indicate consequences in terms of the fundamental reasons for interest in the situation
- **Complete**: include all fundamental aspects of consequences

- **Measurable**: to define objectives precisely and to specify the degree to which objectives may be achieved
- **Operational**: to render the collection of information required for an analysis reasonable considering the time and effort available
- **Decomposable**: to allow separate treatment of different objectives in the analysis
- **Non-redundant**: to avoid double counting
- **Concise**: to reduce no. of objectives needed for analysis
- **Controllable**: address consequences influenced by choice of alternatives
- **Understandable**: to facilitate generation and communication of insights.

Expected Utility, Figure 3, expresses the worth of a consequence. Expected utility is the probabilistically weighted sum of the utilities. In Decision Science under uncertainty, it is considered "rational" to choose between a' and a" based on the value of expected utility. MUAT provides a basis for measuring expected utility but the formulation of the utility function is dependent on many complex independence conditions being established. These independence conditions are difficult to establish and verify and consequently MUAT is considered unsuitable for measuring effectiveness in complex domains, like NCW. (Often these conditions are assumed to be true, thus weakening the validity of the measure). The MUAT formulation is directly analogous to the approaches used in the simple thought experiment, described earlier.

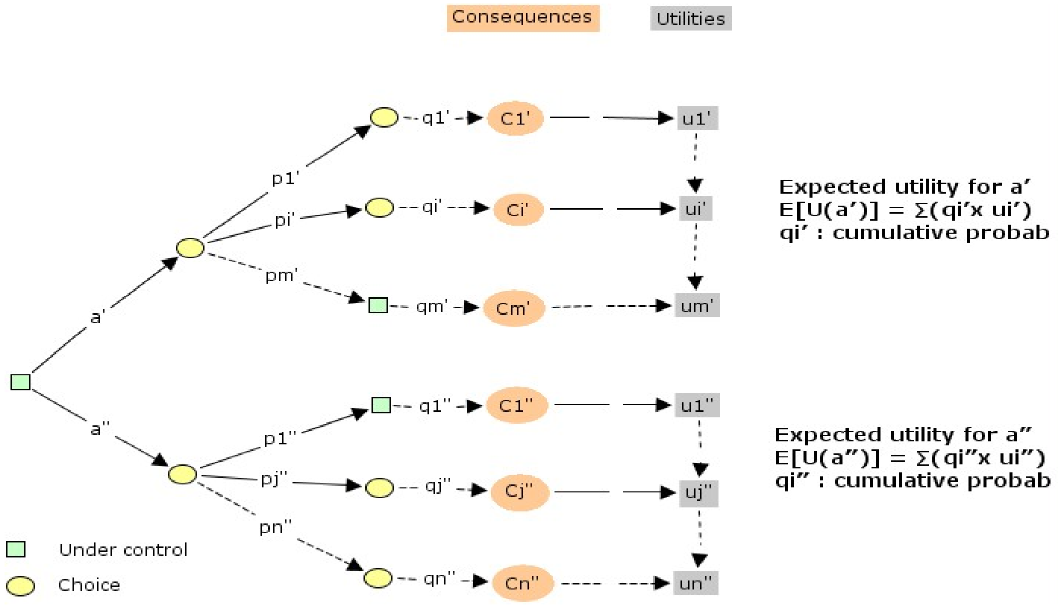


Figure 3 Expected Utility

So in summary, VFT provides a well-grounded, consistent mathematical framework to analyse and model the problem domain to produce a value hierarchy and a means-end network. This is achieved by focussing on the fundamental objective and the values that allow their attainment to be measured. As such it is an ideal first step in an approach to decompose systems to facilitate effectiveness measurement.

Influence Diagrams and Bayes Networks

The VFT model and values provide the characteristics against which effectiveness should be measured. But it is still necessary to provide a mechanism that can aggregate these measures, derived from the value model. Since the VFT describes relationships it is appropriate to find a paradigm which can model relationships and allow their causative effects to be aggregated. Such an approach is Influence Diagrams based on Bayesian Inference (commonly called Bayesian Networks, BN). In common with MUAT, BN provides a well-grounded, consistent mathematical framework which (in addition) supports the forward and backwards propagation of evidence[1]. So it is able to answer the questions:
If I observe something

- ***what may have caused this?***
- ***what outcomes will this influence?***

So within an effectiveness context, this rule can be used to answer the question "what is the effectiveness of a system given the effectiveness of another system?"

BN are acyclic directed graphs [19] with each node representing a variable and each arc representing a causal relation between two nodes [20, 21]. Arcs (or their absence) represent the conditional dependence or independence between nodes. The strength of influence is quantified by a conditional probability distribution for each node given its predecessors [19]. A BN is able to update the probabilities in uncertain nodes (using Bayes rule) given evidence obtained from related nodes. This property and the intuitive way BN model complex relationships among nodes make them a suitable technique for building causative models [17, 18]. There is evidence [22] to suggest that their predictive value is robust against incorrect estimates of the probability values populating the nodes as long as the causative links are correct and have appropriate weighting.

For military systems, and particularly networked systems based on new technology, there is often insufficient data (or operational experience) to quantify a system's effectiveness. So recourse is often made to expert judgement to guestimate effectiveness. In addition, to classical statements of effectiveness, such subjective (qualitative) judgement needs to be handled. Cox's work [23, 24] is accepted as the justification for the use of subjective probability[2] within a Bayesian framework [18]. Cox derived Bayes' rule (and other probabilistic rules) from the rules of logic and two axioms[3] without reference to the frequentist definition of probability. He thus argued

---

[1] ***Bayes law states: Pr(a|b) = Pr(b) x Pr(b|a) / Pr(a)***.
The notation ***Pr(a|b)*** means the probability of a given b.
And by simple rearrangement: Pr(b|a) = Pr(a) x Pr(a|b) / Pr(b).
This justifies forward and backward propagation of evidence.

[2] Cox (1946) claimed that probability theory, in essence, has involved two ideas: "the idea of frequency in an ensemble and the idea of reasonable expectation". Reasonable expectation is the probability of an event which is not based on extensive trials but more on subjective judgement and expert opinion. It provides a "measure of the reasonable expectation of an event in a single trial".

[3] These axioms (1961) are:

(1.i) "the probability of an inference on given evidence determines the probability of its contradictory on the same evidence"; and

(1.ii) "the probability on given evidence that both of two inferences are true is determined by their separate probabilities, one on the given evidence, the other on this evidence with the additional assumption that the first inference is true".

that subjective probability is equally valid for modelling causal relationships under uncertainty [18].

So by measuring effectiveness and using values between zero and one it is possible to aggregate their effects as long as their causative relations can be established. The outputs from VFT provide both these components so it is possible to build a BN to model effectiveness in such a way that total system effectiveness can be inferred from subsystem effectiveness [17]. This assessment of effectiveness can be performed in both a "forward" and "reverse" direction; that is, given subsystem effectiveness total system effectiveness can be determined or if a system is effective, measures of required subsystem effectiveness can be inferred.

The significant outcome from this two stage framework (VFT followed by BN[4], Figure 4), is that MoE (from the Values Hierarchy) is assessed by aggregating MoEs, not MoPs (which are described by the means-end network). That is, the BN models the impact of one component on the effectiveness of another component. To reduce the complexity of the models produced, it is recommended that the high level BN be defined purely in terms of effectiveness nodes[5]. This separation of concerns (only working in the effectiveness domain) greatly simplifies the system model. This simplification is supported by Keeney's argument [15] that many decision assessments fail because the means-end network is intertwined with the value hierarchy.

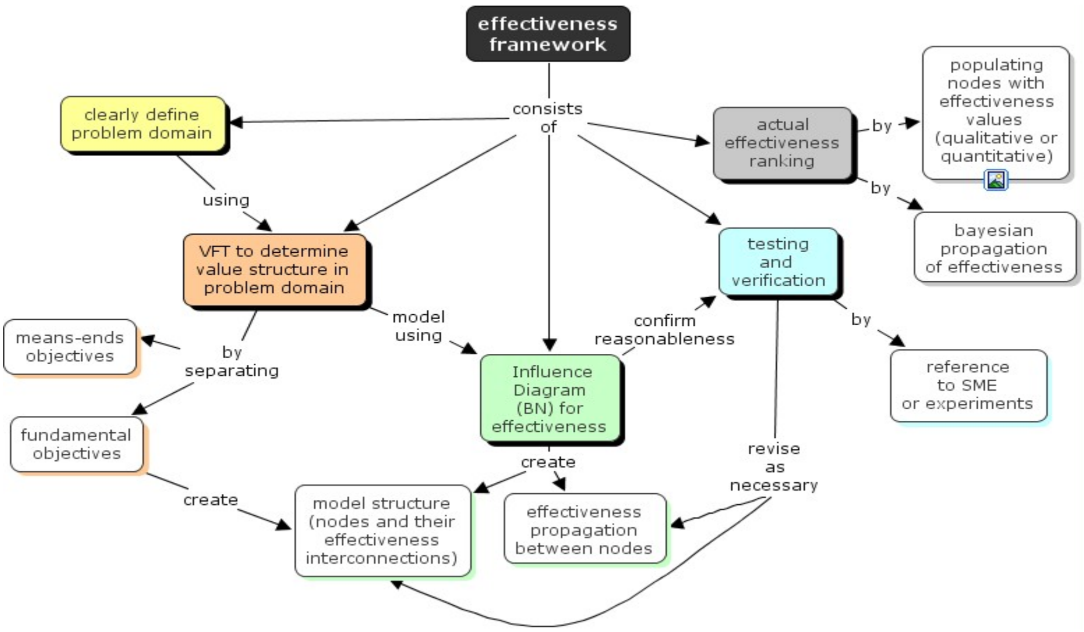


Figure 4 Framework for Measuring Effectiveness

Both these "axioms" can be considered to be valid in the domain of measuring effectiveness; that is, effectiveness and its inverse are related and the effectiveness of a two systems combined is dependent on the effectiveness of the first system, if the systems are causally related.

[4] Maxwell et al [16] describes an approach to using VFT to build BN to assist military decision-makers which proposes a method of eliciting and transforming fundamental objectives hierarchies into BN. Their approach has not been used in the examples that illustrate the framework proposed in this paper as decision-making considers different parameters to effectiveness.

[5] If performance measures can be causatively mapped to an effectiveness measure then the BN approach can be also used to calculate this measure but this mapping should be done within a subsidiary model.

To illustrate this framework, two examples are presented. One is from the domain of film photography where a pre-existing value model is directly mapped into a BN. This example is used to test the validity of MoE aggregation, in a domain that is well understood. The second is from the military domain where VFT is used to define a value hierarchy that is used to develop a BN. Software supporting BN (GeNIe, [25]), is used to illustrate this approach.

Example 1: Feininger's Perfect Negative

Feininger [26] presents a succinct model[6] describing the vital operations and the control devices used in photography to produce a "perfect negative", Figure 5. The arrows indicate cause and effect relations where the node at the head of the arrow is influenced by the node at the start of the arrow.

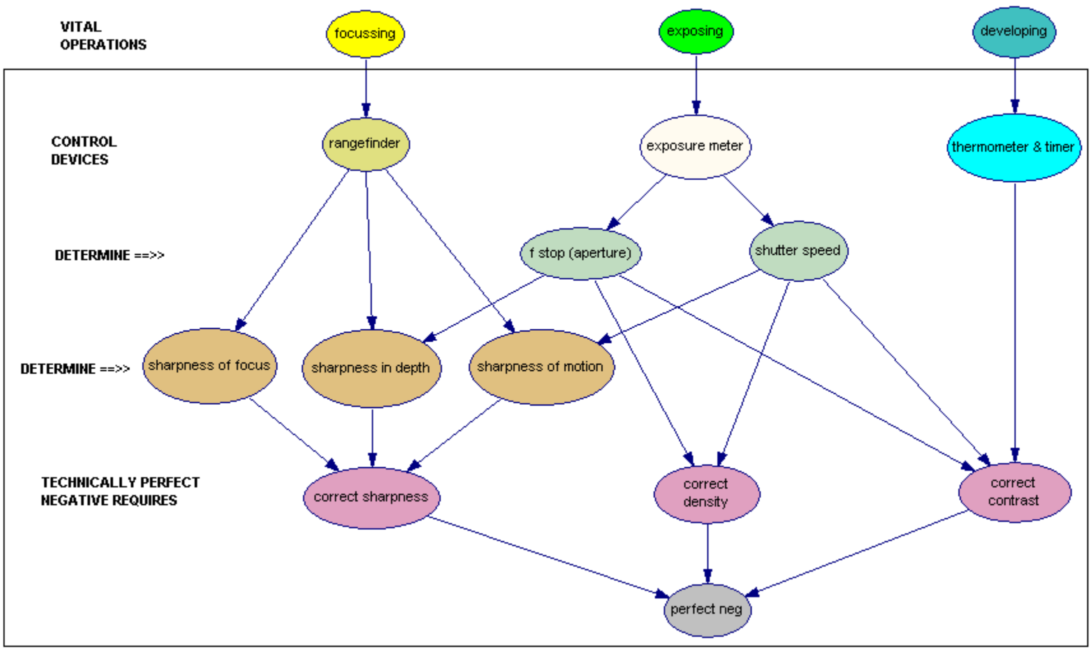


Figure 5 Feininger perfect negative

This model can be directly modelled as a BN as it satisfies the requirement of being acyclic, that is, no successor node links back to a predecessor node. The nodes within the enclosing box express the values (and fundamental objective) required to give a perfect negative. It is particularly interesting that this model is absent of traditional performance measures. The model is very much in the problem domain and largely eschews solution issues, that is, it doesn't specify performance measures like resolution (sharpness), tonal range (contrast, density) etc. These may be used to quantify the value nodes but the value nodes can just as usefully be specified by subjective judgments[7].

[6] It exemplifies a good model in that it clearly identifies the critical aspects of making a negative and is multi-layered with causative connections which "jump" layers (ie has some complexity). That is, it is sufficiently rich to accurately describe the process without being so complex that it cannot be understood.

[7] In the case of subjective judgment, the validity of the BN is largely judged against the BN producing outcomes which match expert expectations.

Feininger's model is largely timeless in the sense that it is just as applicable to the first cameras as it is to the last of the film cameras. His causal model has decomposed the problem space such that at any level, effectiveness can be judged by reference to its preceding nodes, that is, a perfect negative is based on sharpness, density and contrast being correct, etc.

This model (enclosed in the box) can be directly mapped to a BN, Figure 6. The values in the nodes show the probabilistic characterization of the states. They should be interpreted as effectiveness measures. These states, within a node, need to be exhaustive and sum to one (100% in GeNIe)[8].

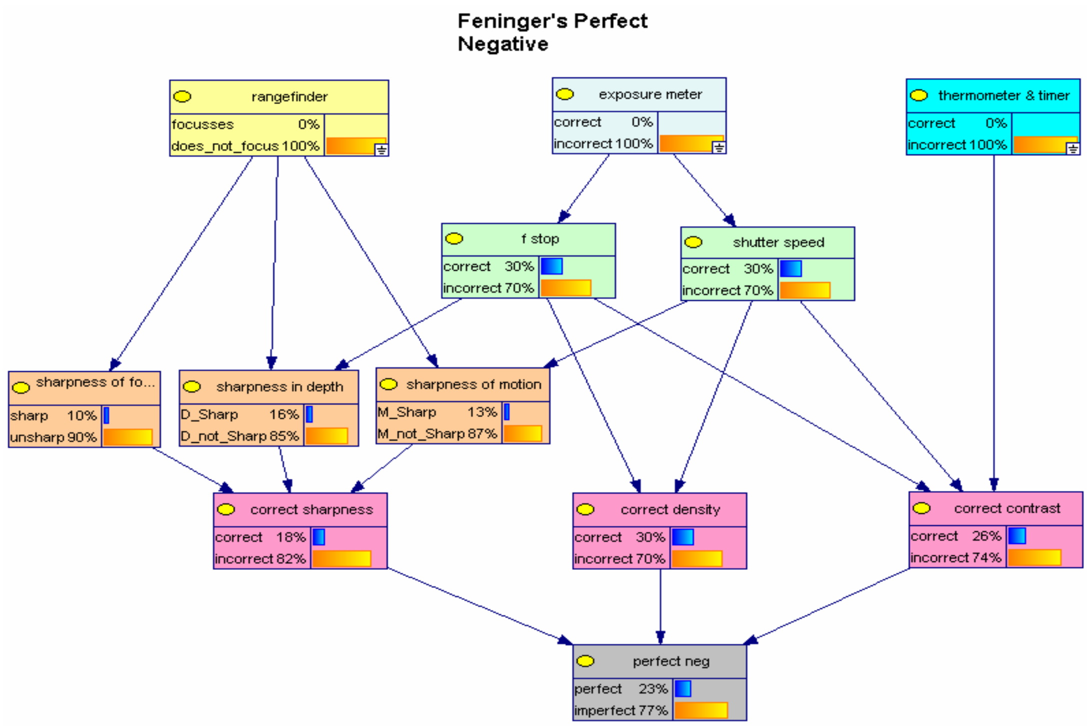


Figure 6 All control devices ineffective

By assigning effectiveness to each node (effectiveness of rangefinder, aperture (f-stop) etc) and the contribution of each node to its successors (including joint contributions) it is possible to measure the extent to which a perfect negative is attained. Because a BN supports backwards propagation (of evidence), the assignment of effectiveness to the perfect negative node can be used to indicate the effectiveness required of the predecessor nodes[9].

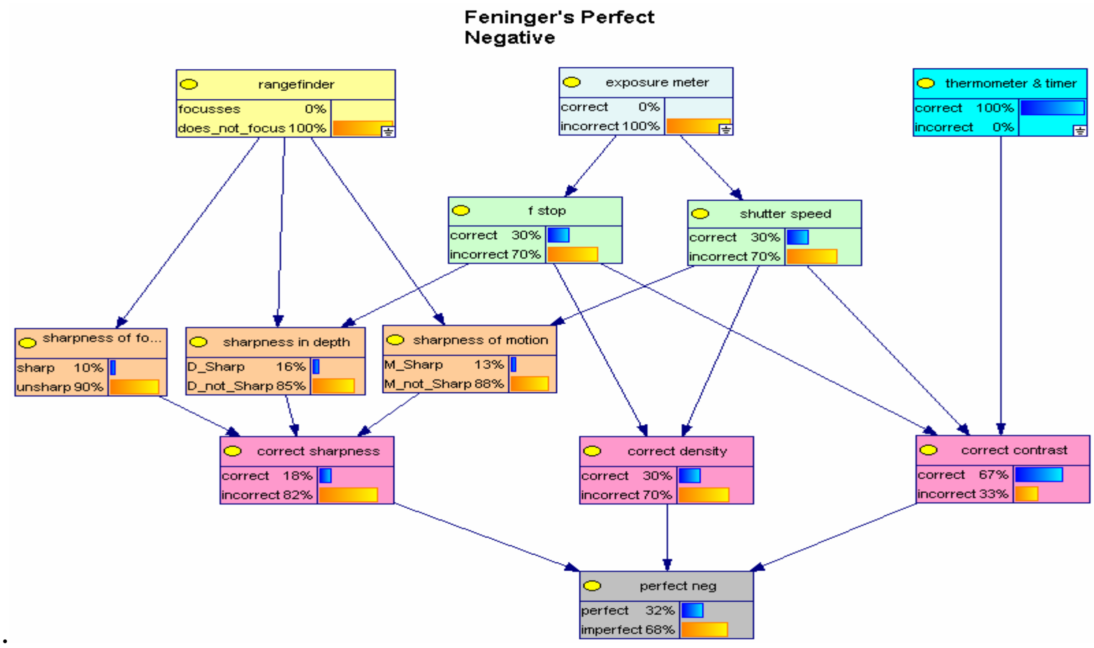
To test the reasonableness of his model various extreme cases can be used. The following figures illustrate what can be achieved by experimenting with various values in the nodes. In Figure 6 all the vital operations are set to failure. The outcome is not 100% "imperfect" as the transfer of effectiveness allow for some imprecision in the measures. In this case, the Bayesian propagation of effectiveness is consistent with intuition.

---

[8] Any minor discrepancies are due to display issues associated with numerical rounding

[9] In a practice, both forward and backward propagation are used to get the values to a steady state (so the assignment of values to some nodes will allow the unassigned nodes to be updated)

By correcting the *developing*, by getting the temperature and development time (*thermometer & timer node*) correct, Feininger's model indicates (Figure 7) that the *contrast* has improved, but an imperfect negative is still produced because the *density* and *sharpness* are still inadequate. This example shows how the effectiveness impacts flow through the BN by the process of evidence propagation.



.

Figure 7 Improved development

The final example (Figure 8) shows the impact of good inputs on the creation of a perfect negative.

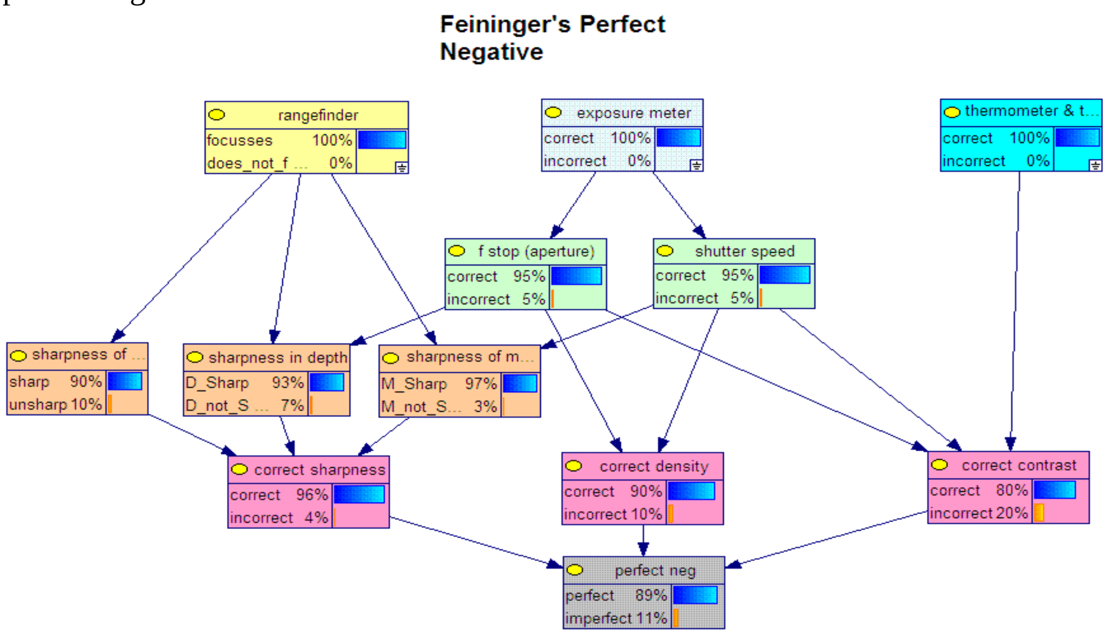


Figure 8 Good Inputs

A good model needs to be a fair representation of the relationship between nodes and the probabilistic values used to indicate the strength of the causal relationship between the nodes need to fairly represent their impact on system effectiveness. Given that this is the case, Feininger's model provides a mechanism to explore the impact on effectiveness of changing the effectiveness of the nodes. Feininger's model[10] reflects the maturity of the photographic field, but its simplicity belies the complexity which would be present if both the MoP and MoE were intermingled.

The goal should be to derive comparable models within the defence domain, by focussing on the problem domain and the relationships between the problem elements.

Example 2: Surveillance and Response

A military example involving multiple systems (both human and technological) networked together is surveillance and response. For surveillance a VFT analysis could produce a decomposition as shown in Figure 9. Note that the arrows here are interpreted as saying that the node at the head of the arrow is composed of or uses nodes at the tail of the arrow. That is, the value hierarchy is not a necessarily a causal structure. The land and littoral nodes are not expanded for brevity.

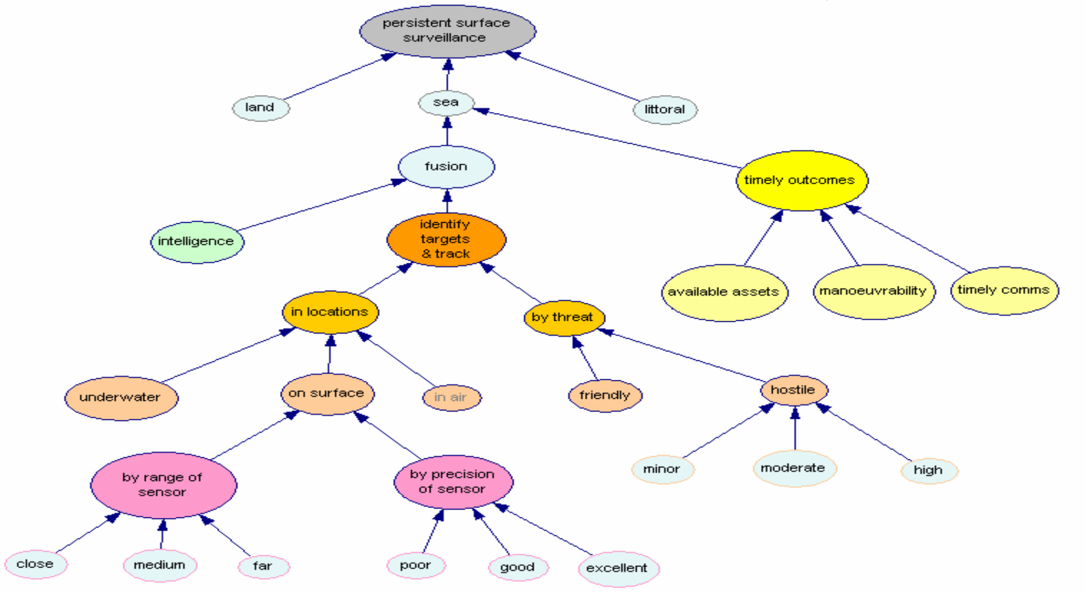


Figure 9 Partial VFT decomposition for surveillance

This model is not complete, but shows that the hierarchy is decomposed through complete enumeration of sub-factors influencing a value node (fundamental objective) as dictated by VFT.

Based on this VFT analysis of surveillance, the following causal model (Figure 10) can be developed (using GeNIe [19]). To include the *response* node it has been necessary to introduce *interpretation* to the model. The yellow nodes indicate a direct causal chain from persistent surveillance through to some response. The blue nodes indicate activities or attributes which can influence this causal chain. Treating each node as having an effectiveness measure in the range [0, 1] and aggregation rules based on its predecessors allows the model to be quantified.

---

[10] The preceding example was predefined by Feininger (without reference to VFT, based on his photographic expertise) and satisfied the conditions applicable to VFT decomposition.

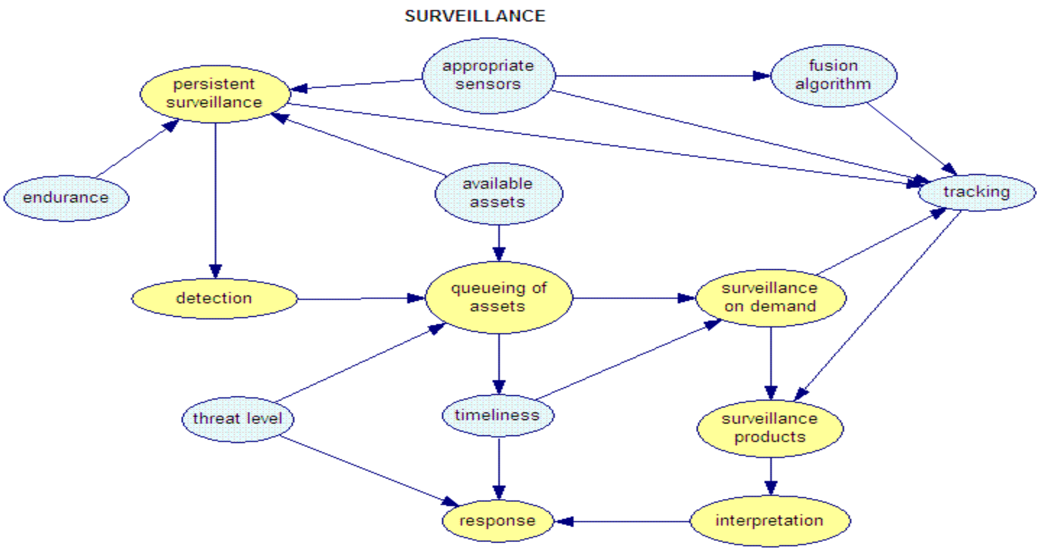


Figure 10 Causal chain for surveillance

To convert Figure 10 to a BN the following assumptions shown (in annotation boxes) in Figure 11 have been made.

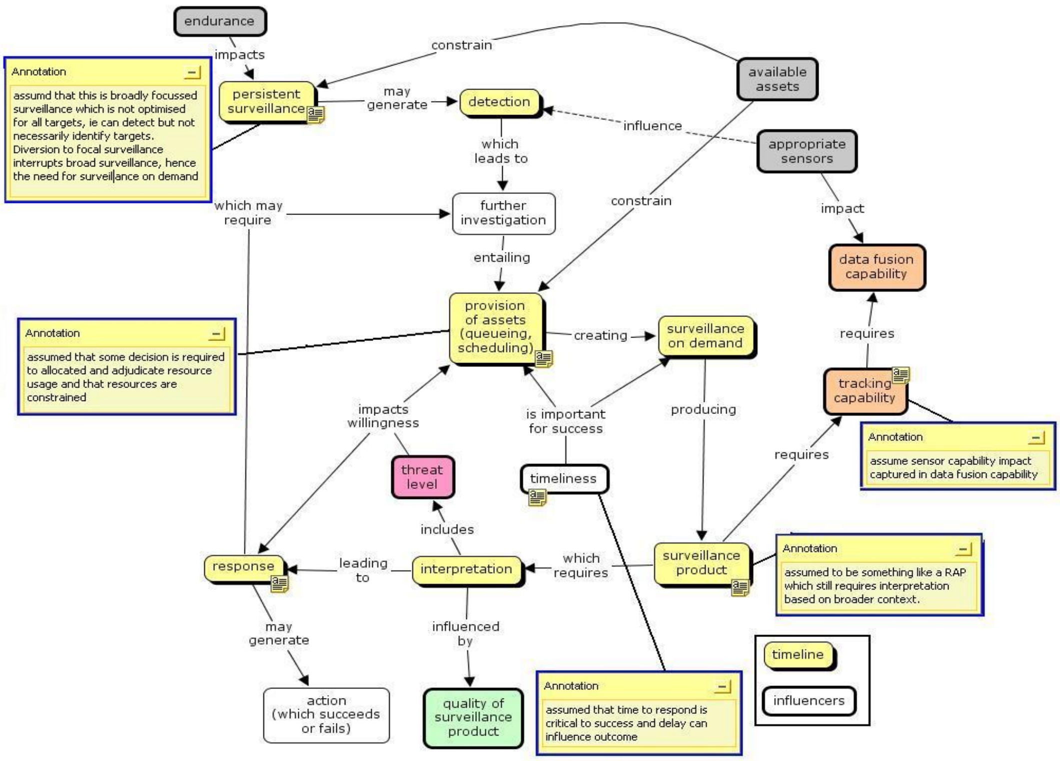


Figure 11 Assumptions used to derive causal BN

The BN nodes are populated with discrete states but they do not need to be binary; more states can be included to increase the fidelity of the model. Extra states increase the effort involved in quantifying impacts (*fusion algorithm* has three states). These nodes are populated with effectiveness values between 0 and 1 (0%-100%) and they need to sum to 1 (100%) to conform to the probabilistic requirements of BN, as shown in Figure 12.

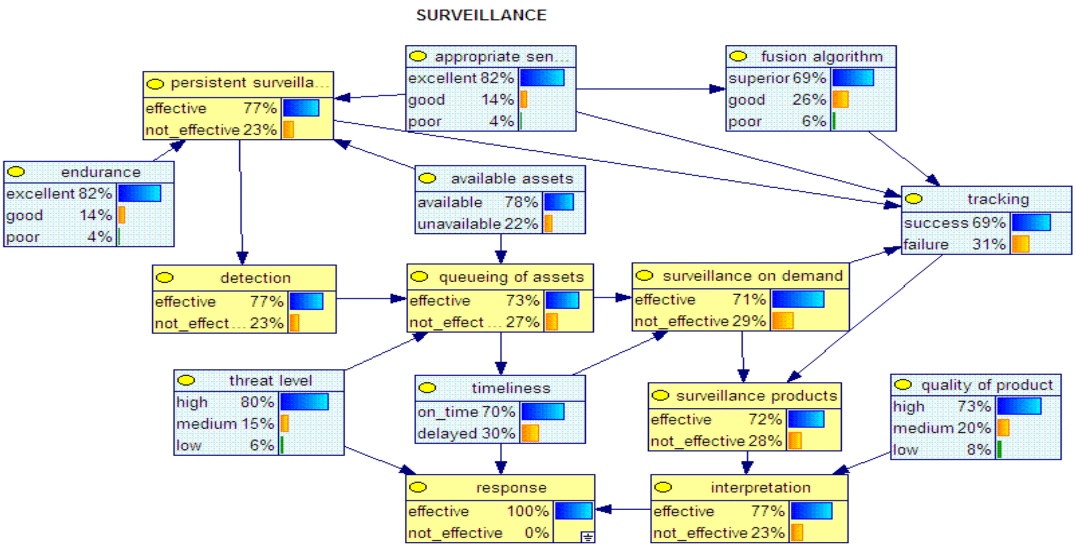


Figure 12 Impact of effective response

By instantiating the model with effectiveness values the reasonableness of the model can be evaluated. By running the BN in reverse (that is, only setting *response* to 100% effective, the effectiveness requirements of the other nodes (to achieve this outcome) are calculated, as shown in Figure 12. By setting the available assets to *unavailable*, Figure 13, and using the default values, the BN shows how this would affect the effectiveness of the other nodes. Note that the *response* is now ineffective.

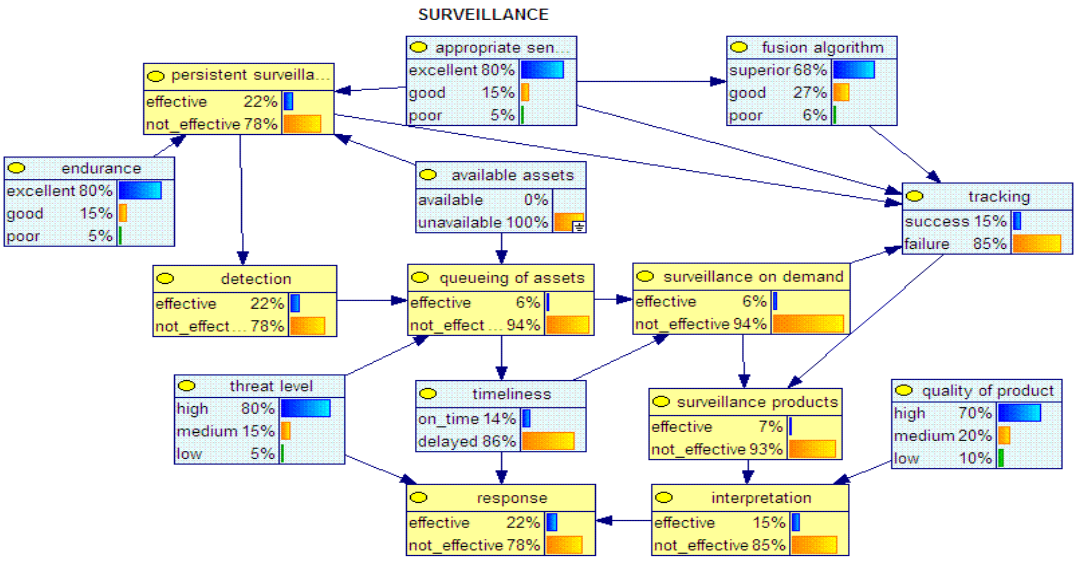


Figure 13 No assets available

A more realistic example of this approach could test the impact of the *fusion algorithm* on the effectiveness of the *response* when some nodes are given fixed values: namely, *appropriate sensors* (excellent), *timeliness* (on time). The outcomes are shown in Figure 14, Figure 15 and Figure 16. These permutations of the effectiveness values are presented to show the propagation of effectiveness values through the BN and confirm the reasonableness of the effectiveness predictions. This example shows *response* improving as the *fusion algorithm* improves.

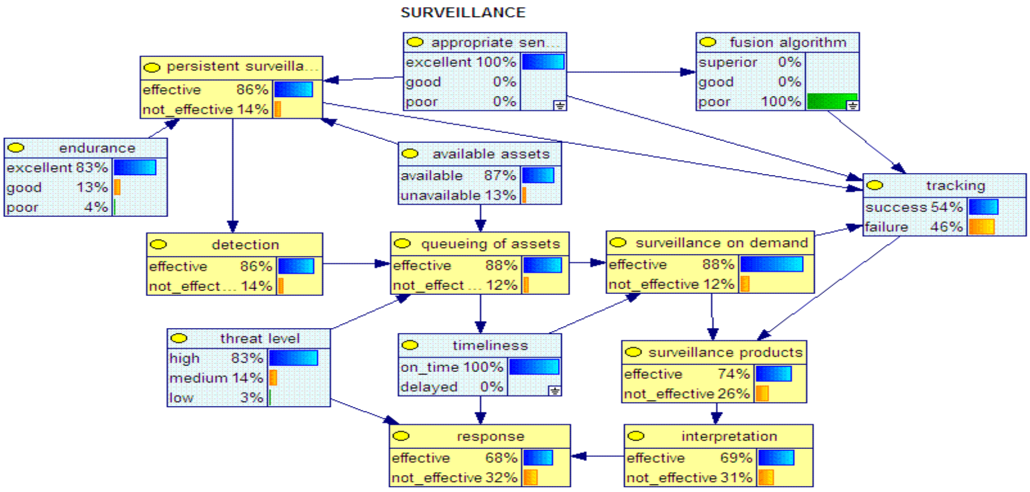


Figure 14 Impact of Poor Fusion Algorithm

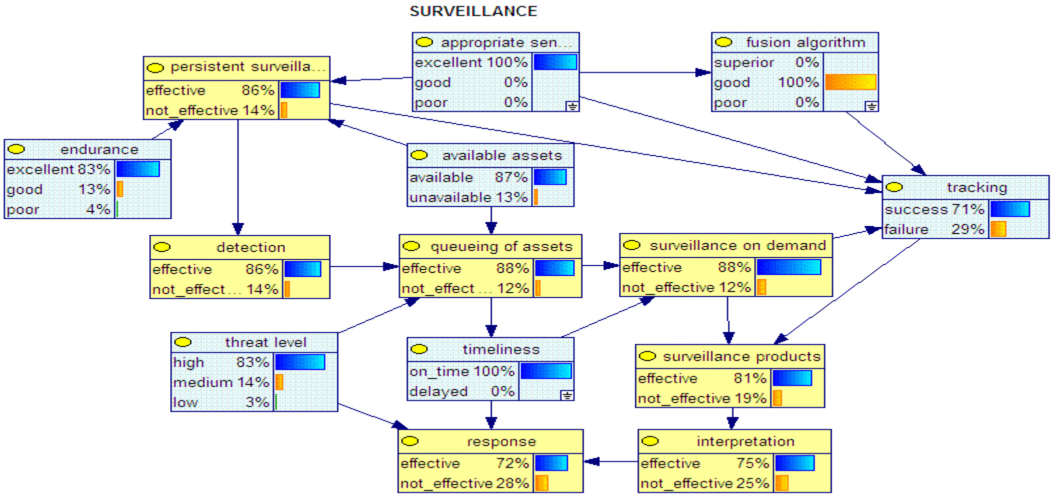


Figure 15 Impact of Good Fusion Algorithm

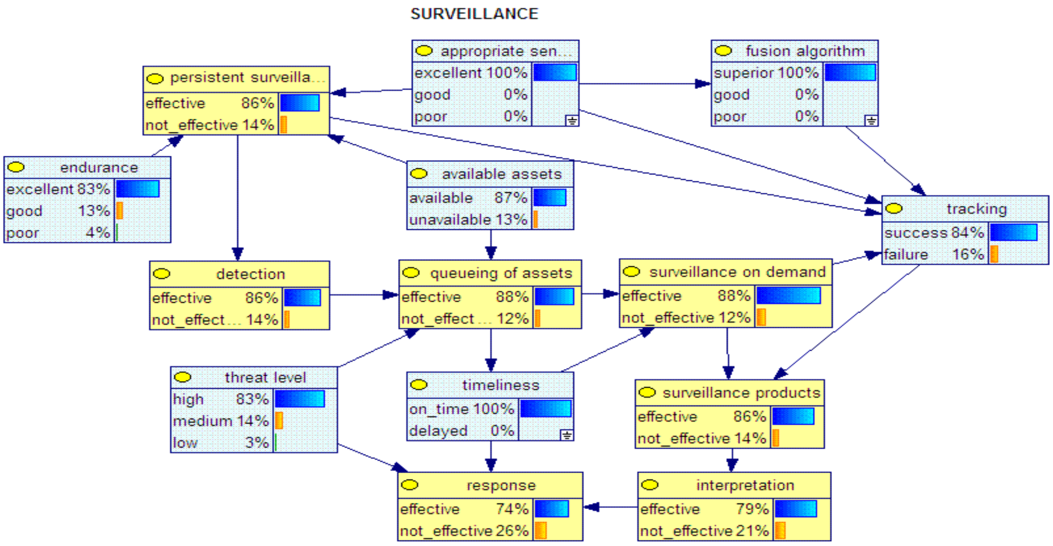


Figure 16 Impact of Superior Fusion

The assignment of effectiveness values to the nodes will be derived from expert knowledge or observation, by reference to subsidiary effectiveness models (other BN) or by aggregating performance measures[11] to indicate a node's effectiveness.

A simple BN for *surveillance and response* has been presented which supports the assessment of effectiveness based on the effectiveness of the nodes which contribute to *response*. With this model it is possible to determine the impact of changing system effectiveness values to observe their impacts on total system effectiveness. These variations are equivalent to performing (high level) comparisons between alternative system configurations. For example: what is the impact on achieving an effective *response* if no assets are available to provide *surveillance on demand* or what would be the impact of using a system with inferior *data fusion* capability.

This approach provides an alternative to other approaches to measuring effectiveness which supports a mathematically valid method for incorporating uncertainty and propagating the impact of effectiveness through a networked system[12].

---

[11] As illustrated by the thought experiment described earlier in the paper, this aggregation of MoP's can only be done when the process is consistent with the tenets of Measurement Theory. This process can be done using BN as long as the causal links and impacts between the measures can determined.

[12] The BN developed is not just a simple network connectivity diagram; it is based on the fundamental objectives of a system and their causal relations.

Applicability of Framework

Table 1 correlates the applicability of this framework to systems according to their characteristics. It defines the system type by nature of the technology, whether a VFT approach can determine its value structure, and whether a MoE which can be developed.

MoE characterization

system type	Values Determined	Model Determined	MoE possible	Example
well defined interactions (causal relationships known)	yes	yes	yes	well known physics; known physical laws, for example: ballistic missile
undefined interactions	yes	partial	only at high level	NCW (now)
disruptive technology	yes(but wrong value structure)	partial	yes, but measures wrong attributes	digital versus film cameras.
sustaining technology	yes	yes	yes	improving radar technology
new approaches	partial (no experience to determine values at low level)	partial (lack of knowledge)	partial(maybe at higher level, using value from comparable systems)	early stage of radar, or totally new surveillance technique. Automated reasoning system
evolving needs (assured technology)	yes	yes	yes(but value trade-off difficult)	stealth fighter (how is effectiveness of stealth combined with existing attributes)

Table 1 Applicability of Approach

Within the context of this framework it is surmised that an inability to create a value model (fundamental objectives and values) means that effectiveness cannot be modelled and measured. Sproles [3] gives examples of surrogate measures of effectiveness, but, within this framework, such measures cannot be validated and verified to be correct.

For the *undefined interactions* category, it may be possible to determine the value structure but the causal (means-end) relations may be difficult to determine.

Disruptive technology (and disruptive processes, like NCW) provide an interesting example of the problems of measuring effectiveness and particularly with placing too much emphasis on performance measures. That is, disruptive technology can be compared to current approaches but are often poorly placed because the traditional performance measures do not capture the value provided by the disruptive technology.

For *new approaches*, high level value structures may possibly be defined but the creation of a means-end network to describe how the value will be created is a major challenge. Consequently it will be difficult to establish the causal relationships.

Evolving needs describes systems where the initial values are no longer challenging and, in fact, the challenge may be the integration of new needs without destroying current value. Such systems create particular challenges in measuring effectiveness. That is, how do you weight the importance of traditional measures against the new measures, particularly as needs are being refined and evolve.

Conclusion

Consistent with the effectiveness definition used in this paper it is critical to maintain a focus on the "needs" of the system stakeholder. These needs define what is required for a system to be effective.

A simple thought experiment highlighted the difficulty in mapping MoP to MoE. This difficulty suggests that there are two problems in measuring effectiveness; one is the mapping of MoP to MoE and the other is: given MoEs, how can they be combined to indicate the overall effectiveness of a system. This later problem is addressed in this paper and the former needs to be developed using the principles of Measurement Theory (which is outside the scope of this paper[13]).

This thought experiment suggested that MoEs should have particular properties that support their aggregation. Given these properties, two approaches from Decision Science were investigated to support this aggregation: VFT with MUAT and BN. Given a VFT decomposition of a system, a BN allows the rigorous propagation of effectiveness values between causally connected nodes. In addition, BN support the use of both qualitative and quantitative effectiveness measures. This is particularly important as new technologies, scenarios and approaches are used by the military.

The increasing complexity and uncertainty created by NCW demands that methods of measuring effectiveness be developed which ensure that a focus is maintained on the problem domain. This is particularly important as technology changes at a rapid pace and existing performance measures lose their relevance. Complexity and uncertainty also increase the importance of being able to combine objective measures and subjective judgement into the process of measuring effectiveness. The framework described here supports both of these goals with broad applicability to military systems.

[13] The approach described here is applicable to mapping MoP to MoE as long as causality and the mathematical requirements of BN are met.

References

[1] Smith, N., Clark, T., *"An Exploration of C2 Effectiveness – A Holistic Approach"* 2004 Command and Control Research and Technology Symposium

[2] Sproles, N., *"Establishing Measures of Effectiveness for Command and Control: A Systems Engineering Perspective"* DSTO report DSTO.GD-0278. 2001

[3] Sproles, N., *"Formulating Measures of Effectiveness"*, *System Engineering*, Vol 5, No. 4, 2002

[4] Malerud, Stein; Feet, Else Helene and Thorsen, Ulf, *"A Method for Analysing Command and Control Systems"*, 1998 CCRTS Proceedings, Naval Postgraduate School, Monterey, California, June 29- July 1.

[5] Pinker, A., Samuel, A.H., Batcher R., *"On Measures of Effectiveness"* Phalanx, Dec 1995.

[6] Starr, S.H., *"C4ISR Assessment: Past, Present, and Future"* presented at 8Th ICCRTS, Washington, 17-19 June 2003

[7] Christensen, C. M., *"The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail"*, Harvard Business School Press, 1997.

[8] Dockery Dr. J. T., *"Why Not Fuzzy Measures of Effectiveness"*, *Signal,* May 1986

[9] Green, John M. & Johnson B.W. *"Towards a Theory of Measures of Effectiveness"* presented at CCRTS June 2002

[10] Smith, N., *"A Proposed Framework to Formulate and Measure Effectiveness"* DORS Conference, Melbourne, Australia August 2005.

[11] Keeney R., Raiffa H. *"Decisions with Multiple Objectives: Preferences and Value Tradeoffs"*. Cambridge University Press, 1993.

[12] Lt Col. Jackson, J., A., Jr., Lt Col. Jones, B.,L., Major Lehmkuhl, L.,J., *"An operational Analysis of Value-Focused Thinking to Future Air and Space Capabilities"* May 1996.

[13] Zuse, H., *"A Framework of Software Measurement"* Walter de Gruyter . Berlin . New York 1998.

[14] Hamalainen R.P., Salo, A.A., *"Rejoiner The Issue is Understanding the Weights"* Journal of Multi-Criteria Decision Analysis Vol 6, 340-343 (1997)

[15] Keeney, R.L., *"Value-focused thinking: A path to creative decisionmaking"*. Harvard University Press. 1992.

[16] Maxwell, D.,T. and Buede, D.,M., *"Composing and Constructing Value Focused Influence Diagrams: A Specification for Decision Model Formulation"* Journal of Multicriteria Decision Analysis; Jul/Oct 2003.

[17] Druzdzel M.J. *"Explanation in Probabilistic Systems: Is It Feasible? Will It Work?"* Intelligent Information Systems V Proceedings of the Workshop held in Deblin, Poland, June 2-5, 1996. from http://www.pitt.edu/~druzdzel/publ.html sighted 30/3/06

[18] Agosta, J. & M,. Gardos, T., *"Bayes Network "Smart" Diagnostics"* Intel Technology Journal Vol 08 Issue 04 Nov. 17 2004.

[19] Kraaijeveld, P. C. & Druzdzel, M.J., *"GeNIeRate: An interactive Generator of Diagnostic Bayesian Network Models"* Decision Systems Laboratory, School of Information Sciences and Intelligent Systems Program.
From http://www.sis.pitt.edu/~pieterk/papers/DX05.pdf

[20] Charniak, E., *"Bayesian Networks without Tears"* AI Magazine Winter 1991
[21] Maxwell, D., *"What every good OR analyst should know about Bayesian Networks"* Phalanx Mar. 1998
[22] Pradham, P., Henrion, M., Provan, G., Del Favero, B., & Huang, K., *"The sensitivity of Belief Networks to Imprecise Probabilities: An Experimental Investigation.*From http://www.sis.pitt.edu/~pieterk/papers/DX05.pdf
[23] Cox, R.T., *"The algebra of probable inference"* Johns Hopkins Press 1961
[24] Cox, R.T., *"Probability, Frequency and reasonable Expectation"* American Journal of Physics Vol. 14, No. 1. 1946.
[25] *Genie*: Decision Systems Laboratory, University of Pittsburgh
[26] Feininger, A., *"The Complete Photographer"* Thames Hudson 1966, Seattle, WA, 1994.