SANDIA REPORT

SAND2004-2902
Unlimited Release
Printed July 2004

Measures of Effectiveness: An Annotated Bibliography

Philip L. Campbell

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865)576-8401
Facsimile: (865)576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800)553-6847
Facsimile: (703)605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2004-2902
Unlimited Release
Printed July 2004

Measures of Effectiveness: An Annotated Bibliography

Philip L. Campbell
Networked Systems Survivability & Assurance

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0875

Abstract

The purpose of this report is to provide guidance, from the open literature, on developing a set of "measures of effectiveness" (MoEs) and using them to evaluate a system. Approximately twenty papers and books are reviewed. The papers that provide the clearest understanding of MoEs are identified (Sproles [46], [48], [50]). The seminal work on value-focused thinking (VFT), an approach that bridges the gap between MoEs and a system, is also identified (Keeney [25]). And finally three examples of the use of VFT in evaluating a system based on MoEs are identified (Jackson et al. [21], Kerchner & Deckro [27], and Doyle et al. [14]). Notes are provided of the papers and books to pursue in order to take this study to the next level of detail.

This report was developed as a customer deliverable.

Table of Contents

1 Introduction 7
2 Reviews 9
2.1 Heartling et al. 9
2.2 Perry 11
2.3 Green & Johnson 12
2.4 Moffat et al. 13
2.5 Doyle et al. 14
2.6 Allen & Demchak 14
2.7 Grier et al. 16
2.8 MORS Study Plan 17
2.9 Hicks et al. 18
2.10 Freeman & Serfaty 18
2.11 Jackson et al. 19
2.12 Kerchner & Deckro 22
2.13 Brown 23
2.14 Keeney 25
2.15 Glenn 30
2.16 Sproles 31
3 Tangential Work 35
4 Conclusions and Future Work 37
Glossary 39
References 41

List of Figures

Figure 1 Connections 10
Figure 2 Figure 1 of Allen & Demchak [6] 15
Figure 3 Five Components of a VFT Value Model 20
Figure 4 The Influence of VFT on Decisionmaking Processes 27

List of Tables

Table 1 MoPs vs. MoEs 33

1 Introduction

The purpose of this report is to provide guidance, from the open literature, on developing a set of "measures of effectiveness" (MoEs) and using them to evaluate a system. The starting point for this report was the following set:

- a literature search on "measures of effectiveness;"
- a literature search on "value-focused thinking;"
- a literature search on work by Richard F. Deckro; and
- a collection of papers, primarily from recent MORS conferences (e.g., Perry [39]).

Along the way I added another starting point after perusing a colleague's bookshelf.

The task is to summarize the above material and, within the time allotted, follow leads to promising material.

I divided the material I considered into two categories:

1. those that I reviewed, and
2. those that I considered tangential to the matter at hand.

The rest of this report is organized as follows. In Section 2 I present summaries of and comments about material I have reviewed. In Section 3 I present brief reviews of material I consider to be tangential. And in Section 4 I present conclusions and future work. This is followed by a Glossary and a Bibliography.

Note:

> In the literature the acronyms, MoP, MoE, MoFE, and MoM, are sometimes written with an upper-case O and sometimes with a lower-case o. I use the lowercase o in this document, as a matter of taste. However, I do not change the case in direct quotes.

2 Reviews

In this section I review the following items:

1. Heartling et al. [19],
2. Perry [39],
3. Green & Johnson [17],
4. Moffat et al. [33],
5. Doyle et al. [14],
6. Allen & Demchak [6],
7. Grier et al. [18],
8. MORS Study Plan [34],
9. Hicks et al. [20],
10. Freeman & Serfaty [15],
11. Jackson et al. [21],
12. Kerchner & Deckro [27],
13. Brown [8],
14. Keeney [25],
15. Glenn [16], and
16. Sproles ([46], [48], and [50]).

Figure 1 shows all of the material, both the material in this Section and in Section 3. The figure shows diagrammatically which item led me to which other item. The figure shows two veins:

1. the literature search on VFT, leading to Keeney [25], and
2. Green & Johnson's paper, leading to Sproles' papers ([46], [48], [50]).

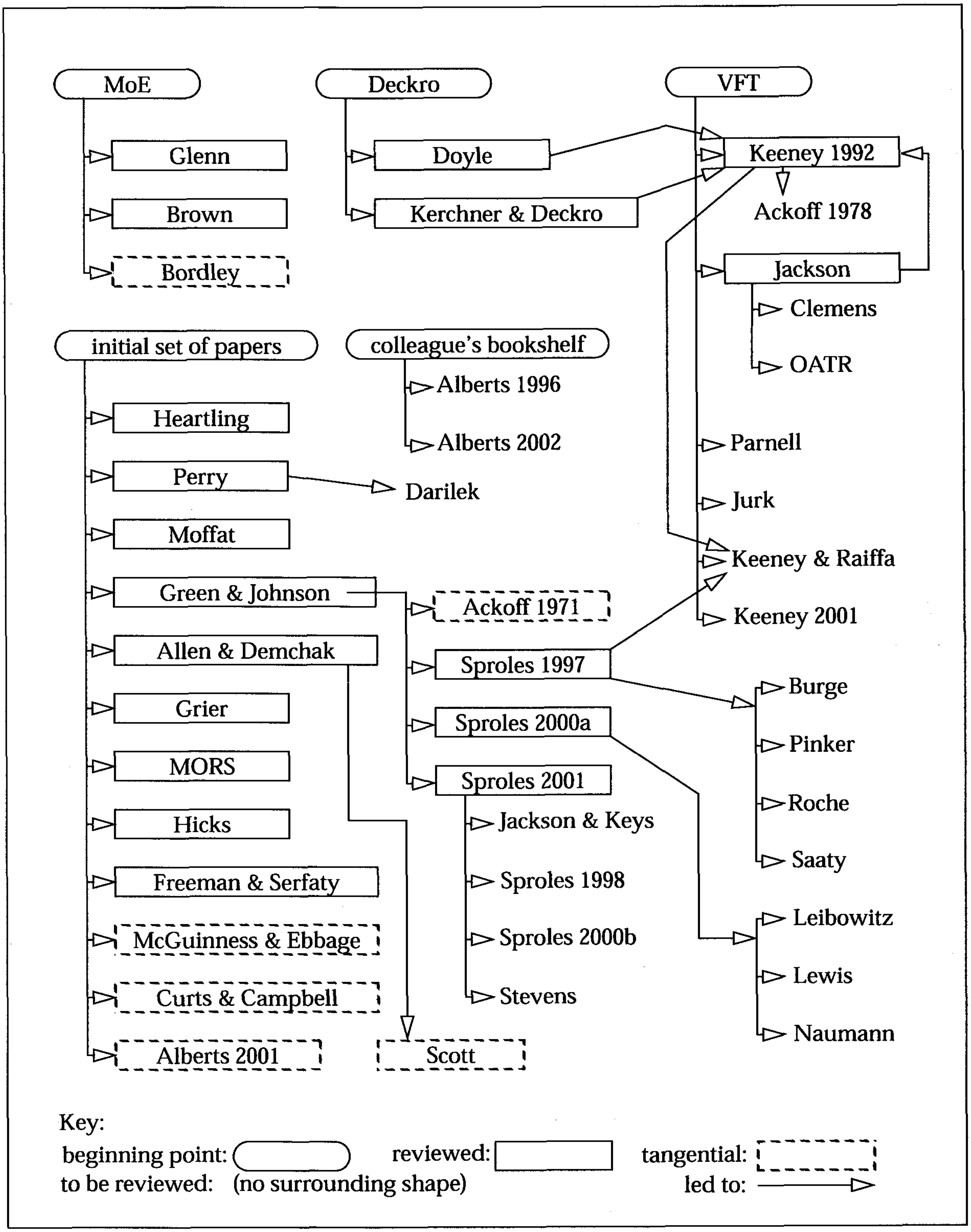
The rest of this Section contains reviews. In each review I present a summary, my comments, and a list of items to which the current item leads.[1] The reviews are shown in the order that I encountered the material, based on my initial impression of the importance of the material relative to the others then available to me. I have not made a second pass through the comments, so my understanding as it came to be by the end of the document is not reflected in previous comments.

2.1 Heartling et al.

According to Heartling et al. [19], the use of Information warfare (IW) weapons has been

1. For simplicity's sake I only note the first time I encounter a new item. For example, if both paper A and paper B reference paper C, and the review of paper A precedes paper B in Section 2, then I show only paper A leading to paper C, not paper B. Some of the references to Keeney's work are an exception to this rule.

Figure 1. Connections



analyzed in "high level theoretical policy" discussions and as they relate to specific targets but they have not been incorporated into the targeting process. Heartling et al.'s purpose is to fill this gap by presenting a model of that activity. Their model is intended to be a "first cut." In their own words, the following is the "true challenge:"

> ...to design MOE's that compare hard-kill physical destruction weapons with soft-kill IW weapons on a level 'playing' field. This includes the development of procedures to measure synergistic effects of using IW weapons as force multipliers. (page 52)

Heartling et al. add one effectiveness measure (diversion) and six attributes (e.g., standoff range) to current sets. They use these and a commander's constraints to "screen" (i.e., remove) possible weapons from consideration of use against a given target. They then use probability of success (Ps), a value that is given for each weapon, to select weapons to maximize a given commander's "aspiration level." They next revise the model to add a second attack that precedes the first in which only those weapons that have a "high degree of post-execution stealth" are considered. They show that this revised model significantly increases the aspiration levels.

Comments:

> Heartling et al. articulate a model that seems straightforward. Further work is required in two areas. First, they need to show that their expanded set of MoEs and attributes is at least sufficient. Second, they need to provide (a) algorithms to measure those MoEs and attributes, and (b) a case for the validity of those algorithms. As Heartling et al. note, additional "extensive" research is required.

Leads to: (none).

2.2 Perry

Perry [39] proposes a "probability model of knowledge," based on Shannon's concept of entropy.

Perry defines knowledge in terms of "situational awareness." For the purposes of this paper he assumes that the Critical Elements of Information (CEI) are knowledge of enemy targets. Within that assumption, some of the important factors contributing to the CEI are

- the number of confirming reports;
- the number of dis-confirming reports;
- the reliability of the sensors and sources;
- terrain occlusions;
- multiple phenomenology (confirmation from N different sensor types is strong than from N sensors of one type); and
- the age of the information received.

"Dominating maneuver" will increasingly be the measure of combat outcome, as measured, for example, by "battlespace control." Speed—the minimum time for any or all of the units to move to another Named Area of Interest (NAI) in the battlespace—is a measure of battlespace control. Perry says that his paper offers "a beginning at understanding the relation between knowledge and its effect on combat outcomes in a quantifiable way" (page 38). Perry argues that increased ability to maneuver ground troops (i.e., dominating maneuver) is the promise of "information-age technologies." (He notes that "information superiority" is relatively clearly defined in Army Vision 2010. However, "information dominance" is murkier but is a proper subset of information superiority.)

Comments:

> How does Perry's paper help us get to MoEs?
>
> Meanwhile, Shannon's model depends on knowing probabilities. We can use Shannon's model for communications written in English because we can compute the frequencies at which the bounded number of letters appear. However, it is not clear how we could compute this for the CEI that Perry enumerates. How, for example, do we compute the probability of the number of confirming reports? Is there even a bound on the number of such reports?

Leads to: Darilek [12] (which appears to be the larger report of which Perry's is the summary).

2.3 Green & Johnson

Green & Johnson [17] argue that a theory of MoEs does not exist. The series of MORS workshops and symposia—early 1984, January 1985, June 1986, autumn 1992, spring 1993—laid the foundation, and Noel Sproles' work built on that foundation. Green & Johnson build on it further using Russell Ackoff's "system-of-systems" model [1].

The MORS work is based on describing a system as a set of processes. The approach consists of a "theory" (a set of definitions) and an "analytic framework," referred to as the Modular Command and Control Evaluation Structure (MCES), that enables determination of the boundary of the system. DPs are used to develop MoPs which, in turn, are used to develop MoEs which depend upon the environment and which, in turn, are used to develop MoFEs. Each MoM could be a MoP or an MoE.

Unfortunately the MORS approach did not propagate to the outside community. What was needed was a way to compute the probabilities of effectiveness of hierarchical (i.e., system-of-systems) sets of processes. (I believe that the importance of the system-of-systems model is that it permits recursive decomposition, until the processes are concise enough to lend confidence about one's ability to gauge the probability of their effectiveness.) This has been provided by Ackoff [1]. For example, Ackoff provides a way to compute probabilities of effectiveness of serial processes, given the probability of effectiveness of each individual process. Similarly, Ackoff provides a way to compute the same for parallel processes, for combinations of parallel and serial processes and, finally, for processes of processes.

Meanwhile, Green & Johnson note a belief—that they attribute to Rudwick [42]—that MoM's are a function of "availability, dependability (reliability), and capability," to which they add a fourth from Marshall [30], namely "survivability."

Comments:

Where is MCES defined?

What support is there for Ackoff's computation rules?

Why are probabilities used? Because this is the real world in which parts fail? Because software is, by its nature, buggy? Because the environment of a system tends to change over time? Or is it because the approach itself is imprecise? Where do the values for the probabilities come from?

How is software hierarchical? It may be developed using language constructs that are hierarchical, but how is an executable binary image hierarchical?

How do we use Rudwick' and Marshall's beliefs about attributes? How do we get from "availability," say, to probability of effectiveness of processes (or the other way)? If we cannot go from one to the other, what is the value of their list?

Leads to: Ackoff [1], and Sproles ([46], [48], [50]).

2.4 Moffat et al.

Moffat et al. [33] present recommendations and additions, for use in the UK, to the NATO "Code of Best Practice for Command and Control." [35] The most important part of Moffat et al.'s paper, from my standpoint, is their review of the relative measures, from Chapter 14 of the NATO document. MoMs consist of DPs, MoPs, MoEs, and MoFEs. DPs are "properties or characteristics of the inherent physical system." DPs, in turn, provide the grist for MoPs "which focus on internal system structure, characteristics and behaviour." MoPs, in turn, provide the grist for MoEs "which focus on the impact of C2 within the operational context." And MoEs, finally, provide the grist for MoFEs "which focus on how a force performs its mission or the degree to which it meets its objectives." The relationships are shown using set notation

MoM = { DP, MoP, MoE, MoFE }

and a diagram

DP ⇒ MoP ⇒ MoE ⇒ MoFE

Comments:

This hierarchy provides a pattern for the relationship between the measures. It is sensible, certainly, but how valuable is it? How characteristic is it of the pattern followed by others? What is there to argue against a different ordering?

Leads to: (none).

2.5 Doyle et al.

Doyle et al. [14] demonstrate a method of ranking courses of action (COAs) for a given scenario based on a decision-maker's "values," using Value-Focused Thinking (VFT). The focus of the demonstration is information operations (IO), which, in times of war, becomes IW, which presents three general targets to attack:

1. "information,
2. information systems, and
3. information-based processes." (page 5)

Doyle et al. decompose these three targets into 19 targets, based on "engineering principles, information theory, military doctrine, expert opinion, and operational experience" (page 6). They add four types of costs, developed by the two decision-makers who participated in the study, arriving at 23 MoMs. Doyle et al. then interview "information operators" to determine what appears to be the effectiveness of their 23 measures, and they also determine the relative effectiveness of the measures, based on the decision-makers. Doyle et al. then compute relative rankings of the three COAs that they consider, by summing the product of relevance and effectiveness for each measure for each COA. Doyle et al. proceed to demonstrate the tools available via VFT, namely

- "characteristic plots," which show the relative rankings of the three, top-tiered objectives (corresponding to the three targets) for the two decision-makers;
- "sensitivity analysis" graphs, one for merit and a second for cost, which show the range, for each decision-maker, within which the relative rankings of the three COAs remain the same (the narrower the range, the more sensitive is the ranking); and
- "direct comparison" graphs, for each pair of COAs, which show the relative ranking for each objective for a pair of COAs.

Doyle et al. claim that their demonstration is evidence of the "viability" of their approach.

Comments:

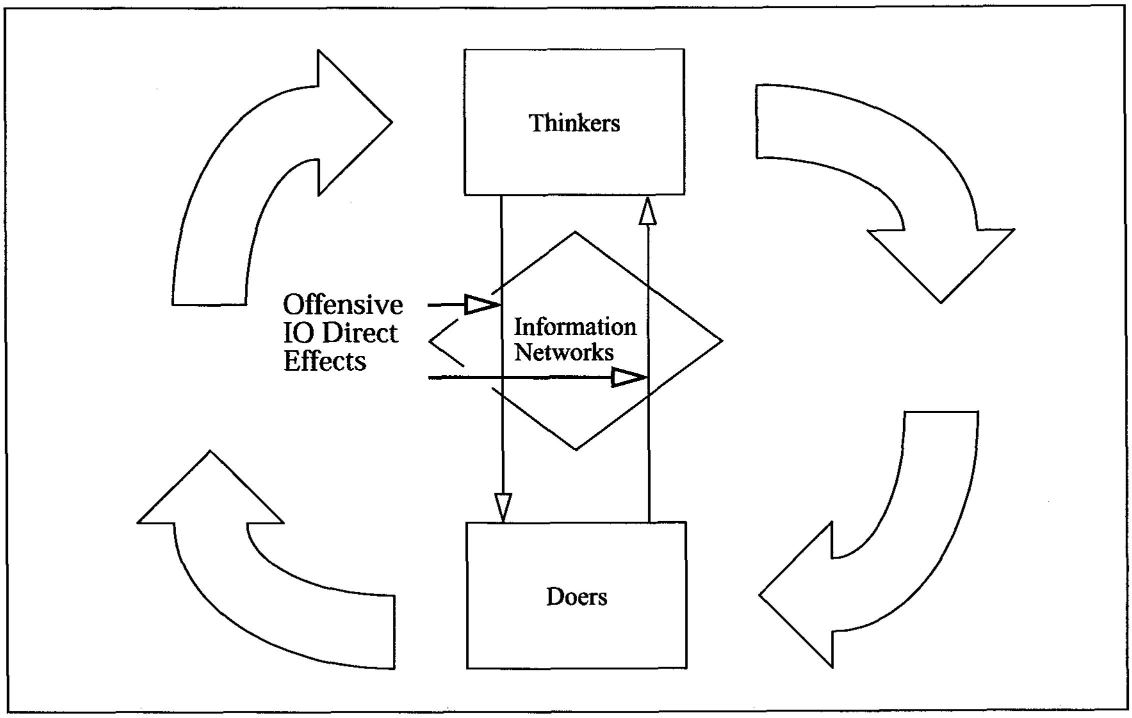
> The general approach is clear but not so with the particulars. Doyle et al. use the same term, such as "objective" and "value," in more than one way, obscuring their study. The road from "values" and costs to the "additive value function" (functions?) appears to be the central idea of the study and is diagrammed in Figure 4 of their paper (page 9) but the path is foggy in places. For example, how is the effectiveness of one of the 19 targets to the achievement of a COA computed? What part does VFT play in all of this? And who are the "information operators?"

Leads to: Keeney [25].

2.6 Allen & Demchak

Allen & Demchak [6] present a "conceptual" model that is intended to include all of information operations (IO). The model is diagrammed in Figure 1 of Allen & Demchak, which is reproduced here in simplified form as Figure 2.

Figure 2. Figure 1 of Allen & Demchak [6] (simplified)



The model consists of four parts:

1. Thinkers,
2. Doers,
3. the Information Networks that connect them, and
4. an "open loop"[2] from Thinkers to Doers and from Doers to Thinkers.

An attacker has four avenues of "direct effects"—"delay, destroy, read, manipulate" (page 11). The attacker can apply these direct effects on an Information Network only. "Indirect effects" are modeled by the open loop around the Thinkers and Doers. Boyd's "OODA loop" (Observe, Orient, Decide, Act) is represented by the Thinker/Doer boxes and their interactions. The conceptual model consists of three categories of models:

- organizational behavior,
- Joint Military Operations Models, and
- Information Network Attack and Defense

2. That is, there is an arrow from Thinkers that points down to Doers, and another arrow that starts where the first one points and continues on to Doers, but the two arrows do not meet. There is a similar set of arrows from Doers to Thinkers. The loop is implied so I refer to it, for lack of a better term, as an "open" loop. Allen & Demchak just refer to "large arrows" and do not describe the loop that the arrows represent.

corresponding to the three elements of the model, respectively. These three models could constitute an "IO Federation of Simulation," according to Allen & Demchak. Organizational behavior models, for example, consist of four types:

- machine (e.g., the Soviet military),
- organism (e.g., the Israeli Defense Force),
- cybernetic (e.g., the US military), and
- social (e.g., coalitions).

IO defense and attack are modeled by slicing the model vertically—the Thinker-to-Doer part of the open loop (i.e., the right side of Figure 2) modeling defense, and the Doer-to-Thinker part (i.e., the left side of Figure 2) attack—though this does not seem clean.

Meanwhile, security levels are needed to enable the model to be both used and protected. Allen & Demchak suggest four partitions, named A through D, for elements of instances of this model, where

- A labels elements that are "well-known and well-advertised,"
- B labels elements for which the method is classified but the effect is not,
- C labels elements for which both the method and the effect are classified, and
- D labels elements, the existence of which could imply a classified capability.

Allen & Demchak's paper is a "very condensed" version of a DARPA report.

Comments:

> The value of this model, it appears, is twofold. First, it provides a view of the entire arena of IO. Second, it provides a frame of reference, establishing relative positions for the constituting sub-models.
>
> But what is a "conceptual" model? Are not all models conceptual? What are the alternative conceptual models? and why is this one superior?
>
> The arrows constituting the open loop have either a source or a target but not both. Graphs do not have such elements. What then do these arrows mean? Why not use two arrows instead of four?
>
> And how does this get us to MoEs?

Leads to: Scott [45].

2.7 Grier et al.

Grier et al. [18] claim a "unique combination of factor analysis and response surface methodology" to "reduce the dimensionality" of a simulation. That is, they answer the question, "What proper subset of factors would provide the same analysis results?"

They list 34 metrics—such as number of tanks destroyed—which they use as grist for their analysis, and nine campaign objectives. Grier et al. then apply their method to a scenario to show its validity. The results show that for seven of the nine objectives, a small set (less than nine) of the 34 metrics are identified that can determine how effectively the objective will be fulfilled. This is the proper subset. Grier et al. state that the "key contribution of our approach is introducing factor analysis as a method for mathematically mapping multiple outputs...to a set of combat indices that capture these objectives" (pages 65-6).

Comments:

> This paper is from a specialized world of warfare simulation. Not knowing what "follow-up fractional factorial central composite designs (CCD)" are, for example, nor "varimax rotation," nor even "Response Surface Methodology (RSM)," I have little basis for comment except that this is a thread into the capability to relate metrics to objectives. But where do we go from here?

Leads to: (none).

2.8 MORS Study Plan

This document is entitled, "Advancing C4ISR Assessment Workshop" [34]. The focus of the document is the design of an approach to answer the following question: What is the value of C4ISR relative to weapons, platforms, and maneuver forces? This question may not be possible to answer, the study argues, without including as a parameter a given decision-maker, since the value of C4ISR is dependent on the decision-maker.

The MoM here is "mission success," as it has always been. But Loss Exchange Ratio can no longer be the focal parameter to that MoM, the document claims. The document lists eight parameters in the following two groups: "mission success" and "success in battle." MoEs provide input to the MoM, and the document lists 29 of them in the following four categories:

1. battle effectiveness,
2. weapon effectiveness,
3. ISR effectiveness, and
4. command and control effectiveness.

Loss Exchange Ratio appears within these categories.

And finally the document lists 10 MoPs. This list, the study notes, is "non-exhaustive." The study does not indicate how the MoPs would be used by the MoEs to provide a MoM.

Comments:

> This is a short study (in number of pages) and unattributed as well. However, it provides succinct measures that might be able to serve as a starting point.

Leads to: (none).

2.9 Hicks et al.

Hicks et al. [20] describe an approach to a "tactical decision" tool that combines differential game theory and classic Bayesian probability theory. This unification of techniques is rare, according to Hicks et al. The authors plan to use a "COTS based decision support architecture" (page 13) known as the "Agent Enhanced Decision Guide Environment (AEDGE™)" to implement a tool that they call a "Submarine Tactical Decision Aid."

Hicks et al. discuss Differential Games as a conflict model from applied mathematics. Two-player games typically involve a "pursuer" and an "evader." The archetypal example is the Homicidal Chauffeur: a person in a car tries to injure a pedestrian; the car is faster but the pedestrian can maneuver better. Unfortunately, as Hick et al. argue, the "complexity of real world situations" typically overwhelms such games: the situations are too big, too complex, and the margin of error in the parameters is too large. One way to address this problem is to constrain the situation. Hicks et al. argue that a better way is to use Bayesian probabilities, which provides a reasoning approach as well (i.e., Bayesian reasoning).

Comments:

> This is a description, not an apology. I presume that a next step, for a subsequent paper, is to describe the implementation of the Submarine Tactical Decision Aid tool and then to provide the results of a study showing the strengths and weaknesses of this tool, as evidence of the value of Hicks et al.'s approach. But again, how does this get us closer to MoEs?

Leads to: (none).

2.10 Freeman & Serfaty

Freeman & Serfaty [15] argue that collaboration may be more cost-effective than hardware. But collaboration is a "very subtle phenomenon" and thus it is difficult to determine its cost-effectiveness. Freeman & Serfaty combine three elements to form a model (and associated measures) of "team collaboration in critical thinking" (TC2T) (page 2):

1. a theory of "collaborative cognition," which defines collaboration primitives;
2. a theory of "individual critical thinking," which identifies the activities that individuals pursue; and
3. a "model-based measurement of team performance."

The model has three "domains:"

1. physical (actions and effects),
2. information (observations), and
3. cognitive (decisions and understanding).

When the understanding part of the cognitive domain is shared between people, they have "common situation understanding and expectations." When the decision part is shared, they share plans. And when the actions are shared, they have synchronization.

The above, and a number of unnamed studies and projects, form the "foundation" of a "measurement system" for Collaboration Technology Measures of Performance, TC2T MoPs, C2 MoPs, and MoEs for "battlefield events, team structure, and team process" (page 4).

This is an "initial" framework.

Comments:

> It is not clear how this framework helps us. It would be instructive to see an application. What are the other forms of collaboration, besides "team critical thinking?" How does the military model of a commander relate to a "team?"

Leads to: (none).

2.11 Jackson et al.

Jackson et al. [21], like Doyle et al. ([14], see Section 2.5), provide an example of the use of VFT. But, unlike Doyle et al.'s, the use of VFT is easier to see in this paper.

Jackson et al. describe an Air Force study that used "operational analysis" to determine the "ideas and concepts" that the U.S. will require in order to have the "dominant air and space forces in the future" (page vi).

The team considered nine analysis techniques. They chose to use VFT instead of any of the following (page 5):

- "Most-to-least dear" with no criteria,
- Qualitative comparison with criteria,
- Simple quantitative comparison matrix,
- Analytical hierarchy process,
- Strategy-to-task,
- Future-to-strategy-to-task,
- Common operational objectives of the armed forces, and
- Cost and operational effectiveness analysis.

Jackson et al. explain VFT but none of the others. (They do not even provide references.)

Based on the description in this paper, VFT is a way to determine the relative value of two systems, represented as a single number for each system. The higher the number, the better the system. VFT directs the building of a model and provides the corresponding method to arrive at that one, final number.

VFT, according to Jackson et al., consists of three pieces:

1. values,
2. a value model, and

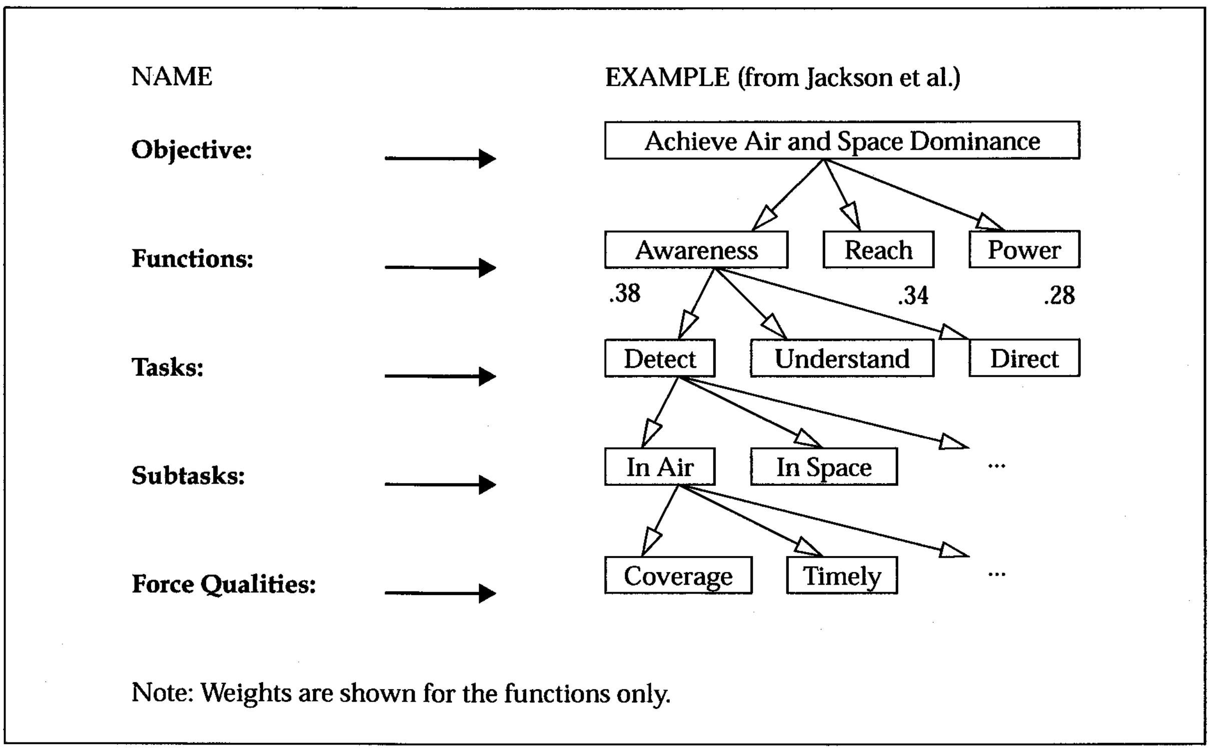
3. weights.

Values are where we begin. They are "what we care about" (which is the definition in Keeney [25], I learned later). A value model consists of seven components:

1. objectives,
2. functions,
3. tasks,
4. subtasks,
5. force qualities,
6. measures of merit, and
7. scoring functions.

The first five of these components can be represented as a tree, where the root node contains the objectives, and the force qualities are the leaves, as exemplified in Figure 3.

Figure 3. Five Components of a VFT Value Model (One set of weights is also shown)



The first four levels represent a grouping consisting of the root node and a functional decomposition of it—what the system does or where or how it does it. However, the fifth level, "Force Qualities," is descriptive, consisting generally of adjectives that characterize the qualities that are of importance to each subtask. That is, a force quality is an attribute of a system that achieves a given subtask. To use Jackson et al.'s value model (see Figure 3), for the subtask "In

Air," two force qualities could be "coverage" and "timely." A force quality should be "measurable, operational, and understandable."

MoMs gauge (i.e., calibrate) system performance. Each MoM translates a force quality for a given system into a number. The MoM is the variable represented by the horizontal axis of a graph of the associated force quality for a given system. Continuing with my example, if the force quality is "coverage," then the MoM could be "percentage of coverage."[3]

Scoring functions dictate how to generate a number for a pair consisting of a given system and a given MoM. There is one scoring function for each MoM. Continuing with my example, if the MoM is "percentage of coverage," then the scoring function could be, "use the results of two test runs at the test facility."[4] It appears that any two MoMs use two, different scoring functions, although there seems to be no proscription against them using the same functions.

Weights are numbers in the range {0..1} that reflect the relative importance of functions, tasks, subtasks, force qualities, and MoMs. The weights for the nodes at the same level in the same subtree sum to unity. Note that weights are not included in a value model. Reason suggests that they are not included in order to facilitate discussion of parameter studies using a given value model. The following is how Jackson et al. use the weights: they develop what I call a "vanilla" value model; then they develop six scenarios; and then they create six value models by starting each time with the same vanilla value model and assigning weights based on a given scenario.

Jackson et al. note that their objective, in the VFT meaning of that term, was already given to them: "achieve air and space dominance" (see Figure 3). This is the root node of the value model. They then developed the next level down, the functions: "awareness, reach, and power." They then enumerated the tasks (eight of these) and subtasks (29 of these), fleshing out the tree. Next they honed a list of possible force qualities—we are not told how many they began with—down to about five per subtask (134 in all). They show all of these top, five levels in Appendix A of their paper. They also show the weights for each of their six scenarios. Unfortunately they describe neither the MoMs nor scoring functions that they used.

Comments:

> I would like to have seen example MoMs and scoring functions. These are important pieces. For them to be absent leaves a gap in the paper.
>
> Why were the alternative analysis approaches not chosen? VFT represented the "best compromise," Jackson et al. say, but best compromise of what? Jackson et al. note that VFT is "well understood and accepted" (page 5) by "the Air University senior leadership" who were involved with the paper; was this the reason VFT was chosen? Were any of the other approaches similarly well understood or accepted?

Leads to: Clemens [10], and OATR [37].

---

3. This is my guess at a measure of merit for the force quality "coverage" and does not come from Jackson et al.
4. As with my example measure of merit, this is my own guess at a scoring function and does not come from Jackson et al.

2.12 Kerchner & Deckro

Kerchner & Deckro present the results of applying VFT to Psychological Operations (PSYOP) [27]. They state that theirs is a "clear demonstration of the potential for using Value-Focused Thinking and multi-objective decision analysis tools for evaluating PSYOP products" (page 20). According to the authors this was the "first time ever" that a "quantitative evaluation and analysis" of PSYOP has been done.

The authors applied VFT in order to generate a value model. Their approach appears to be more informal than that of Jackson et al. ([21], see Section 2.11). For example, the "task" level of their value model does not contain tasks and the "subtask" level seems to contain "force qualities." But the general approach is clear, though, as with Jackson et. al., the particulars for each MoM and scoring functions were glossed over.

Kerchner & Deckro also developed a three-level, "cost" hierarchy that identifies the resources involved. According to the authors, this hierarchy was considered by "several experts" to be "a key contribution" of the paper because "In practice, PSYOP planners spend a great deal of time on the potential benefits of their products and not the drawbacks of those products" (page 9).

The authors used the value model (and the weights) to evaluate five, illustrative PSYOP products in a hypothetical setting, and then they used their cost hierarchy to evaluate the resource consumption of each product, using what they call the "Kirkwood Additive Value Function," which is the sum for each "single dimensional value function for each measure" (i.e., scoring function, in VFT parlance, I believe) multiplied by the weight for that measure (page 9).

They also used what they call "global constraints" (i.e., thresholds). A measure for a PSYOP "product" must be within the global constraints for that measure in order for the product to be considered for use. These thresholds represent "value jumps" (page 21). Below these thresholds, the value of the measure would be zero. For measures with global constraints the authors used a multiplicative, as opposed to an additive, value function, so that the resulting value is zero if any of the global constraints are not met.

This approach helped in two ways. First, it helped to "screen choices prior to development" (page 21) and "as a final check before the PSYOP products are released." Second, it helped as a "tool to help change the PSYOP operator's paradigm" from evaluation of "a single product for a single objective" to considering many objectives for each product. Another prominent benefit from the use of the model, it appears, is that it helped PSYOP planners consider "why a certain product should be used, as opposed to why it was developed" (page 14).

Comments:

> This study shows the value of a quantitative evaluation in general and of VFT in particular. The intent of the authors did not appear to be to provide enough detail for the reader to compute values themselves but rather to demonstrate that a quantitative evaluation can provide more than enough benefit to compensate for the resources required for its application. In this the authors succeeded.
>
> However, it is not clear how the authors determined the value of this approach. I would like some outside confirmation. For example, it would be helpful to see the results of an exercise

with a set of PSYOP planners who used this approach (the experimental group) and a set that did not (the control group).

Leads to: Keeney [25].

2.13 Brown

Brown's thesis [8] is an argument that

1. system effectiveness should be evaluated not only at the "battle outcome" level but also at the "mission" level, and
2. system effectiveness should be gauged using a "cycle" of attributes (availability, reliability, survivability, and capability), based on Marshall's model (see below).

The value of Brown's thesis, as far as I am concerned, is his review of multiplicative and additive models in computing system effectiveness. The multiplicative models equate system effectiveness with the product of several variables, the addition model—Brown presents only one—equates system effectiveness with the sum. I review the multiplicative models first, and then the one additive model, including in parentheses the date of the model. I present the models in reverse chronological order.

The Ball model (1995) uses four variables:

$$MOMS = MAM * (1 - (Ph * Pkh))$$

where	MOMS	is the "measure of mission success,"
	MAM	is the "mission attainment measure" of "offensive capability" in the range {0..1},
	Ph	is the probability of (getting) hit (i.e., susceptibility),
	Pkh	is the probability of being killed if hit (i.e., vulnerability).

The Marshall model (1995) uses four variables—availability, reliability, survivability, and "mission attainment measure:"

$$SE = A * R * S * MAM$$

The Habayeb model (1987) uses three variables:

$$Pse = Psr * Pr * Pda$$

where	Pse	is the probability of system effectiveness,
	Psr	is the probability that the system is ready,
	Pr	is the probability that the system is reliable,
	Pda	is the probability that the system will complete its task.

Habayeb provides a hierarchy that defines each of the three top level attributes in terms of others, such as transportability, supportability, and so on. Brown points out the importance of such hierarchies: they should provide the link between "system attribute metrics" and system effectiveness.

The OPNAVINST model (1987) uses three variables—capability, availability, and dependability:

$$SE = C * A * D$$

The Giordano model (1966) uses just two variables—capability, and performance:

$$SE = C * P$$

Performance, in this model, includes a factor of time.

The model presented in the Weapons System Effectiveness Industry Advisory Committee (WSEIAC) (1965) uses three variables:

$$SE = A * D * C$$

where system effectiveness is a scalar, A, D, and C are 1xN, NxN, and Nx1 matrices, respectively. The system is measured along N independent states. A[i] appears to be the condition of the system relative to state i when the system is called on to fulfill its purpose. D[i][j] appears to be the probability that the system will complete requirements j for state i. And C[i] appears to be the probability that the system can achieve its objectives if it fulfills its requirements for state i.

The additive model, which Brown refers to as "ASDL" (for Aerospace Systems Division Laboratory) (1995), uses five variables

affordability ("life cycle cost"),
capability ("mission capability index"),
safety ("engine-caused attrition"),
survivability ("mission capability index"), and
readiness ("inherent availability") relative to a "baseline system:"

$$OEC = a(LCC/LCC_{bl}) + b(MCI/MCI_{bl}) + c(EAI/EAL_{bl}) + d(Psurv/Psurv_{bl}) + e(Ai/Ai_{bl})$$

where	OEC	is the "overall evaluation criterion,"
	a,...,e	are weights which sum to 1,
	LCC	is life cycle cost,
	LCC_{bl}	is life cycle cost for the baseline system,
	MCI	is mission capability index,
	MCI_{bl}	is mission capability index for the baseline system,
	EAI	is engine-caused attrition index,
	EAI_{bl}	is engine-caused attrition index for the baseline system,
	Psurv	is a measure of survivability,
	$Psurv_{bl}$	is a measure of survivability for the baseline system,
	Ai	is inherent availability,
	Ai_{bl}	is inherent availability for the baseline system.

Brown lists 21 of what he believes to be the "most commonly" used attributes of systems (page 4). However, in the models Brown presents, only a few of these attributes are used.

Comments:

> Brown presents the models without support. Are there *a priori* reasons to choose one over another? What is the empirical evidence? Has some such study not already been done?
>
> None of the models Brown presents sound familiar. Why are they not mentioned in the other papers considered here?

Leads to: (none).

2.14 Keeney

Keeney's book [25] on value-focused thinking (VFT) is the seminal resource for VFT. Its 13 Chapters are partitioned into four "parts," named concepts, foundations, uses, and applications. I review each part below.

Part 1: Concepts

In this part Keeney starts by saying, "Values are what we care about." Values are the "principles used for evaluation" (page 6). He defines VFT as "first deciding what you want and then figuring out how to get it" (page 4):

1. articulate what you want (these are "fundamental objectives"), why you want them, and how the objectives relate to each other;
2. then articulate how to fulfill those objectives (these are "means-ends objectives").

VFT is designed for situations in which there is a real, important, complex decision to be made, with no clear solution at hand (page 22).

VFT is contrasted with a host of methodologies, all of which Keeney places in a category that he calls "alternatives-focused thinking," which I will abbreviate here, for ease of reference, to AFT. In AFT, you choose alternatives presented to you; in VFT, on the other hand, you first augment the list of alternatives, then you create a structure that assists you in the process of deciding one of those alternatives.

Each VFT decision is made within a "decision frame," which consists of a "decision context" (such as "all investment alternatives") and a fundamental objective (such as "maximize financial return of the investment"). A particular decision requires developing a proper subset of the decision frame by developing a proper subset of both the decision context and fundamental objective (such a subset might be "invest in stocks" and "maximize financial return for retirement," respectively).

Part 2: Foundations

In this part Keeney presents how to structure objectives, measure them, and quantify them with a "value model." The most important question to ask is,

> What would I like to achieve in this situation?

This should drive everything else; it is another way of stating the fundamental objectives. The first task is to identify possible fundamental objectives; the next task is to structure them and generate means-end objectives. The fundamental objectives are decomposed into a "fundamental objectives hierarchy," where a hierarchy is a tree.[5] The decomposition terminates when "reasonable attributes"[6] can be found for all of the leaf nodes. For any interior node, the answer to the question, "Why is this important?," is answered by the subtree below the interior node. This hierarchy is based on values. A second structure is needed. The means-end objectives are organized into a "means-ends objectives network."[7] For any interior node, the answer to the question, "How is this achieved?," is answered by the subtree(s) below the interior node. The objectives network is based on facts. The root node of both structures is the same fundamental objective.

These two, qualitative structures may be sufficient. But quantifying them by using a "value model" should be considered: "Quantifying clarifies the meaning of the objectives, and this clarity uncovers hidden objectives and facilitates all aspects of decisionmaking" (page 129). The measurement is done via "attributes"—such as "annual profit in millions of dollars"—which Keeney says some people call MoEs, MoPs, or criteria.

There are three kinds of attributes: natural, constructed, and proxy. Natural attributes are best, followed by constructed attributes; proxy attributes are the worst because they "confound" facts and values (page 119). However, it may be the case that the only type that is available is proxy attributes. Sometimes the latter two types of attributes can be avoided by decomposing the associated objective, but decomposition increases the demand for data collection. I explain each type of attribute below.

Natural attributes are ones that have a "common interpretation to everyone" (page 101), such as annual profit in millions of dollars.

Constructed attributes, which some people call "subjective scale" or "subjective index," have value judgements built-in. An example constructed attribute for the biological impact of a given power plant is a scale from zero to 8 associated with defined levels of loss of farmland or forest or habitat. This example constructed attribute is defined for the given power plant site only. Constructed attributes can, over time, become natural attributes, e.g., the DJIA, the GNP, and the Richter scale.

A proxy attribute exploits a perceived relationship. For example, ambient sulphur dioxide concentration could be a proxy attribute to measure the disfiguration of stone statues due to acid rain. Attributes should be measurable, operational (meaning the attribute describes consequences and maps scores to desirability), and understandable.

A value model is also known as an "objective function." It consists of a set of objectives, a set of attributes that measure those objectives, and a function, v, that maps a set of attribute measurements to a single number. Where uncertainty is involved, v is a "utility function," u. Multiattribute utility theory has developed four "independence conditions," namely "preferential," "weak-difference," "utility," and "additive." The conditions indicate the form of

5. A tree is a directed graph in which every node has exactly one parent node, except the one "root" node which has no parents.

6. Attributes are ways of measuring the fulfillment of objectives, as described below.

7. A network is a directed graph in which every node has at least one parent node, except the one "root" node which has no parents.

u, none of which I will include here.

Keeney notes that the "most common critical mistake" is to improperly gauge priorities between pairs of objectives. That is, saying that objective A is three times as important as objective B may have no meaning if various sample measurements are not investigated. Also, as Keeney puts it, "One of the most useful devices for quantifying values is simply to ask 'why?' questions" (page 150).

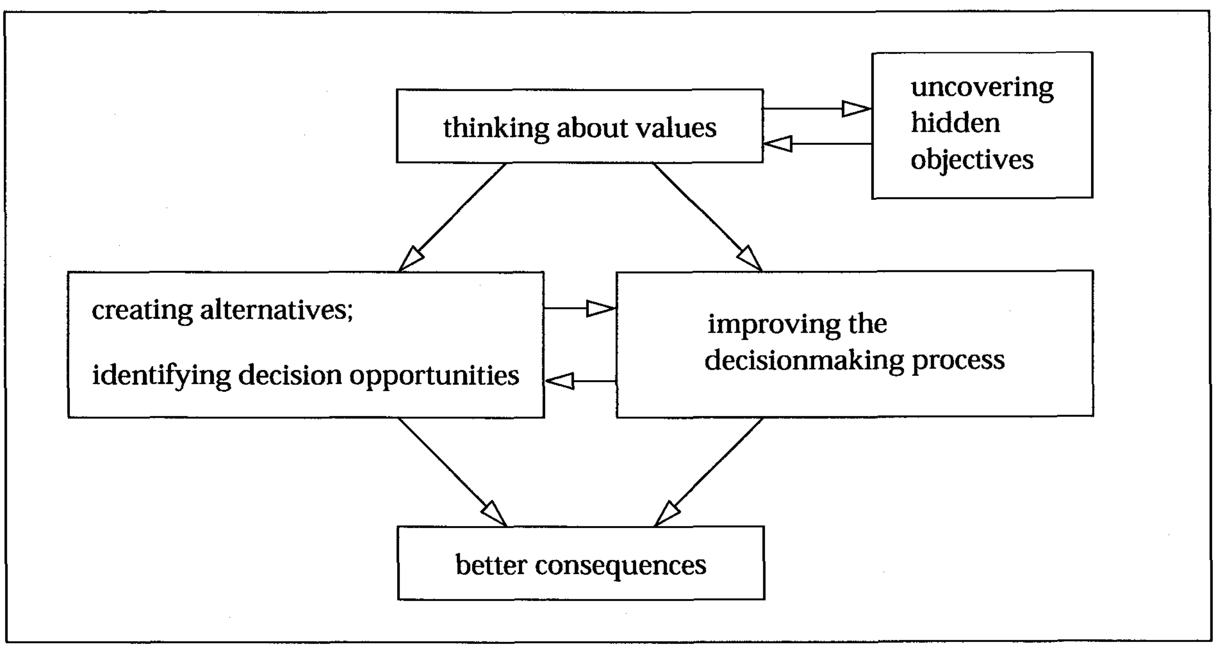
Part 3: Uses

In this part Keeney describes the value of his approach by showing how it can be used. Keeney presents four uses:

1. to "uncover hidden objectives,"
2. to create alternatives,
3. to "identify decision opportunities" (as opposed to decision problems, which are typically forced upon us), and
4. to provide insights.

Keeney presents the relationship between three of these uses in a figure—the most important in the book, in my opinion—reproduced here as Figure 4.

Figure 4. The Influence of VFT on Decisionmaking Processes (Figure 10.1 from [25])



As Keeney puts it,

> Many people think that the only reason to perform a value assessment [i.e., use VFT] is to obtain an objective function that can be used to evaluate alternatives. Used well, however, a

value assessment provides much more. Because values are the reason for interest in a decision, value assessment can provide insights for every phase of the decision process. (page 157)

Keeney notes that, "The qualitative statement of objectives and the quantitative representation of values are complementary in nature and offer complementary insights" (page 157-8). The attributes (quantitative measures) can clarify the meaning of the objectives (qualitative measures). If independence assumptions have been chosen improperly, i.e., the answer provided does not match what the user is expecting for the given objectives for known values, then "either a fundamental objective has been overlooked or means objectives are being used as fundamental objectives" (page 165). VFT helps the user to nudge his own mind to new territory because, Keeney argues, it is in that unexplored territory that unexamined alternatives lie, and it is in amongst those unexamined alternatives that the user can find better alternatives from which to choose. Keeney presents a dozen ways to use a VFT hierarchy and network for this purpose, providing many illustrative case studies along the way.

Part 4: Applications

In this part Keeney describes the benefits of applying VFT by describing examples. In Chapter 11 Keeney presents aspects of six applications. In Chapter 12 he presents, in considerable depth, the first step of an additional application. And in Chapter 13 he presents examples of the use of VFT from his own life.

The six examples in Chapter 11 cover all nine of the uses for VFT from Chapter 1:

1. NASA used VFT to identify and prioritize objectives for missions.
2. The US Department of Energy used VFT to identify and combine fundamental objectives, as well as means objectives, from three different stakeholders on the issue of transporting nuclear waste.
3. A utility company used VFT to develop objectives for research on global climate change.
4. A second utility company used VFT to develop attributes and calculate value tradeoffs from multiple stakeholders concerning air pollution in Los Angeles.
5. A technology company used VFT to compare possible products.
6. And two authors used VFT to explore ways to organize the final revision of a book.

Keeney's interest is in convincing the reader of the range and benefit of VFT.

Chapter 12 recounts the "first step" in the use of VFT at British Columbia Hydro. Keeney helped them develop strategic objectives. This is the sequence Keeney followed with B.C. Hydro:

1. They arrived at six, top-level objectives, the first of which was to "Maximize contribution to economic development."
2. They then developed attributes. The first attribute corresponding to the first, top-level objective was "Minimize cost of electricity use (mills per kilowatt-hour in 1989 dollars)."

3. They then assigned a range for these attributes. The range for the electricity cost attribute was 35 to 55.
4. Next they developed a utility function, with six subsidiary utility functions, one for each of the top-level objectives, followed by scaling constants for those six objectives. The scaling constant for the first, top-level objective was 0.395, indicating that almost 40% of the overall utility function would consist of the measure for the first objective. The scaling constants ranged over two orders of magnitude, with the smallest at 0.004.
5. They then calculated priorities, putting the first attribute at top priority with a weight of 0.253. This enabled them to calculate value tradeoffs, indicating, for example, that "each hectare of wilderness lost is as bad as losing 2,500 of resources [,] and a major electrical outage to 500,000 small customers for 24 hours is as bad as a $6.5 million resource loss (that is, $13 per customer)" (page 368).

Keeney sums up the benefit of this effort at British Columbia Hydro as follows: "An unambiguous and complete statement of strategic objectives can be a guide to identifying decision opportunities that enhance both the likelihood of achieving those objectives and the degree to which the objectives are achieved" (page 370). This Chapter gives the reader an understanding of the process and benefits of applying VFT in a corporate setting.

Chapter 13 is Keeney's review of his use of VFT in his personal life. This has helped him realize decision opportunities, as he calls them. He notes that the effort expended is minimal compared to the reward. VFT enables him to make decisions in a consistent manner, considering consequences explicitly. With his own objectives clear, Keeney can develop alternatives that coincide with others' objectives, even if those objectives are implicit, enabling a win-win situation. This significantly increases Keeney's chances of realizing situations that coincide with his objectives, since one of those "wins" is his. As Keeney puts it, "With the principles established, only the details of the logistics remained. This is the core of value-focused thinking. Understand your objectives first and then figure out how to achieve them" (page 393). And later, "The foundation of value-focused thinking is the idea that objectives should be considered prior to alternatives on important decisions" (page 396).

Book Review

Schuyler [44] provides a favorable review of Keeney's book which he summarizes as follows: "The main message is that substantially more time should be invested in eliciting values and proactively pursuing goals." Schuyler concludes:

> This is an idea book. It is worth the investment solely as an inspiration source for unsticking [sic] decision dilemmas: Clarifying values leads to creating new alternatives. I'm recommending this book for executives and consider it a "must read" for students and MS/OR practitioners concerned that their applications have real-world relevance and value.

The precursor to this book is the book by Keeney & Raiffa [24].

Comments:

After reading Jackson et al.'s [21] description of VFT (or even Kerchner & Deckro's [27] or Doyle et al.'s [14]) I was surprised by Keeney's book. Keeney mentions almost nothing of the structure that Jackson uses. There is no talk about a value model consisting of functions, tasks, subtasks, and force qualities, for example. I was expecting significant regimentation. But Keeney presents what I suppose one would call an "approach" rather than a "model." The approach consists of a "fundamental objectives hierarchy" and a "means-ends objectives network." These are almost freewheeling when compared to Jackson's description.

I appreciated Keeney's clarity on the importance of subjectivity. Values by their nature are subjective. Pretending that they are objective is, I believe, a death knell. So instead of trying to pretend that this subjectivity can be ignored or swept away with enough formulas, Keeney's approach concentrates on making the values explicit and thus provides one of the benefits of objectivity, namely, the opportunity to get "outside" the problem and view it from there. But the values are all still subjective. I believe that this is a sound approach.

However, is "alternative-focused thinking" the only alternative to VFT? Have there been no other decision-making models developed, before VFT, other than alternative-focused thinking? What about the eight that Jackson et al. list? I would like to see a brief history, to see the other models, to see what is done in other cultures, in other places, in other times. This would enable me to see VFT's place and better understand its significance.

I would also like to see a review of what others have done with VFT. What is the genealogy of papers that can be traced to Keeney's book or just to the approach? Is there a consulting group that provides VFT? Is there a VFT advocacy society? Have there been any attempts to standardize it, so, for example, someone could say that they are a "certified" VFT consultant?

Granted, these are questions and additions that require time and the diffusion of innovation. However, there has been no second edition of Keeney's book—the first was in 1992—nor, that I can see, a paper that would provide this kind of information, entitled "VFT at 10 Years," say. So I have no second opinion about Keeney's offering, except Schuyler's favorable book review, which I summarized above. I have no reason to doubt that the approach is valuable—it certainly appears that way to me—but I have no gauge beyond my own conclusions to determine how valuable VFT is.

I would have liked to have seen an Appendix in Keeney's book, with a summary of the approach, a step-by-step guide, something that I could point to that would provide the structure of VFT. I found Keeney's description of "operationality" (page 114) opaque, but I am not satisfied with my own précis of it. Section 4.7, about connecting probability and timing with attributes, was confusing. I do not understand the significant of the "risk attitudes," described in Section 5.3. I covered Chapters 7-10 lightly; these techniques on the use of VFT that are not easily summarized.

Leads to: Keeney [24].

2.15 Glenn

Glenn [16] defines the MoE for a given mission to be the sum of the product of the MoP and the "utility factor" for each function "related to the system's requirements." Both the MoP and utility factor are expressed as a number in the range {0.1..1}. The utility factor quantifies the value of the function to achieving the mission at hand ("the link between the system's MOP and how the system's [electronic warfare] capability offsets the outcome of a battle" (page 3)). Glenn also provides a "confidence weight factor" (page 1), which is again a number in the range {0.1..1}, for each function. This approach, Glenn argues, is superior to the "commonly used qualitative measures of high-medium-low" (page 1).

Glenn provides an example of MoE computation for a "ground control intercept voice anti-jam communication system." The example lends itself to an objective computation of the MoP. The confidence and utility factors are drawn somewhat from the air.

Comments:

> This is a simple approach and that may be its strength. However, the lack of a hierarchy—such as Keeney's objectives hierarchy—to help choose confidence and utility factors means that the approach is unstable, I believe. At their extreme values the confidence and utility factors can spread the same MoP by two orders of magnitude and thus can become the dominant factors. Unfortunately Glenn's approach does not seem to have been picked up by anyone else for no one cites this work (and it was published long ago, in 1981).

Leads to: (none).

2.16 Sproles

Green & Johnson [17] reference three papers by Sproles ([48], [50], [46]), all of which are reviewed in this Section. I requested Sproles' dissertation as well, but Sproles pointed me to these papers instead as they better represent his current understanding and viewpoint. These three papers provide a progression of thought that narrows in on the meaning of MoEs. I present the papers in chronological order of publication date.

Sproles [46]

Sproles [46] argues that "measures" are the standards by which we determine the success of a system. Measures are the "show stoppers." Sproles is concerned that there is no "universal definition" of measures. To make matters worse, he notes that humans are "not quite as rational as one would like to believe" (page 98). We are subject to the availability heuristic, the conjunction fallacy, and framing, to name a few. In light of these irrationalities, how should measures be chosen so that effect of these problems is neutralized? Obviously this is an open question. To make matters worse, Sproles argues that much of the universe is "non-linear" and thus "inherently resistant to quantification" (page 99). How can we measure (i.e., make into quantities) that which cannot be measured? In conclusion, Sproles states, "The fact is that there is no way to distinguish if a measure is good, let alone how good it is, although it is possible to detect a poor measure" (page 100).

As examples of measures, Sproles presents a US Air Force evaluation scale, "Q-factor," from 0 to 10. (Keeney [25] would call this a "constructed attribute.") Sproles also presents example proxy measures (which Keeney would call "proxy attributes"): a measure for ambulance service, and a second measure for taxi service in Adelaide.

Comments:

> I believe that the question that Sproles is honing in on is this: How do we measure our measures? If we cannot gauge a measure, how can we have faith in it? If my measure says the system is good and yours says it is bad, upon what basis do we proceed to a resolution?
>
> Meanwhile, measure developers, being human, are also irrational. Perhaps their irrationality cancels out the irrationality of the decision-makers? It appears highly unlikely but the question requires addressing. (But then, advertisers seem to have learned how to cancel out consumers' irrationality.)

Leads to: Burge [9], Pinker [40], Roche [41], and Saaty [43].

Sproles [48]

The explicit focus of this second paper by Sproles [48] is MoEs. There is no "universally accepted meaning for them," Sproles states. But Sproles proposes a definition:

> Measures of Effectiveness are standards against which the capability of a solution to meet the needs of a problem may be judged. The standards are specific properties which any potential solution must exhibit to some extent. MOEs are independent of any solution and specify neither performance nor criteria. (page 54)

Implicit in the definition is that MoEs are based on a "stakeholder." The stakeholder is in the "preeminent position" (page 50). An MoE is "the standard by which stakeholders[8] establish if a solution will meet the need" (page 50). MoEs are a "management tool" (page 57).

MoEs are formulated by the stakeholder asking himself the following question, relative to a given function, "'If a candidate solution cannot do this, would I reject it?' If the answer is 'yes,' then it is a crucial issue and the basis for an MoE" (page 55). Note that this is a "cognitive process" (i.e., in the mind), and it is performed by the stakeholder. MoEs are the "show stoppers." They will probably be small in number.

Test and Evaluation (T&E) is the process by which a solution is judged based on the MoEs. "The MoEs identify the properties to be tested; Criteria identify the limits within which a successful solution will lay" (page 56). Sproles contrasts MoEs with MoPs. Not surprisingly, there is no "universally accepted meaning" for MoPs. Sproles proposes that an MoP indicate how well a system does something. MoPs are concerned with efficiency, whereas MoEs are concerned with effectiveness, as their name states. MoPs focus on the "inner" environment, the world of the system itself; MoEs, on the other hand, focus on the "outer" environment, the value of the system to the world beyond, as summarized in Table 1.

Comments:

8. Sproles in the next paper, [50], argues that there is always one stakeholder whose opinion is more important than the others, so we can speak of one stakeholder.

Table 1 MoPs vs. MoEs.

Item	MoP	MoE
Viewer	Developer	Stakeholder
Measure	Efficiency	Effectiveness
Viewpoint	Inner	Outer

Sproles' definitions provide a satisfying way to view both MoEs and MoPs. And he provides a satisfying way to develop them. He succeeds, as the title of the paper suggests, in "coming to grips" with MoEs.

Note that Sproles does not address the issue of how well a set of MoEs identify what the stakeholder has in mind. This issue was in the background of Sproles' previous paper, reviewed above. I presume that Sproles presumes that this is up to the stakeholder: if he cannot perceive that he is asking for something other than what he wants, then this is his own problem. The buck has to stop somewhere.

Leads to: Leibowitz [28], Lewis [29], and Naumann [36].

Sproles [50]

An MoE is a "statement," as Sproles explains in this paper [50]. An MoE is generated by the stakeholder, the "one group whose need defines the function of any proposed solution" (page 146). An MoE identifies "what is wanted rather than what must be done" (page 147). An MoE is thus a "tool designed to help establish if a system, when developed as a solution to a problem, performs the mission required of it by a stakeholder" (page 146). An MoE represents the "Critical Operational Issues" (COI) or "show stoppers" (page 147). To qualify, an MoE must be sufficient to determine if a given system meets the stakeholder's requirements. Quantitative objectivity is preferred but qualitative subjectivity is permissible. An MoP, on the other hand, is a gauge of the system from the developer's viewpoint.

Sproles shows that we do not yet have MoEs for Command & Control (C2) systems. Like an ingredient in a cake, a C2 system cannot be isolated. As a result C2's contribution to the larger system cannot be determined, except in extreme situations, such as the "withdrawal from the Gallipoli Peninsula in December 1915" and the "1st British Airborne Division operations at Arnhem in 1944" (page 153). In addition, it is infeasible to "replace" C2 systems to test them. And, unlike a tank battalion, a C2 system has no existence without a commander. In this case our tools are insufficient. Perhaps "behavioral or soft sciences" can help. Perhaps "discursive means" will enable us to determine the contribution made by C2 systems.

Comments:

Sproles has arrived at a clear and satisfying definition of MoEs, how they are developed and the context within which they have meaning. Showing how we are currently unable to develop MoEs for C2 systems highlights his definition: we see what MoEs are not.

It is up to the stakeholders to develop MoEs. Since a set of MoEs is unique for each unique system and since the gauge of a set of MoEs can only be determined by the stakeholders themselves, we cannot develop an independent gauge of MoEs. But then, we no longer need to: it is up to the stakeholders.

Leads to: Jackson & Keys [22], Sproles [47], Sproles [49], and Stevens [51].

3 Tangential Work

In this Section I briefly review the following items (listed in the order that I encountered them):

1. McGuinness & Ebbage [31],
2. Curts & Campbell [11],
3. Bordley [7],
4. Ackoff [1],
5. Scott [45], and
6. Alberts [4].

In addition to the material reviewed in Section 2, I read other material that turned out to be only tangential to the matter at hand. I have included short reviews below of this work. This material is included so that the reader can see what I view as tangential, i.e., where the matter at hand stops. This collection of material defines at least part of the boundary.

McGuinness & Ebbage [31] want to determine the effects of digitization on command and control. They conclude, from experiments in a "synthetic environment," that users want digitization to supplement voice radio communication, not replace it. They present 10 findings, such as digitization reduces a considerable amount of traffic information because digitization provides accurate location information.

Curts & Campbell [11] are concerned with controlling a system, which they might argue is a way of measuring it. They argue that systems engineering provides the processes, methodology, and tools necessary for this task.

Bordley [7] argues that the "dialogue decision process" (DDP), which Bordley says is a form of VFT, is "theoretically equivalent" to Ackoff's "ideal-focused decision process." This paper is more helpful, from my viewpoint, in pointing out the existence of DPP and Ackoff's approach than in arguing for their equivalence.

Ackoff [1] presents a systematic study of systems and provides a taxonomy of them as well. Ackoff makes 33 observations about systems, such as what it means to say a system is "abstract," or what its "state" is, or the general nature of "open" and "closed" systems. Ackoff's is a "holistic" point of view, since a system only makes sense when the relationship between the parts is considered.

Scott [45] was disappointing. Allen & Demchak's [6] summary of organizations was intriguing, so I was hoping that Scott might be able to shed light on how organization structure could serve to identify MoEs. But Scott's book provided little of the sort.

Alberts et al.'s book, "Understanding Information Age Warfare," [4] is an argument for "Information Superiority" and "Network Centric Warfare." The authors propose for measures or metrics "richness, reach, command and control (C2), and value" (page 31). Figure 16 of the book presents what appears to be two lower levels of the hierarchy. Unfortunately there is no approach given on how to quantify those lower level measures nor on how to combine them into the four at the top level. They discuss collaboration and note measures (page 200) but do not indicate how to arrive at values for those measures. Synchronization is another important topic. The authors suggest a measure for synchronization (page 232), but they themselves note it is not "a particularly useful measure" (page 233). Chapter 10 presents the general results of half a dozen studies military exercises that support their main argument in favor of the value of network centric warfare, but the exercises are not measures, though they generate measurements.

4 Conclusions and Future Work

Figure 1 shows two veins:

1. the literature search on VFT, leading to Keeney [25], and
2. Green & Johnson's paper, leading to Sproles' papers ([46], [48], [50]) and beyond.

Sproles provides us with a solid approach to MoEs—what they are, how they are used, how they are developed, how they are gauged, and how they differ from MoPs. Keeney's book [25] on VFT provides us with a way to bridge the gap between MoEs and a system, along with a way to decompose MoEs into a hierarchy that helps us determine the meaning of our MoEs—Keeney would call them values—and how to measure them, and, finally, how to quantify those MoEs for to a given system.

Meanwhile Jackson et al. [21] provide the most explicit example, though incomplete, of using VFT. Their use of VFT seems to be significantly more regimented than it needs to be. This may be personal choice. It would be helpful to become familiar with the eight analysis techniques that Jackson et al. rejected in favor of VFT. (It would also be interesting to follow up on the "dialogue decision process" and Ackoff's "ideal-focused decision process" that Bordley notes.) Clemens' book [10] may provide more needed detail on the model that Jackson et al. used. The technical report [37] of their analysis should provide much needed details. Parnell [38] should help as well. Kercherner & Deckro [27] provide a second example of an analysis using VFT. Doyle et al. [14] is a third analysis using VFT. However, it is more difficult to see the VFT approach underlying this last analysis than it is with either of the other two. Nevertheless, although Doyle et al. are confusing in their terminology, VFT is still visible underneath.

The other papers I reviewed are only of secondary interest. Generally they do not provide sufficient purchase to compete with the ones noted above.

All of the "to be reviewed" items in Figure 1 constitute the frontier to pursue to take this study to the next level of detail. These items are listed alphabetically below:

1. Ackoff 1978 [2],
2. Alberts 1996 [3],
3. Alberts 2002 [5],
4. Burge [9],
5. Darilek [12],
6. Jackson & Keys [22],
7. Jurk [23],
8. Keeney 2001 [26],
9. Keeney & Raiffa [24],
10. Leibowitz [28],
11. Lewis [29],

12. Naumann [36],
13. Parnell [38] (which should shed light on Jackson et al. [21]),
14. Pinker [40],
15. Roche [41],
16. Saaty [43],
17. Sproles [47],
18. Sproles [49], and last but not least,
19. Stevens [51].

Glossary

DP	Dimensional Parameters
MCES	Modular Command and Control Evaluation Structure [17]
MoE	Measure of Effectiveness (sometimes MOE)
MoP	Measure of Performance (sometimes MOP)
MoFE	Measure of Force Effectiveness (sometimes MOFE)
MoM	Measure of Merit (sometimes MOM)
MORS	Military Operations Research Society
PSYOP	Psychological Operations
VFT	Value-Focused Thinking

References

[1] Russell L. Ackoff, "Towards a System of System Concept." Management Science, Vol. 17, No. 11, July 1971, pp. 661-71. 3 refs. (Reference 1 of [17].)

[2] Russell L. Ackoff, The Art of Problem Solving: accompanied by Ackoff's Fables. John Wiley & Sons, New York, NY, 1978.

[3] David S. Alberts, Information Age Transformation: Getting to a 21st Century Military. CCRP Publication Series. 145 pages. 1996.

[4] David S. Alberts, John J. Garstka, Richard E. Hayes, David A. Signori, Understanding Information Age Warfare. CCRP Publication Series. 312 pages. 2001.

[5] David S. Alberts, Richard E. Hayes, Code of Best Practice for Experimentation. CCRP Publication Series. 434pages. 2002.

[6] Patrick Allen, Chris Demchak, "An IO Conceptual Model and Application Framework." Military Operations Research, V6 N2 2001, pp. 9-22. 22 refs.

[7] R. Bordley, "Relating value-focused thinking and interactive planning." Journal of the Operational Research Society (2001) 52, 1315-1326. 53 refs.

[8] Kevin W. Brown, "Measuring the Effectiveness of Weapons Systems In Terms of System Attributes." MS Thesis. Naval Postgraduate School. December 1995. 68 pages. 19 refs.

[9] A. R. Burge, "Test and Evaluation Based on Metrics, Measures, Thresholds, and Indicators," The ITEA Journal of Test and Evaluation, September/October 1996, Vol. 17, No. 3, pp. 47-49.

[10] R. T. Clemens, Making Hard Decisions: An Introduction to Decision Analysis. Boston, Mass. PWS-Kent. 1991.

[11] Raymond J. Curts, Douglas E. Campbell, "A Systems Engineering Approach to Information Assurance operations." CCRTS. Monterey, CA. June 11-13, 2002. 16 pages. 22 refs. http://www.dodccrp.org/Activities/Symposia/2002CCRTS/Proceedings/Tracks/Track_7.htm.

[12] Richard Darilek, Walter Perry, Jerome Bracken, John Gordon, Brian Nichiporuk, "Measures of Effectiveness for the Information-Age Army." 116 pages. 36 refs. ISBN: 0-8330-2847-2. 2001. (Previously appeared as RAND Corporation report DRR-1983-A, 1999.)

[13] Deckro: Literature search on the author "Deckro." September 9, 2002. 52 citations.

[14] Michael P. Doyle, Richard F. Deckro, Jack M. Kloeber, Jr., Jack A Jackson, "Measures of Merit for Offensive Information Operation Courses of Action." Military Operations Research, V5 N2 2000, pp. 5-18. 20 refs.

[15] Jared T. Freeman, Daniel Serfaty, "Team Collaboration for Command and Control: A Critical Thinking Model." CCRTS. Monterey, CA. June 11-13, 2002. 7 pages. 19 refs. http://www.dodccrp.org/Activities/Symposia/2002CCRTS/Proceedings/Tracks/Track_4.htm.

[16] A. B. Glenn, "Electronic Warfare System Measure of Effectiveness." International Conference on Communications." Denver, CO. June 1981. pp. 36.6.1-5. 2 refs.

[17] John M. Green, Bonnie W. Johnson, "Towards a Theory of Measures of Effectiveness." CCRTS. Monterey, CA. June 11-13, 2002. 14 pages. 44 refs. http://www.dodccrp.org/Activities/Symposia/2002CCRTS/Proceedings/Tracks/Track_5.htm.

[18] James B. Grier, T. Glenn Bailey, Jack A. Jackson, "Response Surface Modeling of Campaign Objectives Using Factor Analysis." Military Operations Research, V4 N2 1999, pp. 61-70. 17 refs.

[19] Kenneth P. Heartling, Richard F. Deckro, Jack A. Jackson, "Implementing Information Warfare in the Weapon Targeting Process." Military Operations Research, V4 N1 1999, pp. 51-65. 13 refs.

[20] Jeffrey D. Hicks, Plamen V. Petrov, Alexander D. Stoyen, Qiuming Zhu, "Intelligent Agent-Based Software Architecture for Tactical Decision Aid under Overwhelming Information Inflow and Uncertainty." CCRTS. Monterey, CA. June 11-13, 2002. 15 pages. 16 refs. http://www.dodccrp.org/Activities/Symposia/2002CCRTS/Proceedings/Tracks/Track_4.htm.

[21] J. A. Jackson, B. L. Jones, L. J. Lehmkuhl, "Operational Analysis for Air Force 2025: An Application of Value-Focused Thinking to Future Air and Space Capabilities." June 1996. 60 pages. 4 refs. http://www.au.af.mil/au/2025/.

[22] M. C. Jackson, P. Keys, "Towards a system of systems methodology." J Oper Res Soc 35(6), 473-486.

[23] D. M. Jurk, "Analysis with Value Focused Thinking as a Methodology to Select Force Protection Initiatives for Evaluation." Masters Thesis. Air Force Institute of Technology, Wright-Patterson AFB, OH. 2002. 243 pages.

[24] Ralph L. Keeney, Howard Raiffa, Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Cambridge University Press. New York, NY. 1993. ISBN 0-521044185-4. 569 pages. First published in 1976 by John Wiley & Sons, Inc.

[25] Ralph L. Keeney, Value-Focused Thinking: A Path to Creative Decisionmaking. Harvard University Press. Cambridge, MA. 1992. ISBN 0-674-93197-1. 416 pages.

[26] Ralph L. Keeney, "Modeling Values for Telecommunications Management." IEEE Trans. on Engineering Management, vol. 48, no. 3, August 2001, pp. 370-9. 20 refs.

[27] Philip M. Kerchner, Richard F. Deckro, "Valuing Psychological Operations." 67th MORS Symposium. June 1999. 26 pages. 21 refs.

[28] M. L. Leibowitz, "Metaphysical considerations involved in choosing a measure of effectiveness." Oper Res J Oper Res Soc Am 6(1) (1988), 127-130.

[29] B. Lewis, "Measure for measure: For better performance results, call it what it is." IS Survival Guide, http://www.infoworld.com/pageone/opinions/lewis/rl021966.htm, February 19, 1996.

[30] John Marshall, "Effectiveness, Suitability, and Performance." Proceedings of the 59th MORSS. (Citation 23 of [17].)

[31] Barry McGuinness, Louise Ebbage, "Assessing Human Factors in Command and Control: Workload and Situational Awareness Metrics." CCRTS. Monterey, CA. June 11-13, 2002. 11 pages. 2 refs. http://www.dodccrp.org/Activities/Symposia/2002CCRTS/Proceedings/Tracks/Track_5.htm.

[32] MoE: Literature search on "measures of effectiveness." September 9, 2002. 39 citations.

[33] J. Moffat, G. L. Mathieson, D. Shirley, "A Guide to Best Practice in C2 Assessment." (citation unknown.) August 1999. 18 pages. 1 ref. http://www.mors.org/meetings/CRISR_2000/NATO_Short_COBF.pdf.

[34] MORS MTW C4ISR Working Group, "Study Plan." October 2000. 11 pages. 0 refs. http://www.mors.org/meetings/CRISR_2000/MTW WG Study Plan Oct 00.pdf.

[35] NATO Code of Best Practice for C2 Assessment, AC/243(panel7)TR/8, 1998 Edition. NATO U/C.

[36] E. Naumann, "Creating customer value: The path to sustainable competitive advantage." International Thomson Publishing, 1995.

[37] OATR: "An Operational Analysis for 2025: An Application of Value-Focused Thinking to Future Air and Space Capabilities." Maxwell AFB, Ala.: Air University, 1996. (2025 Operational Analysis Technical Report.)

[38] G. S. Parnell, H. W. Conley, J. A. Jackson, L. J. Lehmkuhl, J. M. Andrews, "Foundations 2025: A Value Model for Evaluation Future Air and Space Forces." Management Science, vol. 44, no. 10, pp. 1336-50, 1998.

[39] Walter L. Perry, "Knowledge and Combat Outcomes." Military Operations Research, V5 N1 2000, pp. 29-39. 8 refs.

[40] A. Pinker, A. H. Samuel, R. Batcher, "On Measures of Effectiveness," *Phalanx*, The Military Operations Research Society. December 1995, pp. 8-12.

[41] J. G. Roche, B. D. Watts, "Choosing Analytic Measures," The Journal of Strategic Studies, Vol. 14, June 1001, pps. 165-209.

[42] Bernard H. Rudwick, Systems Analysis for Effective Planning: Principles and Cases. John Wiley & Sons. New York, NY. 1969. (Reference 30 of [17].)

[43] T. L. Saaty, The Analytic Hierarchy Process. McGraw Hill, Singapore. 1980.

[44] John R. Schuyler, A book review of Keeney's book, "Value Focused Thinking" [25]. Interfaces, vol. 23, no. 6, 1993, pp. 140-2.

[45] W. Richard Scott, Organizations: Rational, Natural and Open Systems. 1998. 4th edition. Saddle River, New Jersey: Prentice Hall Inc.

[46] Noel Sproles, "Getting the Measures of Test and Evaluation." Journal of Electrical and Electronics Engineering, Australia, v. 17, no. 2, 1997, pp. 95-101. (18 refs.)

[47] Noel Sproles, "Measures of Effectiveness: How will I recognise that I have succeeded?" Proceedings of INCOSE UK Fourth Annual Symposium, Hendon UK, INCOSE UK, June 1-2, 1998, pp. 95-102.

[48] Noel Sproles, "Coming to Grips with Measures of Effectiveness." System Engineering: The Journal of the International Council on Systems Engineering, vol. 3, no. 1, 2000a, pp. 50-8. (30 refs.)

[49] Noel Sproles, "Complex systems, soft science, and test & evaluation," Proc Syst Eng Test Eval Conf, Brisbane, Australia, November 15-17, 2000b, pp. 177-186.

[50] Noel Sproles, "The Difficult Problem of Establishing Measures of Effectiveness for Command and Control: A Systems Engineering Perspective." Systems Engineering, vol. 4, no. 2, 2001, pp. 145-55. (35 refs.)

[51] R. Stevens, P. Brook, K. Jackson, S. Arnold, Systems enginering, coping with complexity, Prentice Hall Europe, Hemel Hemstead, UK, 1998.

[52] VFT: Literature search on "value focused thinking." September 9, 2002. 48 citations.

Distribution

1	0415	S. G. Kaufman, 9741
1	0455	R. D. Pollock, 5501
1	0455	R. S. Tamashiro, 5517
1	0708	J. M. Covan, 6202
1	0784	J. M. DePoy, 5512
1	0784	R. D. Halbgewachs, 5501
1	0784	M. J. Skroch, 5512
1	0784	R. E. Trellue, 5501
1	0784	C. M. Villamarin, 5512
1	0785	P. L. Campbell, 5516
1	0785	J. D. Dillinger, 5516
1	0785	D. L. Harris, 5516
1	0785	R. L. Hutchinson, 5516
1	0785	A. J. Lanzone, 5514
1	0785	T. S. McDonald, 5514
1	0785	M. E. Senglaub, 5516
1	0785	J. E. Stamp, 5516
1	0785	M. D. Torgerson, 5514
1	0785	B. P. Van Leeuwen, 5516
1	0806	L. G. Pierson, 9336
1	1030	W. F. Hossley, 12870
1	1351	J. J. Torres, 5517
1	1371	B. G. Varnado, 4142
1	9018	Central Technical Files, 8945-1
2	0899	Technical Library, 9616

1	0612	Review & Approval Desk, 9612 For DOE/OSTI