Research Product 94-03

Case-Based Expert Systems for Combat Performance Analysis

January 1994

Advanced Training Methods Research Unit
Training Systems Research Division

U.S. Army Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Technical review by

J. Douglas Dressel
Douglas Macpherson

NOTICES

FINAL DISPOSITION: This Research Product may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: This Research Product is not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1994, January	Interim Oct 92 - Mar 93

4. TITLE AND SUBTITLE
Case-Based Expert Systems for Combat Performance Analysis

5. FUNDING NUMBERS
63007A
795
2122
H01

6. AUTHOR(S)
Mirabella, Angelo

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-I
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

8. PERFORMING ORGANIZATION REPORT NUMBER
ARI Research Product 94-03

9. SPONSORING MONITORING AGENCY NAME(S) AND ADDRESS(ES)
--

10. SPONSORING MONITORING AGENCY REPORT NUMBER
--

11. SUPPLEMENTARY NOTES
--

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT *(Maximum 200 words)*

This research was part of a larger program to develop methods for selecting brigade training strategies. An essential step in the program was to define performance measures to assess the impact of alternative strategies. Case-based expert system (CBES) technology can help do so. This research product demonstrates how to use CBES to better understand relationships among process and outcome variables. CBES use is described in a step-by-step application to sample data.

14. SUBJECT TERMS
Unit training
Performance assessment
Collective training

15. NUMBER OF PAGES
39

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

Research Product 94-03

Case-Based Expert Systems for Combat Performance Analysis

Angelo Mirabella
U.S. Army Research Institute

Advanced Training Methods Research Unit
Robert J. Seidel, Chief

Training Systems Research Division
Jack H. Hiller, Director

U.S. Army Research Institute for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria, Virginia 22333-5600

Office, Deputy Chief of Staff for Personnel
Department of the Army

January 1994

Army Project Number
2Q263007A795

Training Simulation

FOREWORD

The U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) conducts research on how to design unit training strategies. The Training and Doctrine Command (TRADOC) defines training strategy as the allocation and scheduling of resources across training events. The TRADOC definition includes "gate" measures for defining when units are prepared to move from one training event to the next. This report is part of a program which focuses on methodology for developing brigade training strategies.

A critical part of such methodology is a set of unit performance indicators to assess alternative strategies. To define these measures, we need to explore new ways to analyze the rich store of data emerging from Distributed Interactive Simulation (DIS) and from Combat Training Centers (CTCs). A promising candidate for such exploration is case-based expert system (CBES) technology.

This research product demonstrates how to use CBES to assess relationships among resource, process, and outcome measures in combat and training exercise data bases. It provides guidelines for analyzing the data, interpreting results, improving the quality of the data base, and conducting follow-up analyses. The report lays a foundation for applying CBES to CTC and simulation networking exercises.

EDGAR M. JOHNSON
Director

CASE-BASED EXPERT SYSTEMS FOR COMBAT PERFORMANCE ANALYSIS

CONTENTS

CONTENTS (Continued)

CASE-BASED EXPERT SYSTEMS FOR COMBAT PERFORMANCE ANALYSIS

1.0 INTRODUCTION

1.1 Overview of the Product

a. Objective: Demonstrate the use of CBES-based methodology to assess resource, process, and outcome relationships in combat and in training exercises. Develop guidelines for using CBES.

b. Summary of Guidelines:

(1) Use CBES to produce a rule for predicting battle or training exercise outcomes. Apply CBES to a sample of cases with features (i.e., resource or process measures) that you think influence battle or training exercise outcomes. An example in this paper demonstrates this for a sample of 50 division vs. division battles, fought since 1939. The rule generated was based on (a) subjective measures of tactics, climate, and geography, and (b) a combination measure of offensive and defensive resolution, e.g., penetration/withdrawal.

(2) Next, apply the rule to a new sample of similar cases and generate predicted outcomes. Ask two questions about the results of the application: Does the rule make sense? Do "predicted" outcomes match actual outcomes with high accuracy? If the answer is yes to both questions, you may have a relationship among resources, processes, and outcomes. This is more likely where you (or someone else) have designed and set up a data base to exercise CBES for well-defined questions. A sample application using notional data from National Training Center (NTC) and data from the board game TOBRUK was close to this situation. The target data base had been designed specifically for CBES analysis. The resulting rule was plausible and prediction accuracies were high.

(3) In contrast, low predictions or "implausible" rules are more likely when using an existing, large, diverse data base. For example, the prediction accuracy of CBES applied to the historical data was 41% for attacker outcomes and 65% for defender outcomes. These results and further analyses of the disagreements between predicted and actual outcomes, led to a set of procedures for using CBES to explore data bases systematically. The report outlines these procedures.

c. Use of the Product

This product provides a foundation and guidelines for using CBES to explore process-outcome relationships in Combat Training Center (CTC) and simulation networking data bases.

1.2 Background

a. The Army needs to develop unit training strategies that will effectively combine new training technologies with more traditional approaches. However, it lacks objective methods for doing so, especially for defining frequency of training and level-of-proficiency (i.e., gate) measures.[1] Training exercise data has significant potential for contributing to the development of such methods. Instrumented ranges (Hiller, 1987; Wiering, 1992), network simulations (Alluisi, 1991), and Improved Multiple Engagement Simulation (Kraemer & Koger, 1991) produce large amounts of data not available even 10 years ago. How can we select, process, and use these data to assist commanders in optimizing training strategies? This question is part of the basis for a research program aimed at building a decision support methodology (DSM) for training strategy development. This paper is the first in a series aimed at answering the question.

b. Answers should address persistent obstacles to unit performance measurement (Crumley, 1989; Hiller, 1987; General Accounting Office, 1986, 1991 (February), 1991 (Sept); Medlin & Thompson, 1980).

(1) Lack of a behavior model, valid for nonlinear, interactive combat, has been a prime obstacle to defining relevant performance measures (J. Banks, ARI, Personal Communication, 1992; Medlin, 1979; Mirabella, 1977, 1978). Such modeling is essential for valid, criterion-referenced measurement.

(2) How to aggregate data and set standards has been a related major obstacle (Allen, Johnson, Wheaton, Knerr, & Boycan, 1981; Hiller, McFann, & Lehowicz, 1990; Mirabella, 1977, 1978; Mirabella, Johnson, & Wheaton, 1980). Engagement simulation and unit evaluation research in the 70's set a foundation for much of the current unit training. But that research was constrained by the difficulties of collecting and interpreting combat exercise data.

(3) A persistent problem has been failure to account adequately for effects other than armor and infantry maneuvers. Nash (1991) expressed particular concern about air defense. Combat Service Support proved critical to success in Desert Storm.

---

[1] Training strategy means the allocation and scheduling of resources across such training events as FTXs, STXs, SIMNET, and CTCs. It requires unit commanders to assign tasks to events, specify the frequency of training per task and event, and define "gate" measures (Department of the Army, 1992).

(4) Most critical and intractable has been the problem of relating process and product measures of performance (MOPs), situational variables, and mission outcomes. Past efforts to solve this problem have met with mixed results. Crumley (1989) reported success in relating objective measures of command and staff decisions to field commander decisions. The connection decreased for ratings of task performance. Crumley implicated a reverse Heisenberg effect (Hiller, 1987). The effect occurs when observer/ controllers (O/Cs) influence battle outcomes through interventions. Crumley also noted that battle outcomes may severely bias O/C judgements about specific units. For example, effective armor teams may be downgraded because, overall, the exercise was poorly fought (McFann, 1990). The connection from command and staff performance to outcome measures nearly disappears.

c. In contrast to the mixed results reported by Crumley (1987), for performance within exercises, Hiller, McFann, and Lehowicz (1991) reported notable success in applying traditional parametric techniques to sequential training events. They related tank mileage ("OPTEMPO") and similarity between home station maneuver range and NTC to casualty exchange ratios at NTC. In a sense, they returned to basics of learning psychology for their impressive achievement. OPTEMPO reflects amount of practice (i.e., Thorndike's Law of Effect.)

d. But for break-throughs in unit performance analysis, we need to supplement traditional regression techniques with new ways to explore data bases (Frawley, Piatetsky-Shapiro, & Matheus 1991; Harrison & Hulin, 1989; Hart & Bradshaw, 1985; Strauss & Corbin, 1990; Wong & Chiu, 1987; Ziarko, 1991;). Traditional statistical techniques are well suited to comparing static conditions or making linear predictions with interval or ratio data, especially under experimental control. Such techniques are less well suited to relating the time dependent, interactive, situation specific, adaptive behaviors of units to mission success. (Hart & Bradshaw, 1985).

e. Approaches better suited to variables interacting over time include path, hierarchical, survival, and hazard rate analyses and Grounded Theory Procedures. These might reveal causal chains, which are obscured by more traditional statistical methods (Bart & Lane, 1982; Fichman, 1989; Harrison & Hulin, 1989; Kenny, 1979; Li, 1975; Morita, Lee, & Mowday, 1989; Strauss & Corbin, 1990). Related techniques have been used by the ARI Monterey Field Unit to analyze the interactions of battlefield operating systems (e.g., Root, Nichols, & Johnson, 1990). The techniques are adaptations of operational sequence diagramming, and PERT and GANTT charting.

f. A comprehensive study of how these various methods address the characteristics of unit training data and how they can help validate training strategies is desirable. But the resources--time and personnel required for such a study--are prohibitive.

An alternative is to select one promising method at a time, beginning with one that clearly suits the characteristics of unit training data and is sufficiently mature to be applied easily. Inductive reasoning from case data meets these criteria. Other methods will be explored in future work.

g. The method is designed to explore the kind of interactive sources of information found in combat and unit training data bases (Hart, 1991; Frawley, Piatetsky-Shapiro, & Matheus, 1991). It works well with the categorical and ordinal measurements prevalent in unit training. The method has been implemented in case-based expert system (CBES) programs. CBES can be thought of as a non-parametric version of regression analysis. It requires no statistical and minimal computer skills. CBES can be used to analyze data from training programs or exercises. The results of the analyses then form a basis for creating rules to designing training methods and strategies.

h. CBES has been extensively researched (Frawley, Piatetsky-Shapiro, & Matheus, 1991). It has been available in easy to use programs, for a decade (Export Software International, 1983; Milman, 1984). One of these applications, called 1st Class, was selected on the basis of contractor supported reviews of alternative inductive reasoning approaches.

i. This research product describes the use of CBES for analyzing historical and training exercise data (Hart, 1986; Rakoff, Laskey, Marvin, & Mandel, 1991; Uthurusamy, Fayyad, & Spangler, 1991). The focus is on understanding the methodology and developing ways to apply it to CTC and SIMNET data bases in support of research on training strategy methodology. Current U.S. Army Research Institute (ARI) R&D is aimed at designing a decision support methodology (DSM) to help brigades select training strategies. Critical to this is a set of formulas relating combinations of training events to unit effectiveness.

j. The formulas should weight the events with measures of satisfactory performance. But first we need to define measures that are reliable, valid, and useful to training proponents. The CBES can help us select the best indicators or combinations of indicators of mission success. It can even help organize these indicators into task clusters or task sequences to support training strategy development. For example, CBES analysis can supplement traditional task analyses based on SME surveys (Dressel, in preparation).

1.3 Objective

Demonstrate and explain how to use 1st Class (AI Corp., 1991), a CBES application program, to improve understanding about resource, process, & outcome relationships in combat and unit training data. Demonstrate use of the program with two types of data sources: (1) Pre-existing, large-scale archives and (2) files designed and formatted for case analysis. Develop guidelines for CBES use to explore data bases.

2.0 METHOD

2.1 Overview of the Method

a. 1st Class, a commercial version of CBES, was applied to a pre-existing, large data archive and then to files designed specifically for case analysis. 1st Class uses descriptions of cases and their outcomes to derive rules. The cases, for example, might be instances of illness. The descriptions would be symptoms and results of diagnostic tests. The outcomes would be disease classifications, specified by expert judgement or objective data. The descriptive and outcome data can be numerical as well as categorical. But 1st Class treats numbers as though they were categorical judgements. It forms categories from numbers by computing the midpoints between successive numbers.

b. From this information, 1st Class generates a compound rule (embedded "if" statements). Then, the user provides symptoms and diagnostic information for test cases. The computer identifies the disease for each case.

c. The rule is a decision tree that represents the most efficient set and sequence of diagnostic tests leading to a conclusion. The program evaluates all possible decision trees, using an information theory measure. It picks the most efficient tree. The most potent (i.e., discriminating, efficient) variable is at the top level. This top level variable (i.e., discriminator) branches to less potent, lower-level variables. 1st Class asks for information about the variable or variables at each level. It then gives the user a conclusion based on the information.

d. 1st Class is designed, therefore, to give advice. But it is also a tool for knowledge discovery in data bases (Frawley, et al., 1991). Data which are categorical or which violate assumptions of regression analysis suit the method. The CBES, like regression analysis, identifies potent predictor variables. It may be useful for CTC analyses, since much CTC data is categorical judgement. The INGRES data base at the ARI Field Unit - Presidio of Monterey (POM) is a powerful tool for analyzing the hard numbers. The CBES can be a potent supplement for analyzing the categorical data.

2.2 Demonstration 1. Depuy Data Base

a. The initial step in developing a CBES application to military cases was to demonstrate its use in assessing process-outcome relationships in historical combat data, archived in a large, pre-existing data base. The data base contained battles fought since 1939 (McQuie, 1988). Military historian Trevor DePuy, COL(R), compiled the data. For several hundred battles, the data base organizes information on identification, environment, tactics, results, and force structures. These categories are divided into 45 variables.

b. It quickly became apparent that the first problem in applying CBES to a pre-existing, large, diverse data base is to develop a manageable view of the data (i.e., a selection of cases and variables). The difficulty of developing a view without pre-conceived research questions also became apparent. In this case, it was not difficult to do so since the research focused on showing how to use CBES and on developing ways to apply it to CTC and SIMNET data bases.

c. Division vs. division battles were selected since these were the lowest echelons available. Environment, tactics, and results were selected as sources of predictor and criterion variables, because these were categorical and seemed most relevant to training. CBES works best with categorical variables.

d. The environmental variables were terrain, cover, humidity and temperature. The tactical variables were maneuver, width of attack, presence or absence of major surprise, and defensive posture (Appendix A). The criterion measure was a combination of attacker and defender resolution. The categories of resolution are defined in Appendix A.

e. Fifty division vs. division battles (Appendix B) were used to generate a compound rule (Figure 1). The rule, in turn, was applied to an additional 37 division vs. division battles to get predicted outcomes. I tabulated the joint frequencies of correct and incorrect predictions for offensive, defensive, and combined offensive /defensive outcomes. To estimate consistency of results, I used CBES to generate rules for Cases 1 - 8, 1 -16, and 1 - 30.

```
Rule for knowledge base DEPUY                    2:19 pm  03/24/1992

---- start of rule ----
TEMPERATURE??
 hot:WIDTH??
    <16.:WIDTH??
      <9.:------------------------------------------ brkthrWDr
      >9.:WIDTH??
        <10.5:DEFPOSTURE??
          fortified:------------------------------- no_data
          prepared:-------------------------------- PenWithDR
          hasty:----------------------------------- brkthrWDr
          delay:----------------------------------- PenWithDR
        >10.5:----------------------------------- PenWithDR
    >16.:SURPRISE??
     yes:------------------------------------------ PenStale
     no:------------------------------------------- RepulseStal
 temp:WIDTH??
    <2.5:COVER??
     bare:------------------------------------------ Unknown
     mixed:HUMIDITY??
       dry:------------------------------------------ PenWithDR
         & - - - - - - - - - - - - - - - - - - - - - RepulseStal
       light:---------------------------------------- PenWithDR
         & - - - - - - - - - - - - - - - - - - - - - PenStale
       heavy:---------------------------------------- PenWithDR
     desert:---------------------------------------- no_data
     swamp:----------------------------------------- no_data
     wooded:---------------------------------------- no_data
    >2.5:WIDTH??
      <3.5:MANEUVER??
       frontal:--------------------------------- RepulseStal
       envelopment:----------------------------- no_data
       penetration:----------------------------- no_data
       dbl_env:--------------------------------- RepuleWDr
       river_cross:----------------------------- no_data
      >3.5:WIDTH??
        <17.5:WIDTH??
          <4.5:------------------------------- PenWithDR
          >4.5:WIDTH??
            <7.5:WIDTH??
              <5.5:DEFPOSTURE??
                fortified:------------------- PenWithDR
                prepared:-------------------- PenWithDR
                hasty:----------------------- RepulseStal
                delay:----------------------- PenWithDR
              >5.5:DEFPOSTURE??
                fortified:SURPRISE??
                  yes:----------------------- PenWithDR
                  no:------------------------ RepulseStal
                prepared:-------------------- no_data
                hasty:----------------------- PenWithDR
                delay:----------------------- PenWithDR
            >7.5:WIDTH??
              <8.5:------------------------- PenWithDR
              >8.5:WIDTH??
                <9.5:DEFPOSTURE??
                  fortified:HUMIDITY??
                    dry:----------------------- PenWithDR
                    light:--------------------- RepulseStal
                    heavy:--------------------- no_data
                  prepared:-------------------- PenWithDR
                  hasty:----------------------- no_data
                  delay:----------------------- PenWithDR
                >9.5:------------------------- PenWithDR
       >17.5:------------------------------------ RepuleWDr
```

Figure 1. Rule for Depuy data base

```
64:  cold:WIDTH??
65:     <6.:WIDTH??
66:       <2.5:DEFPOSTURE??
67:        fortified:---------------------------------- PenStale
68:        prepared:----------------------------------- PenWithDR
69:        hasty:-------------------------------------- no_data
70:        delay:-------------------------------------- no_data
71:       >2.5:TERRAIN??
72:        Flat:--------------------------------------- RepulseStal
73:        rolling:------------------------------------ PenWithDR
74:        rugged:------------------------------------- no_data
75:     >6.:HUMIDITY??
76:      dry:----------------------------------------- PenStale
77:      light:--------------------------------------- PenWithDR
78:      heavy:--------------------------------------- no_data
    ---- end of rule ----
```

Figure 1. (continued)

2.3 Demonstration 2. Rakoff Data Bases

a. The purpose of this effort was to demonstrate use of CBES with data bases designed for case-based analysis in training contexts. This purpose contrasts with Demonstration 1 which used a pre-existing, large, diffuse, historical combat archive. In this second application, we would expect more accurate predictions, less likelihood of implausible rules, and therefore easier knowledge discovery. Rakoff, et al. (1991) provided case files which satisfied the purpose of Demonstration 2. Rakoff, et al. generated NTC notional data and data from the board game TOBRUK to demonstrate the use of a neural net technology for assessing battle process vs. mission outcome relationships.

b. The NTC notional data came from 20 hypothetical NTC exercises. Nine variables were given random values (Appendix C). Military experts then judged the exercise outcomes as success (S) or failure (F). Rakoff et al. used sets of 19 exercises to build the predictive model. They applied it to an excluded exercise. For example, they excluded Exercise 1 and used Exercises 2-20 for modeling. Then, in turn, each following exercise was excluded and the remaining exercises used to model. The authors applied a similar method of analysis and test to 30 exercises of TOBRUK. Appendix D lists the variables. Appendix E lists the exercises from which predictor cases were drawn.

c. I repeated their procedures. But I also used subsets of exercises to build predictive models. Rules were generated for NTC Exercises 1 - 5, 1 - 10, 1 - 15, and 1 - 20. For TOBRUK, I built rules for Cases 1 - 15, 1 - 20, and 1 -30. I wanted to examine the consistency of the rules.

3.0 RESULTS

3.1 Demonstration 1: Depuy Data Base

a. Figure 1 shows the rule generated from 50 cases. The top level (Level 1) discriminator is temperature. Width of the attacker's front is the second level (i.e. key contingent) variable. Surprise, cover, humidity, defensive posture, maneuver, and terrain are third level discriminators. This means that the program first asks for information about temperature. Is the temperature hot, temperate, cold? The program next asks for data on the width of the attack. It may now draw a conclusion about resolution. Or, it may continue to pursue additional categories of information.

b. Tables 1 - 4 show the results of applying the rule to 37 division vs. division cases. Tables 1 and 2 show frequencies of various combinations of actual and predicted outcomes for attacker and defender. The diagonal frequencies are correct classifications. Off-diagonal frequencies are misclassifications.

c. The rule predicted attacker outcomes with 41% accuracy. Table 1 shows a major misclassification in the prediction of penetration, where repulse was the actual outcome. If this source were removed, prediction accuracy would be 64%. Table 2 shows major confusion between withdrawal and stalemate for the defender. If this source were removed, accuracy would be 96%.

d. Table 3 is the distribution of correct joint classifications for attacker and defender. If cases with a misclassification were removed, prediction accuracy would be 77%. Table 4 indicates the distribution of four test cases for which 'advice' was not available. Presumably, the original examples did not include patterns which matched the test cases in Table 4.

e. Table 5 summarizes how the rules changed as the number of cases increased. The major discriminators at levels 1 through 4 were identified consistently, beginning with 16 cases. The size of the decision tree is four or five levels until the number of cases reaches 50. There the decision doubles to 10 levels.

3.2 Demonstration 2: Rakoff Data

a. NTC notional exercises. Figure 2 shows the rule generated from 20 exercises. Classification accuracy was 90%. The rule indicated that two variables accounted for the predictions. These were Engagement Ratio (Defense to Offense) and Security of Defense (Sec Def). But defensive security influenced only Exercise 16. Engagement Ratio was a sufficient predictor for every other case. The rule with Engagement Ratio alone predicted every outcome, except for Exercise 16.

Table 1. Frequency of Actual and Predicted Battle Resolution for Attacker

Actual	Predicted		
	Penetration	Break through	Repulse
Penetration	12	2	3
Break through	2	1	1
Repulse	12	0	1

Accuracy: 14/34 = 41%
"Corrected": 14/22 = 64%

Table 2. Frequency of Actual and Predicted Battle Resolution for Defender

Actual	Predicted		
	Withdrawl	Delay	Statemate
Withdrawl	17	0	0
Delay	0	1	0
Statemate	11	0	4
Pursuit	1	0	0

Accuracy: 22/34 = 65%
"Corrected": 22/23 = 96%

Table 3. Frequency of Correct Joint Classification

Actual	Defender		
	Withdrawl	Delay	Statemate
Penetration	14	0	1
Breakthru	1	0	0
Repulse	0	0	1

Accuracy: 17/34 = 50.00%
"Corrected": 17/22 = 77.27%

Table 4. Frequency of "No Advice"

	Actual Outcomes of Test Cases		
	Withdrawal	Delay	Stalemate
Penetration	2		1
Breakthru	1		

Table 5. Rule Structure vs. Number of Cases (Depuy Data)

Rule Characteristic (8 Input Variables)	Cases			
	1 to 8	1 to 16	1 to 30	1 to 50
Number of Variables in Rule	4	4	5	8
Number of Lines in Rule	15	21	44	78
Discriminator (s)				
1st Level	Manuever	Temp.	Temp.	Temp.
2nd Level	Cover	Width	Width	Width
		DefPost.		
3rd Level	Width	Width	Width	Width
		Humidity	Humidity	Humidity
			Maneuver	
4th Level	Surprise	Width	Width	Width
			DefPost.	DefPost.
				Humidity
				Maneuver
				Terrain
5th Level		Surprise	Surprise	
			DefPost.	DefPost.
				Width
6th Level				Width
7th Level				Width
8th Level				DefPost.
9th Level				Surprise
				DefPost.
10th Level				Humidity

Figure 2. Rule for knowledge base (NTC-RAK)

```
     ---- start of rule ----
 1:  OFFENSIVE??
 2:    <4.5:SECURITY??
 3:           <8.5:------------------------------------------ F
 4:           >8.5:------------------------------------------ S
 5:    >4.5:------------------------------------------------- S
    ----- end of rule -----
```

```
     ---- start of rule ----
 1:  MASS??
 2:    <2.65:MOBIL??
 3:           <1.25:SEC_D??
 4:                  <0.9:------------------------------- F
 5:                  >0.9:F_POWER??
 6:                       <1.15:------------------------ F
 7:                       >1.15:------------------------ S
 8:           >1.25:------------------------------------- F
 9:    >2.65:MOBIL??
10:           <1.55:------------------------------------- S
11:           >1.55:MASS??
12:                  <4.85:MIX??
13:                       <0.35:------------------------ F
14:                       >0.35:MOBIL??
15:                             <2.3:MOBIL??
16:                                  <1.75:SEC_D??
17:                                       <0.8:------- S
18:                                       >0.8:MIX??
19:                                            <0.45:-------- S
20:                                            >0.45:-------- F
21:                                  >1.75:---- S
22:                             >2.3:----------- F
23:                  >4.85:----------------------- S
```

Figure 3. Rule for knowledge base (TOBRUK)

b. TOBRUK data. Figure 3 shows the rule generated from 30 exercises. Accuracy was 63%. Combat mass was the primary discriminator. Mobility was the second-level discriminator. Defensive security and mass are third-level discriminators. No variable was excluded by 1st Class. Five out of five tested were discriminators.

c. Stability Data.

(1) Table 6 summarizes how the rules changed as the number of NTC cases increased. In contrast to the Depuy data, the rule for NTC remained nearly constant. An additional variable was identified for the 20-case situation.

(2) Table 7 summarizes how the rules changed as the number of TOBRUK cases increased. The TOBRUK rule varied substantially from 15 to 30 cases. In particular, the decision tree doubled. Nonetheless, the major variable - Mass - was identified throughout. Mobility and Sec Def were consistent for two of the three sample sizes.

Table 6. Rule Structure vs. Number of Cases (NTC NOTIONAL DATA)

Rule Characteristic (9 Input Variables)	Cases			
	1 to 5	1 to 10	1 to 15	1 to 20
Number of Variables in Rule	1	1	1	1
Number of Lines in Rule	3	3	3	3
Discriminator (s)				
1st Level	offensive	offensive	offensive	offen.
2nd Level				security

Table 7. Rule Structure vs. Number of Cases (TOBRUK DATA)

Rule Characteristic (5 Input variables)	Cases		
	1 to 15	1 to 20	1 to 30
Number of Variables in Rule	4	4	4
Number of Lines in Rule	9	11	23
Discriminator(s)			
1st Level	Mass	Mass	Mass
2nd Level	Mix	Mobility	Mobility
3rd Level	Sec Def	Power	Sec Def
			Mass
4th Level	Mobility	Mobility	Power
			Mix
5th Level		Mix	Mobility
6th Level			Mobility
7th Level			Sec Def
8th Level			Mix

4.0 HOW TO USE CBES

4.1 Model (i.e., 'Lessons Learned') for Analyzing CTC and SIMNET Data

The results suggest a set of procedures for future applications. The procedures are outlined below and illustrated using the results of the pilot applications.

a. Examine the details of the rule (i.e., decision tree).

(1) What predictor variables are included? How are they prioritized? What interactions are shown? If the rule doesn't make sense, try to determine why not. For example, the Depuy rule indicates that temperature is the major discriminator. 1st Class showed that the most efficient route to prediction begins with a test of temperature. It continues with a test of the width of the attacker's front. Whether these key variables are plausible or useful is another matter.

(2) Temperature, as a high priority variable, doesn't appear to be either plausible or useful. But it may be a clue to other factors that do make sense or are more useful discriminators for the specific purpose of the analysis. For example, temperature may correlate with theater. Recommended Action: Sort the examples on temperature, with theater as a criterion. If the association is compelling, block on theater. Then, re-analyze to see if the predictions for test cases improve.

b. Examine the frequency distributions of false classifications for clues to performance modeling or analysis problems.

(1) Alternative or additional sources of variance may need to be uncovered (J. Banks, Personal Communication, 1992). Complex interactions may be prevalent, given the non-linear flow of combat (Hiller, 1987). Measurement scales may be deficient. Or, scales may have been applied unreliably.

(2) For misclassifications of attacker outcomes in Table 1, additional sources of variance may need to be uncovered. For example, to continue analyzing the Depuy data base, we might do the following: Isolate the cases in the repulse-penetration cell and examine them for candidate confounding variables. Reanalyze the data with these variables included. Examples of candidates are force ratio and ratio of attacker artillery to defender artillery (In principle, this is easy since 1st Class reads Lotus and dBase).

c. Examine misclassifications for clues to poor definitions or unreliable judgements of outcomes. Recommended actions:

(1) Discard the outcome scales. Consider alternative criterion measures. For example, the Depuy data base provides alternatives (or data for computing alternatives). Some examples are casualty-exchange ratio, duration of battle, and weapon kills.

(2) Revise the outcome scales (Macpherson, Personal Communication, 1992). Macpherson suggests two options. One is to apply psychological or statistical scaling techniques to categorical judgements. This is defensible, if a plausible numerical scale lies beneath the judgements. He believes such a scale may exist for the outcome judgements in the Depuy data base. A second option is to transform numerical outcome scales. 1st Class accepts numbers, but converts them to artificial categories. The numbers are thus forced into a uniform statistical distribution. The underlying distribution may be represented better by a normal, log, or reciprocal transformation. If so, prediction accuracy should improve. These options also can apply to the predictor variables.

(3) Analyze reliability of outcome judgements (if the raw data are available). Or, collect new data with careful attention to reliability issues.

d. Examine Missing Predictions. For the Depuy data, only a few test cases did not generate predictions ("no advice"). In future applications of 1st Class, many cases with no advice available might suggest that the original set of examples was too small. Or, it may have major data gaps. Recommended Action: increase the example set. Document the cells with no data.

4.2 Stability and Sensitivity of the Predictive Models. The results suggest that even with a moderate sample, CBES can identify the most salient predictors of combat effectiveness. A preliminary rule of thumb is 15 cases minimum and two cases per variable. As the number of cases increases, more interactions emerge. But their significance and value is not clear. Level 5 through 10 discriminators may be too idiosyncratic, too subtle to have practical utility. But they may serve a heuristic purpose for research.

4.3 Use in Defining Performance Measures and Training Strategies

a. General Application to NTC, SIMNET, or home-station data. Rules, tables, and reasoning processes similar to those in Section 4.1 could be generated to examine the quality of the data. Rules and confusion matrices then could provide clues to further analyses of measurement methods.

b. Results of these analyses, in turn, may stimulate ideas about training strategies. For example, we might extract hidden sources of variance at high values of operational tempo at NTC. In turn these sources might suggest improved training strategies.

c. Support for refining or using the Bde task analytic data.

(1) For illustration, assume the following. The analysis has documented a command and staff planning operation for Bde maneuver in Movement to Contact. We've identified three planning products. These might include an operations order, intelligence report, and fire support plan. We also have scales for assessing how well the command staff prepared the products and how well subordinate commanders implemented them.

(2) 1st Class would define the relationship between quality of planning products and quality of plan execution. The expert system would identify which of the prediction variables discriminate among levels of subordinate performance. It would prioritize the variables for their potency. These results could help define training events and assign training resources.

d. Support for conducting Bde Task analysis. 1st Class screen variables for use in quantitative task analysis methods. Such methods include path, hierarchical, cluster, and factor analysis. 1st Class would be applied in the ways described earlier but used to hypothesize dependencies among tasks. Given sufficient data from NTC, SIMNET, or quasi-experimental studies, the dependencies could be tested.

4.4 Recommended Applications of CBES

a. Application to CTCs (e.g. NTC, JTRC) Data.

(1) Analysis of Observer/Controller Data at the CTCs. CBES is suitable for analyzing, validating, and improving O/C judgements of unit performance by battlefield operating system (BOS). Instruments for collecting O/C data include the BOS Impact statement and the O/C Training and Evaluation Outline (T&EO) Checklist. The Impact Statement is a narrative evaluation of performance for one BOS, one exercise (Alderman, 1992; Kerins, Atwood, & Root, 1990). The O/C T&EO Checklist is used at the Joint Training Readiness Center (JTRC) to assess task proficiency within BOSs (Thompson, Thompson, Pleban, and Valentine, 1991).

(2) Each of these instruments can be structured in case format. The METT-T Score (Alderman, 1992; Kerins, Atwood, & Root, 1990) provides a criterion measure. An alternative is METT-T separated into offensive and defensive components, (like the criterion score in the Demonstration 1). These components could then be used in a misclassification analyses, similar to those in Demonstration 1. The analyses can be used to identify, screen, and validate performance measures for a decision aid on unit training strategies. They can contribute to assessments of training effectiveness and value-added.

b. Application to data from Pre-CTC training.

(1) The results of Demonstrations 1 and 2 also provide a basis for applying CBES to training events leading to CTC rotations. The purpose of such analysis is to assess the relationships between those events and performance at CTC rotations.

(2) To make the analysis most useful, select measures which are common to home station and CTC training, or which provide a clear link between the two stages of training. BOS measures are good candidates. These and other common measures or schemes of measurement have been discussed by Alderman, (1993), Forsythe (1987), Kerins, Atwood, & Root (1990), Madden (1991), and Root, Nichols, & Johnson (1990).

5.0 CONCLUSIONS

5.1 Case-based expert system technology shows potential for analyzing CTC and SIMNET data bases by identifying and validating performance measures for combat unit performance.

5.2 CBES can be applied to pre-existing, large, diffuse archives or to case files designed for CBES application. This research product provides guidelines for doing so.

5.3 The technology needs only a moderate number of cases (15 or more; 3 cases per variable) to provide stable results.

6.0 REFERENCES

Alderman, I. (1993). Methods for identifying cost-effective training strategies. Military Operation Research Society, 60th Symposium, Contingency Operations in a New World Order, 23-25 June, 1992, Monterey, CA.

AI Corp. (1991). 1st class tutorials and knowledge engineering guide. Waltham, MA.

Allen, T. W., Johnson, E. J., Wheaton, G. R., Knerr, C. M., & Boycan, G. G. (1981). Methods of evaluating tank platoon battle run performance (ARI Technical Report 642). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A159 454)

Alluisi, E. A. (1991). The development of technology for collective training: SIMNET, a case history. Human Factors Journal, 33(3), 343-362.

Bart, W. M., & Lane, J. F. (1982). Advances in the analysis of hierarchical skill structures (ARI Final Report, Contract MDA903-81-C-0394). Minneapolis, MN: University of Minnesota.

Crumley, L. M. (1989). Review of research and methodologies ARI relevant to Army command and control performance measurement (ARI Technical Report 825). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A211 247)

Department of the Army (1992, August 13). Combined Arms Training Strategy (CATS) development (TRADOC Draft Pamphlet 350-xx), Ft. Monroe, VA: Training and Doctrine Command.

Dressel, J. D. (1993). Preliminary analysis of brigade operations survey for brigade task identification (ARI Research Report 1655). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Export Software International, Ltd. (1983). Expert-ease user manual. Edinburgh, Scotland: Darien Books.

Fichman, M. (1989). Attendance makes the heart grow fonder: A hazard rate approach to modeling attendance. Journal of Applied Psychology, 74(2), 325-335.

Forsythe, T. K. (1987). A research concept for developing and applying methods for measurement and interpretation of unit performance at the National Training Center (ARI Research Report 1435). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A181 073)

Frawley, W. J., Piatetsky-Shapiro, & Matheus, C. J. (1991). Knowledge discovery in databases: An overview. In G. Piatetsky & W. J. Frawley (Eds.), Knowledge Discovery in Databases. Menlo Park, CA: AAAI/MIT Press.

General Accounting Office. (1986, April). Army training: National Training Center's potential has not been realized. (GAO/NSIAD-86-130). Washington, DC.

General Accounting Office. (1991, February). Army training: Evaluation of units' proficiency are not always reliable. (GAO/NSIAD-91-72). Washington, DC.

General Accounting Office. (1991, September). Army training: Peacetime training did not adequately prepare combat brigades for Gulf War. (GAO/NSIAD-91-263). Washington, DC.

Hart, A. (1986). Knowledge acquisition for expert systems. New York: McGraw-Hill.

Hart, R. J., & Bradshaw, S. C. (1985). Interactivity theory: Analyzing human environments using linear prediction filters (ARI Technical Report 686). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A172 066)

Harrison, D. A., & Hulin, C. L. (1989). Investigations of absenteeism: Using event history models to study the absence-taking process. Journal of Applied Psychology, 74(2), 300-316.

Hiller, J. H. (1987). Deriving useful lessons from combat simulations. Defense Management Journal, Second and Third Quarter, 29-33.

Hiller, J. H., McFann, H. H., & Lehowicz, L. G. (1991). Does OPTEMPO increase unit readiness? An objective answer. Proceedings of the 1990 Army Science Conference.

Kenny, D. A. (1979). Correlation and causality. New York: John Wiley & Sons.

Kerins, J. J., Atwood, N. K., & Root, J. T. (1990). Concept for a common performance measurement system for unit training at the National Training Center (NTC) and with Simulation Networking (SIMNET) (ARI Research Report 1574). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A230 129)

Kraemer, R. E., & Koger, M. E. (1991, January). Precision Range Integrated Maneuver Exercise (PRIME) user's guide (ARI Research Product 91-05). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 411)

Li, C. C. (1975). Path analysis-a primer. Pacific Grove, CA: The Boxwood Press.

Madden, J. L. (1991). Unit training management system, strategy, and program for Simulation Networking (SIMNET) (ARI Research Report 1592). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A239 916)

McFann, H. H. (1990). Relationship of resource utilization and expenditures to unit performance. Proceedings of the NATO Defense Research Group (Panel VII) RSG-15 Workshop on the Military Value and Cost Effectiveness of Training. Hemingford Grey House, Huntingdon, Cambridgeshire, England.

McQuie, R. (1988). Historical characteristics of combat for wargames (benchmarks) (Research Paper CAA-RP-87-2). Bethesda, MD: U.S. Army Concepts Analysis Agency.

Medlin, S. (1979). Combat Operations Training Effectiveness Analysis (COTEAM) model (ARI Technical Report 393). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A077 839)

Medlin, S., & Thompson, P. (1980). Evaluator rating of performance in field exercises: A multidimensional scaling analysis (ARI Technical Report 438). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A089 264)

Milman, J. O. (1984). Do-it-yourself expert systems with artificial intelligence on a Microcomputer. Proceedings of the IEEE Compucon '84 Fall Conference on the Small Computer (R) Evolution, Arlington, VA.

Mirabella, A. (1977). Criterion-referenced system approach to evaluation of combat units. Proceedings of the 19th Military Testing Association Meeting, San Antonio, TX.

Mirabella, A. (1978). Criterion-referenced system approach to evaluation of combat units (ARI Research Memorandum 78-21). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A077 968)

Mirabella, A., Johnson, E. J., & Wheaton, G. R. (1980). Criterion definition for tactical training of combat units. Proceedings of the 22nd Annual Conference of the Military Testing Association, Toronto, Canada.

Morita, J. G., Lee, T. W., & Mowday, R. T. (1989). Introducing survival analysis for organizational researchers: A selected application to turnover research. Journal of Applied Psychology, 74(2), 280-292.

Nash, T. (1991). The Tactical Air Defence Training Requirement. Military Simulation and Training. Issue 6/1991, 40-43.

National Simulation Center. (1992). Operational Requirements Document for Warfighters' Simulation 2000. Fort Leavenworth, KS.

Rakoff, S. H., Laskey, K. B., Marvin, F. F., & Mandel, J. S. (1991). Conceptual models of unit performance (ARI Research Note 91-19). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A232 743)

Root, J. T., Nichols, J. J., & Johnson, C. (1990). Methodology for sub-BOS development, prototype: Mortars (Unpublished). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Strauss, A., & Corbin, J. (1990). Basics of qualitative research. Newbury Park, NJ: Sage Publications.

Thompson, T. J., Thompson, G. D., Pleban, R. J., & Valentine, P. J. (1991). Battle staff training and synchronization in light infantry battalions and task forces (ARI Research Report 1607). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A245 292)

Uthurusamy, R., Fayyad, U. M., & Spangler, S. Learning useful rules from inconclusive data. In: G. Piatetsky & W. J. Frawley (Eds.), Knowledge Discovery in Databases. Menlo Park, CA: AAAI/MIT Press.

Wiering, G. (1992). A diamond in the rough: The National Training Center. Army Research, Development, & Acquisition Bulletin. March-April, 18-21.

Wong, A. K. C., & Chiu, D. K. Y. (1987). Synthesizing

statistical knowledge from incomplete mixed-mode data. IEEE Trans-actions on Pattern Analysis and Machine Intelligence, Vol. PAMI-9(6), November.

Ziarko, W. (1991). The discovery, analysis, and representation of data dependencies in databases. In G. Piatetsky & W. J. Frawley (Eds.), Knowledge Discovery in Databases. Menlo Park, CA: AAAI/MIT Press.

Appendix A Definitions for Knowledge Base Depuy

Predictor Variables		Attack/Defend Resolution
Variable	Value(s)	
Defensive Posture	Fortified Prepared Hasty Delay	Penetration/Withdrawal Repulse/Stalemate
Width of Attack	Miles	Penetration/Stalemate
Maneuver	Frontal Envelopment Penetration Double Envelopment River Cross	Withdrawal/Pursuit Bypass/Stalemate Breakthrough/Withdrawal
Surprise	Yes No	Repulse/Withdrawal
Temperature	Hot Temperate Cold	
Humidity	Dry Light Heavy	
Cover	Bare Mixed Desert Swamp Wooded	
Terrain	Flat Rolling Rugged	

Appendix B Examples of Knowledge Base Depuy

7:53 am 02/04/1992

	HUMIDITY	COVER	TERRAIN	RESOLUTION	weight
1:	dry	mixed	Flat	brkthrWDr	1.00
2:	dry	bare	Flat	Unknown	1.00
3:	light	mixed	Flat	PenWithDR	1.00
4:	light	mixed	Flat	RepulseStal	1.00
5:	heavy	mixed	Flat	PenWithDR	1.00
6:	dry	mixed	Flat	PenWithDR	1.00
7:	dry	mixed	rolling	PenWithDR	1.00
8:	dry	mixed	rolling	PenWithDR	1.00
9:	dry	mixed	rolling	PenStale	1.00
10:	dry	mixed	rolling	RepulseStal	1.00
11:	dry	desert	Flat	PenWithDR	1.00
12:	dry	desert	Flat	brkthrWDr	1.00
13:	dry	desert	Flat	brkthrWDr	1.00
14:	light	mixed	rolling	PenWithDR	1.00
15:	light	mixed	rolling	PenWithDR	1.00
16:	dry	mixed	rolling	PenWithDR	1.00
17:	light	mixed	rugged	PenStale	1.00
18:	dry	mixed	Flat	PenStale	1.00
19:	dry	mixed	rolling	RepulseStal	1.00
20:	light	mixed	Flat	PenWithDR	1.00
21:	dry	mixed	Flat	RepulseStal	1.00
22:	dry	mixed	Flat	PenStale	1.00
23:	dry	mixed	Flat	PenWithDR	1.00
24:	dry	bare	rugged	PenWithDR	1.00
25:	dry	bare	rugged	RepuleWDr	1.00
26:	dry	desert	rolling	PenWithDR	1.00
27:	light	bare	rolling	PenStale	1.00
28:	light	bare	rolling	PenWithDR	1.00
29:	light	mixed	Flat	PenWithDR	1.00
30:	light	mixed	rolling	PenWithDR	1.00
31:	dry	mixed	rolling	PenWithDR	1.00
32:	dry	desert	Flat	brkthrWDr	1.00
33:	dry	bare	rolling	RepuleWDr	1.00
34:	dry	bare	rugged	PenWithDR	1.00
35:	dry	bare	rugged	RepulseStal	1.00
36:	dry	mixed	Flat	PenWithDR	1.00
37:	dry	bare	rugged	PenWithDR	1.00
38:	dry	desert	rolling	RepulseStal	1.00
39:	dry	mixed	rugged	PenWithDR	1.00
40:	dry	desert	Flat	PenWithDR	1.00
41:	heavy	wooded	rolling	brkthrWDr	1.00
42:	dry	bare	rugged	PenStale	1.00
43:	light	mixed	rugged	PenWithDR	1.00
44:	dry	mixed	rolling	RepulseStal	1.00
45:	dry	mixed	rolling	PenWithDR	1.00
46:	heavy	mixed	rolling	PenWithDR	1.00
47:	light	mixed	rolling	PenWithDR	1.00
48:	dry	mixed	Flat	RepulseStal	1.00
49:	dry	mixed	Flat	RepulseStal	1.00
50:	light	mixed	Flat	RepulseStal	1.00

MEASURES	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Objec. - Dir. of attack 1800- Deflec. degrees from objec.	9	8	7	6	5	4	3	2	1	0	0	1	2	3	4	5	6	7	8	9
2. Defen. - WCTD by weapon sys. ratio of vehicle engaged (% def/% ATT)	6	7	1	0	9	4	3	5	8	2	9	8	7	0	2	1	5	4	3	1
3. Masc - WCTD by weapon sys. force ratio (ATT/DEF)	7	8	5	6	2	0	1	6	7	2	5	5	3	1	5	8	7	6	9	2
4. Econ. of force sec. ratio (main /second. ATT)	9	7	0	5	2	8	1	6	0	4	4	3	8	7	5	2	3	5	4	8
5. Maneuver - Speed (Avg KM/WR)	9	5	9	8	7	2	1	8	2	4	3	9	6	3	3	7	4	5	0	1
6. Unity of Com.- Span of Control (% unit within dist)	8	2	5	9	4	7	5	6	0	2	1	5	9	6	7	0	2	9	8	1
7. Security - ADA coverage (% force covered)	5	7	7	6	9	8	0	1	6	5	3	3	4	6	8	9	1	1	4	8
8. Surp.- Defense readiness (% of def moving)	4	3	8	9	8	7	4	2	0	2	4	2	5	6	7	7	2	1	0	3
9. Simplicity - scheme of maneuver (# groups/direc)	8	9	9	2	5	6	2	6	5	3	4	2	5	4	3	7	8	7	8	6
OUTCOME	S	S	F	F	S	F	F	S	S	F	S	S	S	F	F	S	S	F	F	F

*(Rakoff et al., 1991)

Appendix D Variables for Tobruk Data Analysis *

Variable	Definition
Mass	Attacker to defender vehicle ratio at end of turn
Mix	Number of attacker tanks/total attacker vehicles at end of turn
Firepower	Number of attacker vehicles fired in turn/number of defender vehicles fired in turn
Mobility	Total hexes moved by attacker/total attacker vehicles at beginning of turn
Security-Defender	Percent of defender with flanks protected at beginning of turn

* (Rakoff et al., 1991)

Appendix E Tobruk Data Input*

Case/Measure	Mass	Mix	F'Power	Mobil	Sec-D
1	1.3	1.0	0.7	2.3	0.9
2	2.0	1.0	2.0	0.0	0.8
3	3.6	0.5	1.0	1.6	1.0
4	3.4	0.4	3.3	0.7	1.0
5	3.8	0.5	1.8	1.6	0.6
6	2.8	0.4	1.0	1.6	1.0
7	3.3	0.4	3.0	1.2	1.0
8	2.3	0.3	1.5	1.2	0.8
9	1.8	0.4	1.0	1.2	1.0
10	5.0	0.3	1.0	2.1	1.0
11	3.3	0.3	1.5	1.4	0.7
12	2.2	0.3	2.0	1.7	1.0
13	3.8	0.7	1.0	2.1	0.3
14	2.5	0.7	2.0	1.5	0.8
15	3.5	7.0	0.0	2.1	1.0
16	1.8	0.6	0.5	1.3	1.0
17	3.3	0.9	1.3	1.5	0.3
18	2.5	0.7	0.4	1.9	1.0
19	3.3	0.8	0.0	2.5	1.0
20	6.3	0.1	2.0	1.9	1.0
21	5.0	0.4	1.0	1.8	1.0
22	4.7	0.3	1.5	1.6	1.0
23	3.3	0.5	0.0	1.9	1.0
24	1.8	0.6	1.0	0.9	0.8
25	2.3	0.6	3.9	1.6	1.0
26	3.0	0.1	2.0	0.9	0.7
27	3.3	0.2	1.0	0.8	1.0
28	3.3	0.7	4.0	1.1	0.8
29	2.0	0.7	1.3	1.2	1.0
30	2.8	0.2	0.0	1.8	1.0

*(Rakoff et al., 1991)

Appendix F Comparisons of Results Provided by CBES and Neural Net

a. The application of neural nets by Rakoff et al. provided an incidental target of opportunity to make some comparative observations on neural nets and CBES. Formal, rigorous comparison was not a purpose of the research and is not the intention of the following comments. Such a study would be very useful, especially if it showed how to use the methods in complementary ways. But it's beyond the scope of the present effort. The following observations and comments may suggest some questions to be answered by a rigorous comparison.

b. The predictions of 1st Class in Demonstration 2 and neural net (in Rakoff et al., 1991) seemed to provide about the same information. For example, they were within 10% of each other. But 1st Class is easier to understand and apply. The mathematics for neural nets is formidable. The user needs to understand some of it. In contrast, clerk-typists can use CBES by typing in case data and selecting menu options. 1st Class provides easy to learn spread sheet templates for entering variables and case data. Alternatively, the user can import dBase or LOTUS data.

c. CBES generates a decision tree which lets the user see the relative importance of the predictor variables. The more potent the variable, the higher it is on the tree. Moreover CBES shows specific values of the variables which distinguish success from failure. This feature May help define standards of performance. The neural net also shows relative importance of the predictor variables. In fact, it does so more precisely by assigning weights. It may not be as useful in identifying specific predictor values which separate success from failure.

d. CBES also shows how variables interact. It showed interactions in each application. Rakoff et al. reported no interactions with the neural net. Whether it's less sensitive to interactions is not clear. CBES excludes variables which do not contribute to outcome decisions. Neural net does not. However, weights close to zero indicate such variables. An example is Mix, at -1.05.

d. The CBES and neural net analysis agree partially on the predictability of specific input variables for TOBRUK.

(1) Both rank Mass and Mobility 1st and 2nd.

(2) They reverse the 3rd and 4th level discriminators. Neural net puts firepower slightly above Sec Def. 1st Class reverses this order.

(3) They disagreed, notably, on Mix. Mix did not influence prediction accuracy in neural net. 1st Class, on the other hand, included Mix in its rule.

e. The disagreements are constructive. They flag variables which may need to be examined further in one or both models. For example, I eliminated Mix from the 1st Class model. Model accuracy remained unchanged. Thus the models agreed that Mix was not potent.

f. The foregoing discussion suggests that CBES and neural nets provide about the same knowledge discovery results. There may be innovative ways to combine them (Rakoff et al. 1991). How to do so is a suitable research issue for the artificial intelligence (AI) community. Given the relative ease of using CBES, it appears to be an appropriate choice for use in this program (T. Dunlap, STATCOM Inc., personal communication, 1991).