CRM D0007581.A1/Final
February 2003

The Practice of Military Experimentation

Brian McCue

4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
FEB 2003		**00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
The Practice of Military Experimentation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **CNA Analysis & Solutions,Center for Naval Analyses ,4825 Mark Center Drive,Alexandria,VA,22311**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**116**	

Approved for distribution: February 2003

H. Dwight Lyons, Jr., Director
Expeditionary Systems and Support Team
Integrated Systems and Operations Division

This document represents the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Department of the Navy.

Approved for Public Release; Distribution Unlimited. Specific authority: N00014-00-D-0700.
For copies of this document call: CNA Document Control and Distribution Section at 703-824-2123

Contents

Preface

Because it especially deals with practical matters, the present document is strongly based on the author's own experience of 4 years at the Marine Corps Warfighting Laboratory (MCWL). To cite some negative examples, as I do, may appear to be criticism, but it is not: part of the Lab's mission was to learn how to do military experiments, and the examples—positive and negative—are selected because they are instructive.

An *intellectual* is somebody who is *excited by ideas.* In the course of my 4-year assignment at MCWL, I met and spoke with Marines of every rank (unless I missed one of the levels of warrant officer), from newly joined Privates to 4-star Generals. One thing I noticed was that *every Marine is an intellectual.*

Introduction

This paper is part of CNA's project on military experimentation. The project's products are:

- *The Art of Military Experimentation*
- *The Practice of Military Experimentation*, and
- *Wotan's Workshop: Military Experiments Before the Second World War.*

The different products are intended to serve different readers' purposes. The military officer (active duty or otherwise) newly assigned to an organization devoted to military experimentation is advised to start by reading *The Practice of Military Experimentation*. The newly assigned civilian analyst might want to start with *The Art of Military Experimentation*. Either should then read *Wotan's Workshop*, and then the other's starting point.

A separate effort has resulted in an additional product,

- *Analysis Planning for a Domestic Weapon-of-Mass-Destruction Exercise.*

A reader with so strong an interest in the topic as to read all the documents will note some commonality among them, especially in the early sections.

The key to the practice of military experimentation

The key to the practice of military experimentation is that, contrary to outward appearances, *an experiment is not an exercise.* This point was noted as early as 1946, when Morse and Kimball wrote,

> This idea of operational experiments, performed primarily not for training but for obtaining a quantitative insight into the operation itself, is a new one and is capable of important results. Properly implemented, it should make it possible

> for the military forces of a country to keep abreast of new technical developments during peace, rather than have to waste lives and energy catching up after the next war has begun. Such operational experiments are of no use whatever if they are dealt with as ordinary tactical exercises, however, and they must be planned and observed by trained scientists as valid *scientific experiments*. Here, then, is an important and useful role for operations research for the armed forces in peacetime.[1]

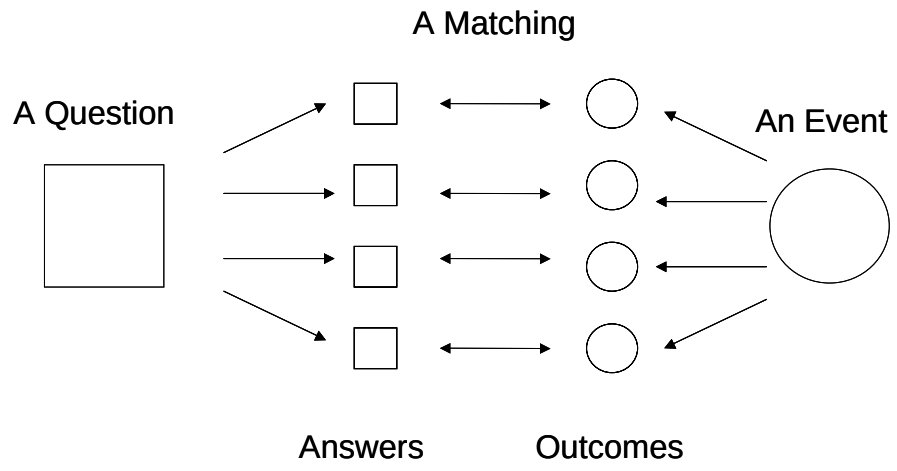
It is no less true today.

Experiments

As shown in figure 1, an experiment consists of

- An *event* that can have multiple outcomes,
- A *question* that could have multiple answers, and
- A *matching*—almost always *pre-stated*—between the outcomes of the event and the answers to the question.

Figure 1. Schema of an experiment



1. Morse and Kimball, page 129.

A familiar example is the use of litmus paper to test the pH of a sample. The *event* is that the litmus paper is dipped into the sample and turns color. The multiple outcomes are that it can turn either of two colors. The *question* is, "Is the sample an acid or a base?" The *pre-stated matching* is that the color red indicates an acid whereas the color blue indicates a base.

Note that this account of experimentation does not require an experiment to have a hypothesis, a control group, a statistically valid number of trials, quantitative measurements, or any of a number of trappings sometimes associated with experiments. An experiment *may* have some or all these things, but if it does, they are part of the definition of the set of outcomes, and the matching of the outcomes to the answers.

What makes the practices of military experimentation[2] so difficult is that a large number of real-world influences act on the experiment, preventing the experimenter from doing exactly what he or she would like. Therefore the problem must be worked from both ends: the experiment must be designed to fit the question, but the question may also have to be adjusted so as to fit the experiment.

In this process, two important traits must be retained:

- There are multiple possible outcomes, not just a single outcome that is guaranteed to happen.
- There is a matching between event outcomes and answers to the question, and normally it is pre-assigned.

If there is only one outcome, or if there are multiple outcomes but they are indistinguishable, the event is a *demonstration*, not an experiment. If the meaning of the outcome is determined only after the experiment is over, then it is an *exploration*, and very possibly not an experiment.

---

2. And, probably, in most other kinds as well, including all but the most "scientific" *and* well-funded.

Experiments in contrast to tests

Experiments, even the test-like limited technical assessments (LTAs) described in the next chapter, differ from tests.

In a test, a strict protocol is followed, almost no matter what, whereas in an experiment, the goal is to obtain knowledge and the experiment can be adapted somewhat if events warrant.

For example, one MCWL LTA addressed the accuracy of a precision targeting system (PTS).[3] After a few attempted uses, it was clear to all involved that something was very wrong with the piece of gear. In a true test, no change would have been allowable, and the test would have continued as scheduled, and concluded that the system didn't work at all. The LTA being an experiment, by contrast, the on-scene experiment team decided to turn the system over to the on-scene person sent by the manufacturer; in a few minutes, the piece of gear was working and by the end of the experiment, sufficient data had been gathered that the system's accuracy could be assessed.

Serendipity

An experiment can lead to an unexpected discovery. For example, MCWL's Urban Warrior experiments led to the discovery that the urban tactics being taught to Marines had basic flaws; this unexpected finding was more important than any other finding of Urban Warrior, and arguably more important than all of them put together. In another example, some of MCWL's experiments yielded unexpected discoveries so compelling as to obscure the points that the experiments were supposed to elucidate, and the serendipitous findings came to be their main product.

3. PTSs find the location of a target by measuring the range to the target with a laser, measuring the bearing to the target with a compass, and feeding this information to a device that knows the sight's own location and does the resulting trigonometry to find the position of the target.

From that idea, it was but a short step to the idea that experimentation didn't really require "all that hypothesis stuff," and that simply fielding the ingredients of an experiment would suffice, because serendipitous findings would result. This notion is pernicious, and is accordingly discussed in the chapter, "Obstacles to Effective Military Experimentation," below, under the heading "'False Serendipity'."

A final problem with serendipity is that it is possible to discover something, and yet not know what it is. During one major MCWL experiment, a new event—radically different from those in the rest of the experiment—was inserted during execution, mostly to mollify a senior advisor. This event turned out differently from what almost everybody had expected, but we couldn't tell why, or what to make of it.

Question for discussion

One day a stranger came to town, and said he was a doctor. The townspeople wanted to believe him, because they needed a doctor in town. But they wanted to be sure he really was a doctor, so they said to him, "Prove you are a doctor: give this dog an appendectomy."

He did so, was welcomed to the town, and gave many years of good service; the story of how he was hired spread far and wide. Eventually, though, he joined the Gold Rush, and the town again needed a doctor. Another stranger came to town, and said "I'm a doctor, and I'll prove it: watch me give this dog an appendectomy." They rode him out of town on a rail.

What was the difference between the two men's claims?

Why military experiments are needed

One may well ask, "why do we suddenly need to do military experiments when we got along just fine for so long without doing any?" Several answers are possible:

- Military experiments are no longer new: the CNA Occasional Paper *Wotan's Workshop* describes some experimentation that took place before the Second World War.[4] Morse and Kimball devote the seventh chapter of their seminal 1946 book, *Methods of Operations Research*, to military experiments. Experimentation was regularly a part of Fleet Exercises through the 1960s, though it fell off thereafter.[5]
- Normally, wars have been frequent and military change has been slow, so that the characteristics of the next war could readily be anticipated by considering the previous war. Presently, as was the case before WW II, huge technological change has occurred since the last war, so experimentation is needed to understand how the next war will work. As Secretary of Defense Donald Rumsfeld wrote in the Quadrennial Defense Review of September 2001, "Exploiting the revolution in military affairs requires not only technological innovation, but also development of operational concepts, undertaking

4. Experimentation, *and the documentation of results*, is ancient: the first chapter of the Book of Daniel describes a dietary experiment, complete with a hypothesis and a control group. And when Saul equipped David with armor, helmet, a coat of mail, and a sword, David complained of a lack of experimentation ("I cannot go with these; for I have not tested them"), and went to fight Goliath with his trusty sling. (First Book of Samuel, 17:38-39.)

5. Ervin Kapos, "The Rise and Decline of Fleet Operations Analysis: Exercise and Real World," presentation at NPS Tactics Symposium, 30 May 2000.

organizational adaptations, and training and experimentation to transform a country's military forces."[6]

- Critics of warfighting experimentation sometimes say that the great battles of history were won not by new technology or even clever tactics, but by courage, training, and tenacity. This is probably true, but the great battles of history (e.g., Waterloo, Gettysburg, Belleau Wood) are implicitly defined as those that were "a near-run thing" and hugely productive of casualties. Better to fight and win battles that are remembered as walk-overs: Crécy (1346, in which infantry armed with the newly-perfected longbow proved capable of defeating heavily armored mounted knights, though the French learned the wrong lesson and dismounted their knights, only to have them again defeated by longbowmen, at Agincourt, 1415),[7] Quebec (1759, in which General Wolfe won by the novel tactic of sending his troops up a cliff), the various German blitzkrieg and U-boat victories early in the Second World War, or Desert Storm.

- Maybe we didn't "get along just fine" heretofore. For example, the trench warfare of the First World War, and the ineffectiveness of the battlecruiser, the obsolescence of the battleship in the Second World War, the necessity of trans-Atlantic convoys in each World War, and the effectiveness of guerilla tactics in Viet Nam were surprises that we would have been better off without.

Less polemic points are worth considering as well: experimentation can be useful in the establishment of cause-and-effect relationships, in the investigation of major innovations, and as an adjunct to the study of military history.

Establishment of cause and effect relationships

Richard Kass states that only experimentation can establish relationships of cause and effect.[8]

---

6. Rumsfeld, page 6.
7. Brodie and Brodie, pp 39-40.
8. Kass.

His case might be illustrated by the Battle of Lissa (1866), in which the Austrian fleet defeated that of the Italians.[9] Naval experts at the time attributed the success of the Austrians to their use of the ram, and ram-bowed ships (such as those of the U.S. Navy's "White Fleet") accordingly remained in vogue. Naval historians now attribute the success of the Austrians to superior gunnery and leadership, but in a very real sense we will never know, because history does not allow us to find out what the battle would have been like without rams or with equal leadership on each side. Experimentation, were it to be practicable (and if anybody still cared), could hope to provide an indication of *cause and effect*—battles could be fought with and without rams, and with and without major differences in gunnery and leadership, and examination of the outcomes could point to one factor as a cause, or perhaps at least *allow a factor to be ruled out.*

It is important to keep in mind, however, that from a strictly logical standpoint, causality can never be *proven*, only *inferred.* Iteration can improve the inference's accuracy.

Investigation of major innovations

Judgment-based military decision-making works best when it has a strong basis in experience. Almost by definition, there can be no strong basis in real-world experience if the question at hand regards major innovation. Today's standard military equipment was yesterday's innovation, and last week's hare-brained scheme. The tank, the airplane, the radio, and the machinegun were each, in their infancy, decried as useless, and yet today they are deemed essential. The rigid airship, the battlecruiser, and the tank destroyer were supposed to be great ideas, and yet they are now remembered for their disappointing results.

Just as important, even the partisans of the successful innovations did not necessarily recognize how best to employ them. The Imperial Japanese Navy kept trying to operate submarines in support of surface vessels, whereas other navies had determined before the war that this

9. Brodie, *Guide,* pages 251 and 279, and *Seapower,* pp 86-88.

idea was unworkable and found alternate ways of employing their subs. The Germans, as we will see, resorted to experimentation to validate Admiral Dönitz's radical "wolf pack" idea.

It is important to remember that "innovation" need not mean, "technological invention": the practitioners of blitzkrieg had tanks that were no better, technologically, than those they faced, but they operated them in an innovative way.

We are fortunate to live in a period of relative peace, and in a period of rapid technological progress, but these positive aspects combine disadvantageously: we are faced with an enormous number of proposed innovations, technological and otherwise, and lack the ability to decide about them based on judgment alone. Almost any alternative will fit some definition, perhaps very broad, of "military experimentation."

An adjunct to history

Study of historical battles and campaigns has long been a staple of military education, and rightly so.

There exists a major split between those who investigate military matters via the study of history and those who do so via models.[10] The most adamant supporters of the historical approach are generally quite critical of models, with much of their criticism based on the unrealism and inaccuracies inherent in the latter. Yet surely the machineguns and tanks of the models, however mis-specified they might be, correspond more closely to the real thing than do the slings and elephants of the historical cases.

---

10. An exception showcases the rule: the late Trevor N. Dupuy (who also had a large amount of personal combat experience, even by the standards of his generation) was one of the very few who have attempted both approaches. His work, sometimes airily dismissed as "trying to predict the past," has never been accepted by either camp. See also McCue, "A Chessboard Model of the Battle of the Atlantic," for more on history and modeling.

A limitation of the study of historical battles is illustrated by a comment made by a reader of a book that described the Battle of Midway as having contained some improbable events: "But I read it six times, and the battle turned out the same way every time!"

Less apocryphally, RADM Bradley Fiske made the same point, in the context of wargames:[11]

> War games and war problems have not yet been accepted by some; for some regard them as games pure and simple and as academic, theoretical, and unpractical. It may be admitted that they are academic and theoretical; but so is the science of gunnery, and so is the science of navigation. In some ways, however, the lessons of the game-board are better guides to future work than "practical" and actual happenings of single battles; for in single battles everything is possible, and some things happen that were highly improbable and were really the result of accident. … The game calls our attention to the influence of chance in war, and to the desirability of our recognizing that influence and endeavoring to eliminate it, when reasoning out the desirability or undesirability of a certain weapon or a certain method. … The partial advantage of the game-board over the occurrences of actual war, for the purpose of studying strategy, lies largely in its ability to permit a [number] of trials very quickly.

Strictly speaking, the historian is hard-pressed to defend rigorously any statement about happenstance or luck, since his material consists only of a set of events, each of which happened exactly once. To refer to luck, or to counterfactual ("what-if") events, the historian must, logically, appeal not only to the historical record, but also to intuition or to reasoning by analogy. Intuition is not rigorous, and historians tend to look askance at the use of analogies, perhaps recognizing that they are tantamount to models.[12]

---

11. Fiske, pages 181–182.

12. For a comprehensive treatment of the logic of historical reading, as well as some remarks highly critical of simulation and gaming, and of counterfactuals in general, see Fischer. For an intriguing use of reasoning by analogy, see Stolfi.

The historian, or other student of history, is then left to think of every event as one-of-kind, shaped by unique circumstances, and improbable. This may be a philosophically correct appreciation of history (an even more philosophically water-tight version would throw out causality altogether), but it makes the "lessons of history" into a set of "Just So" stories, from which no overarching generalizations or underlying truths can be gotten.

Without turning our backs on history, then, we may cast about for alternatives, and find one in experimentation. Because experiments can be repeated, we can, through experimentation, build up a defensible account of what is unusual and what is not. We can also, as noted above, at least begin to discern causal relationships, and therefore to learn what is important and what is not.

The hierarchy of military experiments

Remarkably, the Services' present usage is largely consistent regarding a set terminology for military experiments, though it must be recognized that such terms as "large" and "small" are relative, and a "large" Marine Corps experiment might be the same size as a "small" Army experiment.

A useful policy, in place at MCWL, holds that experimentation on a given topic or piece of equipment should advance through stages of wargame, LTA, LOE, AWE. The wargame is likely to represent an entire military operation, undertaken as of some future time and using many innovations, hardware and otherwise. The game will focus interest on certain points and raise certain questions, usually *not* technical in nature, and in creating a vision of the eventual Advanced Warfighting Exercise (AWE). Innovative technologies to be used in the AWE are then vetted by a series of LTAs, each treating just one technology, or two strongly related technologies. The Limited Objective Experiments (LOEs) then explore particular *topics* in the context of a simulated force-on-force engagement, perhaps using some of the new technologies and perhaps using *surrogates*, with the real emphasis being on the *Tactics, Techniques, and Procedures* (TTPs). The AWE enacts a large portion of the original wargame's operation, in a simulated force-on-force engagement and with the troops, technologies (surrogate and otherwise), and TTPs represented concretely. To these stages, the hierarchy presented here adds the "thought experiment," at the beginning, and experimentation in actual combat, to the end.

Each step is likely to expose imperfections in the question, so one must be ready to revise the question at each stage.

Thought experiments

Strange as it may seem, the experimental set-up can be used purely mentally, with the "experimenter" structuring his thoughts as if for an experiment, and considering each of the possible outcomes in turn.[13] The envisioned experiment may not even be practicable to execute. This style of thinking, termed a "thought experiment" and made famous by Einstein, in fact goes back to ancient times—thought experiments abound in Plato's writings. It usually works by revealing a contradiction that was not earlier apparent.

The following is an example of a military thought experiment.

Sun Tzu wrote, in his famous treatise, *The Art of War,* "Know the enemy and know yourself; in a hundred battles you will never be in peril. When you are ignorant of the enemy but know yourself, your chances of winning or losing are equal. If ignorant both of your enemy and of yourself, you are certain in every battle to be in peril."[14]

This statement certainly seems plausible: knowledge of one's own side and/or the enemy's are good things, and ought to help in battle. But let us perform a thought-experiment to see if this statement can be true. Suppose that two generals fight each other. If each knows himself (i.e., understands his own side) and not the enemy, they are evenly matched, at least in this regard, and one can imagine that each would then have a 50 percent chance of winning, consistent with the second sentence of the quote. If each knows the enemy and himself, they cannot both win, but the aphorism doesn't say they will—it says they will not be imperiled, and maybe the reason is that each will know enough not to fight. But if each is ignorant of his enemy and himself, the aphorism says that each will lose, which is clearly impossible.

Thus, after the thought-experiment, it is clear that Sun Tzu's statement cannot be completely true—something that is not obvious to most people before the experiment.

---

13. Sorenson treats thought experiments extensively.

14. Sun Tzu, Chapter III, "verses" 31-33 (page 84 of Griffith edition).

A thought-experiment helped shed some light on a problem once proposed at MCWL. Interest centered on Unmanned Air Vehicles' search for rare, high-value targets such as SS-1 SCUD mobile missile transporter-erector launchers. Because of these targets' rarity, any given UAV mission would be lucky to find even one. A question was proposed for experimentation: "Would the search effort be aided if the UAV's mission plan could be altered in flight?" At first, the answer seemed self-evident: "of course—any additional capability will help, and the only question is if it will help enough to be worth the effort." One analyst, however, disagreed: he had read in a book on search that "... a well-planned search cannot be improved by a redistribution of search made at an intermediate stage of the operation in an attempt to make use of the fact that up to that time the target had not been observed,"[15] i.e., in-flight re-planning can't help. The book, however, reached this conclusion mathematically, and the analyst wanted a more accessible line of reasoning. He resorted to a thought experiment. In it, a UAV was an hour into a so-far fruitless SCUD-hunt and the mission planner, given the ability to alter the mission plan in mid-flight, had re-optimized the remainder of the mission accordingly. The analyst asked the mission planner, "So why didn't you plan to fly the mission this way in the first place?"

Thought-experiments are, in one sense, a special case; normally, one must undertake a physical event of some kind to have an experiment, and the something-for-nothing deal offered by the thought experiment is a rare exception.

In another sense, however, thought-experiments are quite common. Indeed, it can be said that every experiment starts, or ought to start, as a thought-experiment. The experimenter considers the proposed question and its answers, the proposed event and its possible outcomes, and the proposed matching. He or she mentally reviews the identified outcomes to check whether or not they exhaust the possibilities, and to check that each outcome really is possible. Then he or she considers the answers to which the resulting outcomes point. If they all point to the same answer, or to answers that will result in the

15. Koopman, page 151.

same conclusion or course of action, then the experiment need not be physically performed.

If the possible outcomes do indeed point to multiple and distinctly different answers, then the experimenter has designed an experiment, but further thought should be given to refining it. A good guide is the question, "*What do I want to be able to say when the experiment is over, and what needs to happen for me to be able to say it with conviction?*"

Wargames and simulations

The term "wargame" carries a large number of meanings, ranging from a seminar-type decision-making game to a tabletop or map exercise, or to a computer-assisted version of the same. Many of the issues and considerations regarding wargames have not changed since the publications of the books by Allen and Perla (see bibliography), or even since the much-earlier book by Wilson, and the reader who is totally unfamiliar with wargames would do well to refer to these works.

Computerized or not, many wargames aspire to create realism "constructively," i.e., by starting with a detailed, physics-like knowledge of the speeds, ranges, and other capabilities of the people, weapons, and platforms involved, and then combining these to create a model of their interactions. The players (if any) then give orders, combat is joined, and a supposedly realistic outcome eventuates.

The process described in the previous paragraph is, however, fraught with difficulty and does not necessarily produce valid results. In some applications it has been noticed that all of the realistic details are really tangential to the benefit of the game, which is to instigate and capture the decision-making and the discussion that goes into it. This line of reasoning has resulted in the "seminar" or "course of action" wargame, in which the players of each side separately convene, caucus to consider their options, reach a decision, and communicate them to the game's officials. The officials render a judgment as to the upshot of the two sides' actions, and report to each side whatever information it would realistically get regarding this result. This process might be iterated a few times, completing the game.

In Marine Corps parlance, the term "wargame" is often used to refer to the limiting case of such a seminar wargame, which is really just a systematic discussion of alternatives.

The seminar wargame—with its judgment-based approach to creating the consequences of the players' actions—is often chosen as the means of addressing the decade after next, or other situations about which little is known, on the grounds that the factual basis for creating a detailed simulation is not available. Paradoxically, however, the seminar method may be the *least* effective means of wargaming the unknown: however difficult it would be to create a detailed simulation of the future 15 or 20 years hence, it is even harder to run a judgment-based game about it.

The term "simulation" usually refers to a wargame in which great attention has been devoted to the faithful replication of equipment performance. One such simulation is the Joint Conflict and Tactical Simulation (JCATS), designed for battalion-level battle staff training. JCATS uses computers to administer a real-time map game, in which the units (to include individual infantrymen) move and shoot at realistic rates and—perhaps most important—each map reflects only what the commander of a particular unit would see, based on the disposition of his men amid the terrain.

LTAs

A limited technical assessment (LTA) has many points in common with a traditional "field test":

- The focus is one or more pieces of equipment.
- There is no opposing force (OpFor) or scenario, merely the use of the equipment in a controlled, repetitive way.
- The goal is to see if the equipment works, or how well it works. In the latter case, the answer is likely to be quantitative, e.g., a CEP or a probability of hit.

However, there are some important points of difference as well. In a traditional test, the personnel are likely to be intimately familiar with the equipment, whereas in an LTA, the personnel are usually Service-

people who have just been through a day or two of training. Their performance is likely to be more similar to that of the actual users than would be the performance of professional testers or the people who built the equipment.[16]

However, the biggest difference is in the conduct of the experiment. In a traditional test, the goal is to conduct a fair evaluation. To ensure fairness, the test will proceed in a pre-determined way almost regardless of how it is going, with the only exceptions being safety-related. In an LTA, the goal is to learn as much as possible, and if the test article fails in each of the first 15 attempts, there is no point in putting it through another 85: the LTA will be halted, something will be changed, and then the LTA will resume.

MCWL LTAs typically involve about a dozen people for a few days.

LOEs

A limited objective experiment differs from an LTA in that it has a scenario, an OpFor, and at least a little opportunity for "free play," i.e., decision-making on the part of the participants.

LOEs, especially those devoted to testing technologies, can include numerous sub-experiments. Sometimes, however, it can be difficult to disentangle the sub-experiments from one another—for example, if the Blufor's performance improves when they are given a number of futuristic technologies, which of the technologies made the difference?

MCWL LOEs show great variation in size: the "Blackhawk Down" LOE involved only about a few dozen Marines and one civilian for a few days, whereas Capable Warrior LOE 6 involved well over a hundred Marines and as many civilians, and took a month.[17]

---

16. Herman Kahn cites an extreme example of this, in which the German testing of an anti-aircraft gun showed that one in four rounds might be expected to hit; the wartime average was one in 5,000. Kahn ascribes the difference in large part to the test personnel, whom he characterized as "athletes with Ph.D.s in physics." See also McCue, *Wotan's Workshop.*

AWEs

An AWE is a large experiment that is in principle "complete" in the sense of involving everything that would be needed in a real operation. In practice, such completeness is probably unattainable, but an AWE should at least address all major areas such as (to consider a Marine Corps example) force-on-force ground combat, Close Air Support, fires (to include Naval Surface Fire Support), logistics at least at the Combat Service Support level, and Command, Control, Communications, and Intelligence, and involve at least a whole company of ground troops.

Some have decided that AWEs are inevitably so mired in VIP and media considerations that no actual experimentation is possible. This is certainly a very real risk, but it is not inevitable: certainly the Warfighting Laboratory's Hunter Warrior AWE (March 1997) managed to include actual experimentation.

Experimentation in real-world operations

The mostly likely area of experimentation during real-world operations[18] is in electronic countermeasures; indeed, at some level electronic warfare is characterized by constant experimentation—"OK, so now let's try this."

As early as 1946, Morse and Kimball turned to electronic warfare as a vehicle for discussion of methods of experimentation during hostile operations.[19] The present author lacking experience in this area, the topic is mentioned primarily for completeness, though it is

17. See the MCWL archives for reports on these experiments.

18. The topic of this section does not include experimentation, e.g., with new equipment, by operational units during training that they undertake while deployed. That topic is addressed in a later chapter: the present section refers to experimentation in the context of real-world operations involving interaction with non-cooperative—if not hostile—forces other than one's own.

19. Morse and Kimball, page 98 and following.

interesting to note that Morse and Kimball explicitly emphasize the need for pre-stated criteria by which to judge outcomes.

Components of a military experiment

Certain components are almost always present in military experiments. To set the stage for later discussion, these will be defined here. Upon close reading, the definitions embody a considerable number of assumptions regarding the structure of experiments. For example, there would not really have to be an "experimental force" and an "opposing force"—an experiment could be designed with experimental tactics or equipment on each side, and the fact that it is not usually done that way is an important observation about military experimentation.[20]

This section will also recount some of the basic findings of the MCWL experience regarding these ingredients of an experiment.

Questions, ideas, and hypotheses

The biggest difficulty with questions, ideas, and hypotheses is finding sources of good ones.

A common source of new ideas is new technology: a new device, or a new idea for one, comes along, and a military experiment is designed around it. The ease with which technologies lead to ideas for experiments has made technologies the most common inspiration for experiments, and has also made many people think that technology is the only possible focus of military experimentation. This view is strongly associated with critics of experimentation, and insofar as their message is that there is much more to military success than technology, they are doubtless correct.

---

20. See also *Wotan's Workshop*: U.S. interwar experimentation sometimes pitted a traditionally equipped "Blue" force against an innovative adversary.

An interesting alternative source of new ideas is *existing* technology. Admiral Dönitz availed himself of this source of ideas when he realized that his First World War submarine tactics had been conceptualized as if for perfect submarines, but were being implemented with decidedly imperfect ones. His idea of the wolfpack came from asking himself how to use craft which, though fast, stealthy, and deadly, could at any one time be only fast *or* deadly, and had nearly as much trouble seeing other vessels as other vessels had in seeing them.[21]

Another source of ideas is the Services other than one's own. The Experimental Combat Operations Center with which MCWL experimented borrowed ideas heavily from the Combat Information Centers found aboard U.S. Navy ships.

Ideas can also come from outside sources such as law enforcement, from the Services of foreign countries, or even from science fiction, or the behavior of animals. They can come from wargames or the agent-based computer "worlds" developed by Andrew Ilachinski and others.

Finally, one can generate ideas based on the results of earlier experiments. For example, MCWL's 2-year Urban Warrior series of experiments suggested that the Military Operations in Urban Terrain (MOUT) tactics being taught to Marines were grossly sub-optimal in several respects. A new set of tactics was developed, and then a new training package to inculcate those tactics, and then experiments were done to see if the training was effective, and if the new tactics were better than the old ones.

With an idea firmly in mind, it is then easy to think of questions:

- Will the idea work?
- How well?
- Is it better than what we have?

These questions can be rephrased into declarative hypotheses if desired.

---

21. See *Wotan's Workshop*.

Again, it can be useful to ask one's self, "What do I want to be able to say when this is all over?"

After it is all over, one is not allowed to change the goal(s) of the experiment to match what has taken place and hope to be taken any more seriously than would any commander, teacher, coach, or politician who revised a statement regarding goals so as to match what had occurred.

It is best to avoid asking the users simple, isolated questions, such as, "does this new piece of gear help?" or "is it too heavy?" Almost anything will help, and to the person who must carry it, almost anything is too heavy: the important question is whether the help will be worth the costs, burdens, risks, etc. The end user may or may not be able to make that assessment.

Finally, one should avoid having one's experiment be driven by the question of whether the individual participants liked the equipment or, worse, the unit liked the experiment. This point is addressed below, in the chapter entitled, "Obstacles to successful experimentation."

Who's who in a military experiment

Most experiments include most of the following groups of people. Some may lack an opposing force, and many will lack the civilian role-players.

Experimental force

The experimental force is the force that is using the experimental equipment (or surrogates for it), or tactics, doctrine, etc. It is often abbreviated "ExFor." In most American experiments, that is the American side, so it is often called the Blue Force, or "BluFor," but it is worth noting that one need not necessarily equate the American side with the experimental side: in some of the pre-World War II experiments recounted in *Wotan's Workshop* the Blue side was the Americans, and the other side had the new equipment and tactics.

MCWL policy rightly considers it important to do LTAs with representative Marines as the equipment users. These Marines could therefore be called the experimental force, but in practice the term was not used in experiments that did not have human "players" on both sides.

Opposing force

The opposing force is the force that is not the experimental force. Typically it would be configured to represent an adversary that American forces might encounter in a real-world situation. The name is sometimes abbreviated "OpFor." The OpFor are critical to the success of the experiment, so correct choice and management of OpFor are in turn critical as well.

In MCWL's Urban Warrior series of experiments, there was a desire to create an OpFor that represented the low-grade infantry supposedly expected to be encountered in real-world contingencies. For this reason, USMC combat engineers were chosen. They turned out not to be as inept as had been hoped, because:

- Usually cast in the role of defenders, they could improvisationally bring their engineering skills into play, creating formidable obstacles in and around the buildings;
- The same unit was used for the first four experiments, held at Camp Lejeune, and they learned, whereas they were always operating against Blue forces that were in an experiment for their first time; and
- Perhaps engineers aren't such bad infantry after all.

One problem with the opposing force is that in a realistic land-war scenario, the U.S. force is likely to be outnumbered, but few organizations will be willing to have the majority of their strength assigned

to the opposing force. Various approaches can be used to solve this problem, notably "recycling" "dead" opposing force troops[22] and defining some regions as impassable by virtue of being held by constructive opposition troops.

Most OpFor will pursue their duties with vigor *unless* they become convinced that their defeat is a foregone conclusion.

Civilian roleplayers

While, in a sense, all the participants (even the Blue Force) are "roleplayers," the term was generally reserved for people who played the role of civilians. The scenario would call for these "civilians" to perform certain acts, such as lining up for food at an aid station, or moving about as if on their daily business. The ExFor and OpFor could order the Roleplayers around.

Particular roleplayers can be assigned specific roles, e.g., that of a leader, and the masses can be divided into factions with various leanings. In some experiments, MCWL took this idea one step farther and allowed the roleplayers to change their minds in response to actions taken by the ExFor and OpFor during the experiment. A poll after the experiment, compared to the pre-assigned leanings, therefore became a good way of measuring the effect of the two sides' actions on these third parties.

MCWL variously used Marines, family members of Naval Postgraduate School students, and Hollywood "extras" as roleplayers. We were pleasantly surprised to find that, no matter what the source, the roleplayers tended to take their task seriously, and that the occasional

22. This is not as easy as it sounds, and is usually done poorly. The total constructive size of the OpFor is usually calculated as its original size, plus the number "killed" additional times on the grounds that these recycled. But this calculation omits the (often substantial) number of OpFor who, as of the end of the scenario, are back in after having been "killed" and re-entered. Moreover, recycling does not really create a larger force, just a longer-lasting one: an attack by 200 men who can die twice, with a time-consuming re-cycling process in between, is different from an attack of 400 men all at once.

instance of ham acting added more than it detracted. The use of Marines as roleplayers though, did suffer from some drawbacks: too few females, too little variety in appearance, and too much aggressiveness in confronting the armed forces. The Hollywood extras were considered to be a great success, and certainly cost-effective, and only a few were unwilling to put up with the hardship sometimes expected of them.

Experiment control

Experiment control, or ExCon, would look quite familiar to those accustomed to administering field exercises. Working from a master scenario events list (MSEL, the individual entries of which are inevitably, if nonsensically, also referred to as MSELs, pronounced "measles") of scheduled inputs, and at times using optional "injects" to push the course of events onto a desired path, Excon controls the experiment as it unfolds.

Other forces, notably the actions of the players, may also control unfolding of events. An important aspect of experiment design is to decide the degree to which they are going to be allowed to do so. This point is addressed in a later section, under the heading, "Scripting."

Observer-controllers and "firewalkers"

Data collection is of primary importance. Although the participants in MCWL's experiments were responsible for collecting a good deal of data via forms that they filled out after the event, their activities during the event had to be tracked and recorded. In most of Urban Warrior, this was done by hand, by Marines tasked with helping to control the experiment as well as to observe it; they were termed Observer/Controllers, or "O/Cs." We found that each fireteam, each platoon or company commander, and each vehicle needed an O/C. Thus, something like a fifth of the available manpower needed to be dedicated to data collection. This is a daunting requirement, and occasionally it would be skimped, inevitably leading to shortfalls in data collection.

MCWL analysts also positioned themselves around the battlefield, at the Blue side's Combat Operations Center, and in Experiment Control (see below) taking notes.

Roleplayers were able to collect data on themselves.

When the data-collection manpower requirements were satisfied, I estimated that we could have 90 percent confidence of knowing each fireteam's activity to an accuracy of 50 feet and two minutes. This estimate was impressionistic, not analytic. See also under Instrumentation, below.

The conduct of the experiment required a good deal of "adjudication" and other intervention, so the data collectors (apart from the analysts) were empowered to perform those functions, and thus became "observer controllers." It being difficult to find large numbers of senior Marines without anything better to do than to be observer controllers, there were occasional problems in which a participant would attempt, on the basis of rank, to evade the authority of an observer-controller.

In addition to observer/controllers, MCWL used "firewalkers," a higher form of observer/controller. These were Majors, and responsible for the use of flash-bang artillery simulator devices, at the behest of Exercise Control, to simulate indirect fire artillery, and for the use of the "God-gun" devices capable of setting and re-setting MILES (Modular Integrated Laser Engagement System) gear, and of receiving records from the recording device in the new MILES 2000. They had a secondary duty, seldom required, of intervening if an observer-controller was failing to control a participant because of a disparity in rank.

Analysts

The present author contends that those who will reconstruct and analyze the experiment ought to be present in person. This prescription may seem self-evident, but there are those who see analysis as a purely mechanical process, to which the analyst's own observations can add nothing, or from which his or her observations could even detract by introducing prejudices that would interfere with the objectivity of the

analysis. Therefore some see the analysts' presence as optional, and one group of experimenters even went so far as to (in the name of objectivity) bar the analysts from witnessing the experiment, allowing them only access to post-processed data.

But if prejudice can interfere with analysis, then analysis is not a purely mechanical process, and the issue becomes a matter of whether the analyst's personal observations will enhance or detract. The author's experience has been that while personal observation may have introduced a few prejudices, it also informed the analyst of a host of important facts that would otherwise never have come to light, and the latter far outweighs the former. Use of multiple analysts helps even more, since they tend to disabuse one another of prejudices.

An LTA usually involves only one or two analysts: in a force-on-force LOE or AWE, many analysts will be needed. Each is assigned a topic and must give careful thought to how to observe the action in order to best cover his or her topic. Analysts must recognize the need not to interfere, either actively or even passively, e.g., by becoming observable and thereby calling attention to the position of those whom they are following. On the other hand, they must be given the freedom to go wherever they want, with only their good sense restraining them.

In the Second World War-era dawn of operations research, Winston Churchill said that scientific experts should be "on tap, not on top." The same is true today, but surely the role of the scientist at a command devoted to military experimentation ought to differ from that of a scientist assigned to an operational command. The assignment of a scientist to an operational command indicates a belief that a scientific outlook might help with operational matter. But the creation of a "laboratory" devoted to military experimentation indicates an initial and conscious choice that a scientific approach is to be used, and therefore the scientist should be treated as a native guide, not a foreigner with an important alternative perspective, as might be appropriate at an operational command.

Experiment-unique equipment

In an experiment, the participants will likely use much of what they would use in a real battle. They may also use *experimental* equipment—real, prototype, and/or surrogate—*exercise* equipment such as MILES gear, and *instrumentation* equipment that might well not figure in an exercise.

It is important that the participants understand the nature of such equipment, and the need for it.

Surrogates and prototypes

Surrogates are pieces of equipment that, possibly in conjunction with experiment procedures or "rules," represent or provide the functionality of another piece of equipment.[23]

Sometimes surrogates are used for reasons of safety, for example, the familiar MILES gear allows the participants' service weapons to shoot laser beams rather than bullets, and therefore makes possible force-on-force engagements. An alternative is Simunitions®—9mm paint rounds, fired by a reduced charge.[24] A special-purpose barrel and upper receiver adapts the standard-issue M-16 to fire these; for safety's sake, this adapter is unable to fire standard non-paint 9mm ammunition.

The principle drawback of Simunitions® is their short range, variously estimated at 25 meters or less. They are thus suitable for use only in the replication of urban combat.

However, the need to discuss surrogates here arises from their more central use in experimentation, in which the surrogate represents a piece of equipment that is not yet available, as opposed to one (such as the M-16) which is available, but cannot be used for reasons of safety.

---

23. See also Karppi and McCue.

24. Simunition® is a registered trademark of SNC Technologies Inc.

Often, experimentation begins when new technology is visible on the horizon, but not yet available. For example, the American and German armies began to experiment with tank warfare before they had tanks, but after they had enough insight into what tanks would be like that they could create meaningful surrogates and use them. Admiral Dönitz had a clear idea of what submarines would be like (based on experience in the First World War) but in the mid-1930s Germany had no submarines because of the Versailles treaty.[25] Thus Dönitz's experiments may have used destroyers or torpedo boats as surrogates for submarines.[26]

Similarly, no vehicle yet existed that embodied the ARMVAL experimenters' "helicopter-liftable mobile protected weapons system," so they had to create surrogates from Army M551 Sheridans.[27]

The futuristic command-and-control capability under test in MCWL's Hunter Warrior did not yet exist, so the Lab had to create surrogates using off-the-shelf radios, GPS devices, and palmtop computers.

A frequent problem with surrogates is that people mistake them for prototypes; the better the surrogate, the more likely this is to happen. Thus people may have merely snickered at the wooden antitank guns of the Louisiana maneuvers or the plywood tanks of the German experiments, but they devoted considerable effort to trying to explain why the command and control device used in MCWL's Hunter Warrior was not battleworthy, when in fact it was only an "eclectronic" surrogate made by combining consumer products.

Another problem with surrogates is that some actually work better than the systems they represent. For example, if a remote electro-optical sensor is surrogated by a person with a radio, the performance is likely to be far better than any current or imagined sensor system can provide: the person's visual acuity is likely to be better than the sensor's, and his or her ability to pre-process what is seen, report only

25. Treaty limitations also delayed the introduction of tanks into the German army.

26. These examples are treated at more length in *Wotan's Workshop*.

27. See Thompson.

what is of interest, and respond to spoken queries would be the envy of many an Artificial Intelligence project. It would be quite a technological challenge to make a surrogate that works exactly as well as the system it represents: the solution is to make it work somewhat better, and then limit it via a rule or procedure. For example, a manned helicopter was a surrogate for certain Unmanned Air Vehicle (UAV) in a sequence of MCWL experiments. The helicopter carried the UAV's sensor, but of course it also carried a pilot, and in early experimentation the pilots often helped out by scanning the terrain visually and telling the sensor operator where to point the sensor. They even engaged in dialogue with the ground operators while doing so. Thus the surrogate was vastly more capable than the real UAV would be.[28] Of course, the human pilot has to be present, so the solution is to have a rule that he cannot help out with the sensing process.

Though it is bad to have a surrogate that works better than the system it represents, it is worse to have one that does not work as well: the experimenters can create a rule eliminating unwanted functionality, but they are much harder-pressed make a rule that will restore missing functionality.

Surrogates that are created administratively, such as remote sensors operated by ExCon or direct-fire weapons whose effects are created by adjudication (as opposed to, say MILES or Simunitions®) should be created so as to have some realistic imperfections, even if the exact parameters required to do so are not known. For example:

- In one MCWL experiment, there was no provision for scattering the fire of helicopter-mounted Hellfires and 20mm guns. Consequently, these were always adjudicated as hits. Afterwards, one write-up of the experiment observed that

28. This case provides a good example of the Inverse Surrogate Test, which asks, "If this were a perfect surrogate, what system would it represent?" In this case, the helicopter with the helpful pilot represents a surrogate with automated voice recognition capability and speech response, and a second sensor (almost fully directable, with a field of view of about six steradians, about a half an arc-minute of resolution, and full color capability) that can be used to cue the grainy black-and-white IR camera.

"helicopter-mounted Hellfires and 20mm guns proved remarkably effective."

- The surrogate sensors—people—of another LOE were plentiful and "functioned" admirably, especially in that they returned no false alarms. As a result, the best the LOE could do was to prove that can be helpful to have a large number of perfect sensors. This is not a very informative result. It would have been better to use a more realistic number of sensors, and to assign them—arbitrarily, if necessary—a miss probability and a false alarm rate. Then, if the set-up worked, these parameters could serve as design goals, at least until something better came along; if it didn't work, then analysis could try to determine where the weakness was—in the density of the sensors, their tendency to miss the target, or their propensity to create false alarms. A later experiment could be done with an "improved" sensor, until tolerable parameters were found.

Occasionally one hears the cry, "No more surrogates!" Usually this arises when an experiment has been conducted with surrogates, and a conclusion has been reached, and then it is pointed out that the conclusion applies only to the surrogates. The experiment should have addressed the *concept* under test, instead of devolving into a meaningless *test of the surrogate*. "Experimenting on the surrogates" is addressed further in a later section, in the chapter "Obstacles to successful military experimentation."

Instrumentation

MCWL, with the assistance of SRI International, developed the Integrated GPS Radio System (IGRS, pronounced to rhyme with "tigers") to track experiment participants and vehicles. The device resides in a fanny-pack and uses the Global Positioning System (GPS) to detect the participant's location. A radio system then polls participants' IGRS units in rotation and updates their locations in a central computer. The result is a display that proved to be of great use to ExCon,

and a position database that serves as the basis for replay software that is immensely useful to analysts reconstructing the events of the experiment. [29]

IGRS interacted with the older version of the MILES system, allowing participants' "deaths" by MILES fire to register at ExCon (and in the database) and allowing ExCon to induce "deaths" remotely, simulating the effect of indirect fire. The follow-on MILES 2000 system is not interoperable with IGRS in real time, but it does create a database of its own. SRI software can meld the MILES data into the IGRS database, giving the analysts a replay display that shows participants' status as alive, dead, or wounded, and shows MILES shots that hit or are "near misses."

GPS does not penetrate indoors, but in its absence IGRS units attempt to make ultrasonic contact with small boxes that can be preplaced to "tag" individual rooms. The display software can be set up to include the deck plans of buildings, and to indicate the room (albeit not the location within the room) in which an IGRS is signaling itself to be located. (Though the GPS signal does not propagate into buildings, the IGRS signal propagates out with relative ease.)

Paint-filled Simunition® rounds serve an "instrumentation" purpose as well as fulfilling their role as a surrogate (as discussed below) for bullets: their colors record their shooters' sides (and thus show instances of fratricide), and—unlike MILES—they indicate which portion of the victim's body was hit.

Models and simulations

Computer models can play a role in field (or fleet) experimentation. In fact, they can play two roles. One, obviously, is in the administration of wargames; a computer model can keep track of the entities locations and resolve combat outcomes. The second role, for which JCATS was used extensively by MCWL, is as an adjunct to live force-on-force experimentation: supporting fires, and aircraft in general,

29. See also the Marine Corps Warfighting Laboratory "X-File" on instrumentation.

were done in JCATS and the results transposed to the live entities by ExCon, via a signal sent to their MILES gear, via their IGRS instrumentation.

Methods

In a sense, this whole document addresses methods, but this section will treat particular methods for accomplishing particular experimental goals. A host of additional methods, at a lower level of detail (e.g., how to use dice to make constructive deviations in adjudicated mortar shots) will not be addressed here.[30]

Base case v. experimental case

Experience has shown that if people remember only one thing from school about experiments, it will be the idea variously known as a "base case," "control," or "baseline."

Early MCWL experiments (like most military experiments—see also *Wotan's Workshop*) did not have any such feature, and some people dismissed them out-of-hand on this basis. Yet an experiment need not have a base case, because it need not be engaged in *comparison*: for example, it could be engaged in *measurement* (e.g., of the accuracy of a weapon) instead. The rationale for structuring the early MCWL experiments without a base case was that the hypothesis was that something (a tactic, a set of technologies) would work or not, and in that context a base case was meaningless.

Later MCWL experiments benefited from the presence of a base case.

As a practical matter, it can help to *do the experimental case first.* One reason to do the experimental case first is that if experimental case involves technology, and the technology fails catastrophically, the base case needn't be done at all. Another reason is that the participants will learn during the experiment, and if the experimental case is done second, it is possible that any improvement is ascribable to

30. See also the section entitled "Methodology," in *The Art of Military Experimentation.*

learning and not to the experimental tactics or technologies. If the base case is done second, any learning will act to lessen the apparent improvement caused by the technology or tactics, so if performance in the experimental case is nevertheless superior, a strong case can be made for the technology or tactics. If the base case is done second and performance is superior in it, then any benefit conferred by the experimental equipment or tactics is less than the training benefit of having done one case, and is thus probably negligible.

Resetting

Sometimes, the events in an experiment unfold in such a way that all value may be lost. The most obvious example would be an early defeat of the ExFor at the hands of the OpFor. If this impends—or after it has occurred—there is really no choice but to start the experiment all over and hope that chance or learning will cause the Experimental Force to do better the second time.

However, there are two great dangers in this course of action. The first is that the opposing force will conclude that the experiment will be repeated until they lose, and accordingly decide to exert minimal effort, the better to lose forthwith. The second is that afterwards the instance in which the experimental force was swiftly defeated will be viewed as an aberration that doesn't really count, and that only their ensuing victory on the second try will be remembered or used in analysis.

These dangers can be avoided if it is made clear to one and all that the first try, in which the Experimental Force was defeated, will be treated as no less valid than the second. The Opposing Force—tired after their efforts, and then frustrated upon seeing victory snatched away from them administratively—will need an especially clear, patient, and understanding explanation of this point. In attempting to give such explanations, more than one analyst has resorted to the science fiction concept of "branching time streams," and been surprised by how comfortable the young Marines were with this idea.

However, it is still all too easy to dismiss the disasters, on the grounds that "Excon said that didn't really happen," and all the technologies,

TTPs, and other experimental innovations will check out as having performed well. Therefore the written report of the experiment will have to make very clear that the "road not taken," however disappointing, was at least as valid an outcome of the experiment as the one dictated by Excon—more valid, in fact, because it resulted from force-on-force free play instead of from Excon fiat.

Even if the written report is very clear about the two outcomes, there remains the point that to the observers and participants, the events that unfold on the ground are more real than those they are told would have happened if experimentation had not been halted. These people must be careful not to allow this bias to creep into their briefings and discussions (especially those that are part of the *assessment process*, described below), as these can be at least as influential as the analysts' final report.

Scripting

Scripting of the experiment's scenario (if any—an LTA does not normally have a scenario) must be done with some care so as to set up the desired conditions for the experiment, yet avoid prejudicing the outcome.

Experience has shown that the participants will need explicit guidance not only as to what they must and must not do, but also as to where their own decisions and free actions are required. Otherwise, during the debrief, one is likely to ask why some puzzling action was undertaken, only to be told, "We thought you wanted us to do that."

Adjudication

Not withstanding the availability of MILES, Simunitions®, and the like, the effects of some weapons will be reproducible only through *adjudication*. Adjudication relies on observer/controllers to realize that a weapon is being used, quickly make an assessment of its likely effect, communicate that assessment to the victims, and ensure that they react accordingly.

Adjudication works adequately for short-range weapons such as hand grenades, for which one observer/controller can perform all the steps listed above.

Adjudication breaks down seriously for longer-range weapons, because the controller who is near enough to the weapon to know that it is being used, and against whom, must make radio contact with a controller who is near enough to the victim(s) to impose the effect. The time needed to complete the adjudication is usually long enough that, in the interval between their "deaths" and their notifications thereof, the victims have had time to do something. In the worst case, they have killed somebody with MILES, who in turn will expect to be revived when he finds out that his killer had been supposed to be dead.

Statistics and sample size

Most people have an intuitive understanding that if one is trying to understand a system in which chance plays a role, one ought to make multiple trials. Obvious examples would include the testing of a new weapon for accuracy, or reliability. Only slightly less obviously, the testing of new tactics, or equipment, in force-on-force experimentation requires repeated cases as well: combat outcomes notoriously depend upon chance as well as upon tactics and equipment, and a host of other variables that the experimenters can at least hope to hold constant.

The number of trials needed is called the "sample size." Sample sizes calculated on the basis of "cookbook statistics" can be useful in LTAs that test equipment for basic hardware traits such as for accuracy or reliability. But for force-on-force evolutions (and even in some LTAs), textbook sample sizes will be far in excess of what most experimentation efforts will be willing to undertake: most force-on-force experiments are hard-pressed to attain five iterations, whereas the statistics-book approach will demand many more.[31] In addition, the announced intent to do the same event more than once invariably

31. See, for example, Crow, Davis, and Maxfield, page 52.

inspires planners to think of variations that could be built into the later repetitions, all in the name of experimentation—but undercutting the goal of attaining statistical significance. See also *The Art of Military Experimentation* for how and why a small number of trials may be made to suffice, despite the statistics "cookbook."

It is not normally possible to do all the trials at once, so instead they are done sequentially, leading to the difficulty that they are then not really all the same, because some were done earlier than others. In particular, the participants may (and in fact, almost certainly will), *learn* from one iteration to the next. The solution of using new participants at each stage is seldom possible, and introduces the added complication that the separate groups of participants may differ.

As mentioned above, one can immunize the experiment against anti-base case bias by doing the base case *second*, so that it—and not the experimental case—benefits from any learning that may take place.

Debriefing

Regardless of the data-collecting abilities of the observer-controllers, analysts, and automated data-collection systems, a debriefing of the participants is needed. This debriefing needs to occur immediately after the event (i.e., immediately after the event is concluded and the leaders have accomplished their personnel and equipment accountability checks), because memories will fade rapidly. At that time, the participants will be tired, dirty, probably either too cold or too hot, and probably either hungry or thirsty, or both. Yet the debrief must occur. In some cases, it can beneficially be done during time that would otherwise be spent waiting for transport or the like.

This debriefing should be conducted by the analysts, whose approach will be dictated by the goals of the experiment and their own personal styles. One MCWL analyst found it useful to hold a meeting of all the participants for which he had cognizance (usually a company, or most elements thereof), and then to ask for a volunteer from each squad—*other than the squad leader*—to give his squad's view of what had happened. These little speeches usually led automatically into a useful general discussion. The analyst is well-advised to take notes.

Questionnaires will be necessary as well. Again, the requirements will vary from experiment to experiment, but at least one analyst found it useful to construct his questionnaires so as to be filled out by a squad leader and the observer/controller(s) attached to the squad. Such a questionnaire could be filled out during the debriefing session, each squad's operational summary being given (as mentioned above) by somebody other than the squad leader.

And as another analyst observed, the reverse side of the questionnaire should be left blank, because it may thereby turn out to be more useful than the front.

Accuracy, realism, fidelity, reality, truth, and cheating

Today, "models" of warfare are automatically assumed to be *computer* models. Many people understandably assume computer models of warfare to be of questionable validity despite their impressive graphics, and tend to reject findings based on them. To them, field experiments or fleet experiments are alternatives to modeling, and perhaps attractive for that very reason.

However, it is important to realize that the activities undertaken in the field, at sea, or in the air are themselves warfare models, albeit not resident in a computer. Just like a computer model, this model should be examined critically, and judged on factors other than appearance.

Accuracy and realism can be quite troublesome to build into military experiments. Everybody agrees that more is better, but there is sometimes disagreement on how much is enough. A philosophical split underlies the disagreement between research-oriented analysts and exercise-oriented military people.

Analysts and accuracy

Military personnel tend to look with skepticism on computer models of warfare. One retired Marine officer working at MCWL wrote: "M&S [modeling and simulation] is the black hole of Calcutta, it will consume billions of dollars and produce very little." If they consider the matter at all, MCWL workers tend to see their live experiments as an alternative to modeling. But when used as a basis for analysis, the exercise-like force-on-force warfare (with MILES, Excon, O/Cs, adjudication, and all the rest) is a model of actual warfare as well—just not a *computer* model. As such, it is not reality, and the analysts need to consider, just as they would for a computer model of combat, the level of fidelity with which it reproduces reality. Their scientific training

leads them to think of calculations, and therefore models, as being no more accurate than their least-accurate part: "a chain is as strong as its weakest link."[32] Therefore, efforts to increase accuracy must always be devoted to improving the accuracy of the least accurate portion. When looking at a model—including MCWL's "model" of urban warfare created through the use of MILES, O/Cs, firewalkers, ExCon, JCATS, and all the rest—analysts automatically use the same logic and tend to reject any accuracy-increasing proposals that do not address the least accurate portion of the model.

For example, M-16 engagements were relatively realistically portrayed by MILES gear, and hand grenades were surprisingly well portrayed by blue bodies[33] and an adjudication procedure. Medium and heavy direct fire weapons (e.g., SMAWs (Shoulder-Launched Multipurpose Assault Weapons), M203-launched grenades, and medium and heavy machine guns) however, proved difficult to handle: adjudication of shots with these weapons required (as discussed in the present document, under "Adjudication") the coordination of multiple observer/controllers, and took too long. Analysts' suggestions for improving realism tended to focus on how to improve the adjudication of the medium and heavy direct fire weapons, because of all the shortfalls in realism, that pertaining to medium and heavy weapons was the greatest.

Retirees and realism

The planners of the Lab's experiments have had a background in exercise-planning, and are interested in *realism*. They are roughly consistent, across the different parts of the experiment, in the amount of *trouble* they will tolerate for the sake of realism. There is some level of trouble which, when reached, is "enough," and beyond it no further effort to increase realism is to be made. These planners

32. This account of error propagation is simplistic, but it is often a reasonable guide and in any case the point here is that scientists are inculcated to think in this way, which is certainly the case.

33. "Blue bodies" are practice grenades that are fuzed like the real thing, but make only a firecracker-like bang.

have shown little interest in improving the adjudication of medium and heavy weapons, because any improvement (other than obtaining suitable MILES gear, if any becomes available) would be a great deal of trouble, and they had already gone to enough trouble regarding these weapons.

When applied to *additive* quantities, such an equalization of the "threshold of pain" across alternate endeavors results in a maximized total,[34] so the planners' approach arguably results in *maximum overall realism.*

Thus the analysts and the planners tend to talk past one another when discussing how to set up an experiment: analysts see it as a calculation (albeit an analog one), and hence only as good as its least accurate portion, while the planners see it as an exercise, and hence as good as the sum of its portions' realisms.

Operational fidelity

In a model, however—especially a live-action model such as ours—what one should seek to maximize may be neither the analysts' "accuracy" nor the exercise planners' "realism." The point is not so much the words as the maximization processes (weak link v. threshold of pain) with which they are associated: neither is appropriate to experimentation.

34. This point, though not quite intuitive, is a staple of freshman microeconomics classes. It assumes "diminishing [marginal] returns to scale," which is almost always a reasonable assumption. Example: a farmer has two cornfields, one with better soil than the other, and needs to allocate his available irrigation water between them. The correct allocation (i.e., the one that gives the biggest total harvest) is the one that equalizes the increase in value attributable to the last day's watering of each field.

Instead, I propose "operational fidelity." In this phrase, "fidelity" is used as in connection with home stereo systems, and "operational" is used to mean "defined in terms of feasible actions and measurements,"[35] not in the military sense of the word.

The creators of the distributed interactive simulation originally known as SIMNET referred to "selective fidelity" in describing how they decided what to put into the model (including, but not limited to, its video-game-like human interface). For example, sound effects (including subsonic vibrations) are relatively inexpensive, yet give people a strong feeling as to their surroundings: the sound effects in SIMNET are accordingly well developed. But they did not elect to concentrate on sound simply because it was cheap: they kept in view the goal of maximizing the extent to which the people in the model did what they would do in real life, and found that sound was a "best buy" in terms of evoking correct behavior on a limited budget. "Proportion of correct behavior" was thus the SIMNET developers' standard in deciding which aspects of fidelity were worth seeking and which were not.[36]

What standard ought to be used in military experimentation? Correct behavior on the part of the experiment participants is nice to have, but because we are interested in experimentation, correct behavior is not the bottom line for us that it was for the training-oriented SIMNET developers. For us, the gauge of fidelity is, or should be, *the degree to which the connections between the experimental outcomes and the answers to the experimental question are preserved.* Physical scientists routinely, perhaps even unconsciously, apply this rule: in designing a laboratory experiment, for example, they might know that they must pay great attention to whether the table is level, but that it doesn't usually[37] matter what the table is made out of.

---

35. CNA's institutional forebears, the early operations researchers, used this sense of the term, e.g., in defining "operational search rate," doubtless because of their training in the logic of modern physics, as expounded by Percy Bridgman. See also Morse and Kimball, *Methods of Operations Research.*

36. Voss, pages 5 and 17.

37. For an exception, see Rhodes, pp. 217-218.

Selective reality

Results of MCWL experiments have oftentimes been dismissed on the grounds that the experiment wasn't "real." But not all of the experiment needs to be real, only certain parts.

In an aviation-related LTA, we measured the CEP of a candidate weapon in actual drops on an urban-like target array. The question arose as to what CEP would be good enough, and an analyst cited earlier MOUT experiments (described in more detail in the course of another example, below) in which CEPs on the order of 100 meters had proven inadequate, while CEPs of 1-3 meters were satisfactory. The person responsible for the aviation LTA responded, "But those LOEs weren't real." What he saw as unreal about those LOEs was the application of fires—in the LOEs, fires were of necessity adjudicated (because there were live players on both sides), whereas in the LTA, weapons (albeit inert ones) were being dropped from airplanes. But in showing what was needed in terms of accuracy, the fidelity belonged in the ground combat: what needed to be real, or at least realistic, was the situations in which fire was called, and the distribution of the resulting simulated impacts on the battlefield. The experiment had to be real, or at least realistic, in those places where it needed to be in order to answer the question, and not necessarily in others.

So, in order to decide if one's experiment is sufficiently realistic, one must first know what question it is supposed to answer, and the rationale for associating the various possible experimental outcomes with the various possible answers to the question. This knowledge entails the application of a theory. MCWL's Hunter Warrior experiment, for example, dealt with a proposed style of warfare in which supporting

fires (to include CAS) took on predominant importance. Accordingly, the Hunter Warrior experiment was designed around observation and fires-calling, with little provision for direct-fire small-arms engagements.[38]

Sometimes there is a role for "gratuitous reality." For example, one experiment tested the concept of having a large number of sensors deployed as a "cloud" to detect the movement of critical mobile targets such as Scuds. Originally the sensors were to be surrogated by the scientists who were developing them, the most realistic means possible of having surrogate sensors that would act the way the real ones will. But when the scientists couldn't come, some of the sensors were surrogated by Marines and most were simply played in ExCon. Under these circumstances, it seemed odd to persist in the use of real vehicles (mostly rental vans marked with recognition panels): given that most of the sightings would be made by ExCon-surrogated sensors (and ExCon could have taken over the few remaining Marine-surrogated sensors, so that all sensor reports were really from ExCon), why have the vehicles at all? Why not just have ExCon move pennies around on a map and call in sightings accordingly? Although the experiment certainly could have been done on that basis, we noticed a benefit to the use of the real vans—they provided an inarguable "ground truth" in a way that pennies could not have.

"Gratuitous reality" can also pay off by providing pieces of background realism that turn out to be needed to assure the validity a serendipitous result.

Truth

Without complete accuracy, we can't be sure that the combat in the experiment will turn out in the same way that real combat would. In fact, in light of some of the comments made above, we can nearly be

38. In addition, the experimental force consisted of squads, who were calling in fire on a battalion, so if a squad were ever to be found and engaged by the battalion, no particular ingenuity would be needed to adjudicate the result.

sure that it *won't*. So then how can we hope to find the truth through experimentation? Is this not a case of "garbage in, garbage out"?

There are a number of reasons to hope that, despite all of the inaccuracies and artificialities, the truth can be found,[39] but the fundamental reason is this: *we do not require that the fighting in the experiment's event turn out as the real fighting would, we only require that the outcome of the event be the one that is matched to the true answer to the question.*

Cheating

In exercises, a certain amount of leeway in regard to the rules, sometimes summarized by the phrase, "If you ain't cheating, you ain't trying" is expected and allowed.[40] There even exists a respectable rationale: exercises are so artificial and constrained that cheating is the only opportunity for the kind of creative thinking necessary for success in actual warfare, and some amount of cheating ought therefore to be allowed.

However, *an experiment is not an exercise*. It is hard enough to construct a valid experiment without having to allow for the possibility that the participants might deliberately violate the rules. The response, "Well, in warfare, there aren't any rules," is thoughtless. Certain courses of action that would make a great deal of sense in a real war are forbidden in an experiment, whether for reasons of safety, geographic and temporal limitations, or the very nature of the event as a deliberate attempt to gain knowledge. Regardless of the rationale given, the frustrated participants are very likely to say, "they wouldn't let us do it because they said it wasn't fair," and to follow up on this observation with a detailed discussion of why considerations of "fairness" have no place in armed conflict. This attitude almost always comes from mistaking a fully justifiable desire to have a "fair experiment" for a misplaced desire to have a "fair fight."

---

39. This topic is treated at some length in the companion piece, *The Art of Military Experimentation.*

40. Typical applications of this phrase appear on pages 35 and 40 of the novel by DiMercurio.

Characteristics of effective military experimentation

To be effective, military experiments must be correctly set up in terms of a question and its possible answers, and an event and its possible outcomes, as described above. They must be well-planned, well-observed, and in general well-done in a host of obvious ways, some described heretofore in this document. But effective military experiments also tend to share a number of other characteristics that are not to be taken for granted. This chapter is devoted to some of them.

Conceptual basis and grounding in theory

A conceptual basis includes, but is hardly limited to, the experiment's hypothesis. For example, the first MCWL MOUT LOE had the hypothesis that the new tactics of penetration, thrust, and swarm would help in urban combat. This hypothesis was part of the conceptual basis: other parts included the definitions of the tactics, the idea of what a likely urban mission for Marines might be, and so on.

An important part of the conceptual basis is a *grounding in theory*.

The word "theory" has a variety of meanings. It is variously used:

- As if synonymous with "hypothesis," or even "speculation," as in, "I have a theory."
- As the antonym of "practice," as in "That's all very well in theory, but it would never work in practice."

- To mean "systematically organized knowledge applicable in a wide variety of circumstances, especially a system of assumptions, accepted principles, and rules of procedure devised to analyze, predict, or otherwise explain…,"[41] as in "music theory" or "game theory."

Especially in the military, the widespread derogatory use of the term in the first two senses has not only detracted from its use in the third sense, but also may even have deterred it. In fact, much of what passes for "military theory" is either platitudinous ("Inflict the maximum casualties on the enemy while suffering the least possible level of casualties to one's own force,"), without empirical foundation (the famous 3:1 ratio of offense to defense has surprisingly little),[42] or both.[43]

However, there do exist some useful military theories, such as that of John Boyd, who thought in terms of the Observe-Orient-Decide-Act ("OODA") Loop, or the "energy" theory of fighter combat, which takes as its starting point the sum of the kinetic and potential energies of the aircraft. Possible ground combat theories include those based on firepower, those based on attrition, and those based on maneuver.

Different theories of warfare would lead to different ideas for experiments, but they would also lead to different ways of conducting experiments. A maneuver-warfare theorist's experiment would concentrate on maneuver, possibly using just headquarters vehicles to move about in a large region and represent their forces, as in a Tactical Exercise Without Troops (TEWT). An attrition theorist would require that all the troops be represented, along with a means of eroding their numbers. A firepower theorist would insist on some system that accurately reflected the firepower of different weapons.

A person who lacked any theory of warfare would not know where to begin in conducting an experiment. Worse, he or she would not be

---

41. *Webster's II New Riverside University Dictionary*, page 1200.

42. Dupuy, 1987.

43. This point is forcefully made by Davis and Blumenthal in their RAND report, *The Base of Sand Problem.*

able to assess the implications of the points of difficulty that will inevitably emerge. For example, in the urban experimentation to which MCWL has devoted considerable effort, it turns out to be quite difficult to simulate shots through the walls of buildings. Absent any theory of warfare, one cannot determine whether this is a minor matter that will not change the outcome of the event (and thus the answer to which this outcome points), or a major point that must be resolved in order to have meaningful experimentation. Some argue that such a point of unrealism is of no consequence because it applies to both sides. But this argument, if pursued to its logical extreme, could be used to justify anything—MCWL could send its Marines home and conduct experiments in urban warfare in the MCCDC building, using civilian analysts armed with Nerf® weapons.[44]

Informed participants

Because of experiments' superficial resemblance to exercises, the two are often mistaken. But *an experiment is not an exercise*. In the early stages of Urban Warrior, MCWL made a point of giving a presentation, to one and all, that drew the distinction. This presentation also made the participants aware of the questions that the upcoming experiment was designed to answer, and of the distinction between surrogates and prototypes.

At some point MCWL fell out of the habit of giving this presentation, on the premise that the individual Marine did not need to understand these highfaluting ideas. But this premise underestimated not only the Marines' curiosity and ability to absorb abstractions, but also the benefit of converting them from being subjects of experimentation to being partners in it: experimentation suffered, and the presentations were reinstated.

Iteration

The term "iteration" could be used to describe the process of getting an acceptably large statistical sample, as described in the chapter on

44. Nerf® is a registered trademark of the Hasbro Corporation.

"Methods," above, but it also arises in a different way, and on a different scale of time: an entire sequence of experiments.

The idea of proceeding from wargame to LTA to LOE to AWE embodies iteration in this sense, as does the idea of having multiple LTAs and, especially, LOEs.

For example, in Urban Warrior LOE 1 there was no attention given to the adjudication of indirect fires, with the result that the fires always went exactly where they were aimed. The analysts noticed that calls for fire were often made from positions extremely close to the intended target.

In LOE 2, we used realistic present-day CEPs,[45] adjudicated via a simple dice-rolling system, for indirect fire weapons, with the result that there were a large number of Blue-on-Blues and civilian casualties and little damage to the enemy.

In LOE 3 we used futuristic CEPs of 1-3 meters for the same weapons; the adverse results decreased enormously and the effect of the fires increased. The conclusion was that in urban fighting there would be a big payoff from smaller CEPs. This conclusion could not have been reached simply from the initial, zero-CEP experiment.

Prior to Kernel Blitz (Experimental), which was really a collection of four LOEs, repeated LTAs had been done regarding Precision Targeting Systems (PTSs—these combine a compass, a laser rangefinder, a GPS unit, and a computer so as to create a fast and accurate means of finding a visible target's grid coordinates). One of the Kernel Blitz LOEs contained a precision targeting piece, and some inveighed against this on the grounds that "we've already experimented with PTS so much." But the LTAs could only show how accurate the PTSs were, not how much they would help: for that, a force-on-force LOE was needed.

The term "iteration" also describes the key practice of conducting planning in a sequence of loops, rather than as a straight-through

45. The circular error probable (CEP) is the median miss distance.

process. One group of planners, for example expended months of effort on a two-stage wargame (first a seminar game, then a JCATS-assisted game that was really an experiment in itself) as the first step in planning for an AWE-level experiment in sea-based operations at the battalion level, only to be told that experiments could not be larger than company-level in size.

Long-term effort

Because of the need for iteration, in the sense just described, military experimentation requires a sustained, long-term effort. Most of the successful pre-WW II efforts to which today's efforts are often compared took place over periods of years.[46]

To qualify as long-term effort, what is needed is not simply long-term continuity of an institution, or of spending, or of involvement by key people, though all of these can help. What's needed is long-term continuity of effort *and purpose*, and a means of documenting the work and ensuring long-term availability of this documentation (see also below). For these reasons, the *Code of Best Practice for Experimentation*[47] refers to experimentation "campaigns."

Two MCWL successes have been UCATS (the Universal Combined Arms Targeting System), which allows a FO (Forward Observer) to find the position of a target using a laser-based Precision Targeting System (PTS, mentioned earlier) and then to transmit that grid to a Fire Support Coordination Center or aircraft as part of a larger preformatted digital call for fire, and the predecessor ACASS (the Advanced Close Air Support System), which similarly allowed a Forward Air Controller to locate a target and transmit its location to the aircraft along with the rest of a digital nine-line briefing. Yet the creation of workable ACASS and UCATS systems, now resident as separate pieces of software within a common piece of hardware, took a matter of years, and was threatened at various times by people who thought it was taking too long and not really progressing.

---

46. See *Wotan's Workshop*.

47. Alberts and Hayes.

Focus on revolutionary improvement

Successful military-experimentation efforts of the past focused on the creation of revolutionary, not incremental, improvements. Many people cite the Germans' blitzkrieg method of combined-arms warfare as an example, though in fact their creation of U-boat "wolf pack" tactics is probably a better example. American examples would include the pre-WW II development of operational art for fast carriers, and the development of USMC landing doctrine.[48] These efforts created whole new ways of fighting, not just improvements on old ways of fighting. They also took many years of effort.

Quantification

With reference to his studies of inter-species competition, the naturalist Gause remarked, "Apparently every serious thought on the process of competition obliges one to consider it as a whole, and this leads inevitably to mathematics."[49] The same could certainly be said with "combat" replacing "competition." Even those who profess deep distrust of quantification can usually be heard to resort to it when expounding their views of military matters, using such terms as "more," "fewer," "every," "most," "the majority," and "none," if not actual numbers.

Others use the term "quantitative" as if it were a synonym for "objective," which it is not.

Thus the planner of a military experiment finds near unanimity that the experiment should produce quantitative results.

The opposing view, an extreme position held by few, is that one ought not to derive quantitative results from any experiment more complex than the most test-like LTA: the only use for such numbers, according to this view, would be their incorporation in a computer model, but

48. See *Wotan's Workshop.*

49. Gause, page 7.

because nobody believes such models, there is no point in deriving such results.

On the whole, the derivation and presentation of some simple quantities (e.g., casualty counts and ratios thereof, times and distances, and locations of "hits" on the body if these are recordable), is justifiable. These quantities *are* good for more than just inclusion in models, e.g.:

- They can be compared from one scenario to the next, or even one experiment to the next, to show changes. The use of the casualty data to show the utility of accurate supporting fires, described above under the heading "Experiments in contrast to tests," is an example of such a use.
- Later experimenters might want to use them for reasons of their own. Workers in MCWL's Project Metropolis, a follow-on to Urban Warrior, compared Urban Warrior casualty data to their own, to argue that the lower casualties in Project Metropolis showed the value of the training package they had developed.[50] In such cross-comparisons, great care must be taken to ensure that the two sets of data are, in fact comparable. This cannot, generally, be done after the fact: the second experiment must be designed specifically with a view to comparability with the first experiment. Imagine doing the second experiment, finding that casualties were lower, and then finding that because the OpFor strength was less in the second experiment than in the first, one cannot ascribe the reduction in casualties to a difference in training!
- Quantifications, e.g., of casualties, are available regarding historical battles, and even those who see themselves primarily as consumers or producers of "seasoned military judgment" can compare these to quantitative characterizations of an experiment's mock battle to give themselves a sense of where it fits into the constellation of historical cases, or into their own experience.

50. Project Metropolis.

In addition, of course, one should derive specific quantities that relate to the experimental objectives, but it is important to recognize that in the larger picture, the minority view is correct: the quantifications are almost all only the means to a non-quantitative end.

Once again, the precision targeting systems (PTSs, mentioned above) furnish a good example. First, a series of LTAs found the CEP of the devices. The CEP characterized the performance of the system (and could have gone into a computer model, had one been under development), but the much of the LTAs' value came from the objective, and yet non-quantitative, facts that they revealed about the system, e.g., that tall grass could introduce severe errors in range. Upon being given a figure for the CEP, people almost invariably wondered how the PTSs' CEP compared to that of traditional Forward Observers' sightings. A figure for the latter was available from some much earlier Army experiments, and showed that the PTS represented a major improvement in CEP. The natural reaction to this information, in turn, was to wonder how much good the PTS would be in a combat situation. An LOE, with a PTS-less base case and an experimental case done with PTSs (and, ideally, everything else the same as in the base case) could help answer this, with the answer probably cast in terms of such quantifications the reduced time to accomplish the objective, reduced casualties, or increased number of enemy troops killed by the weapons targeted using PTSs. But none of these analytical measures is really the bottom line: in the end, some measure of military judgment must be introduced to fill in the picture that the analysis has outlined. A later section will address the *assessment* process, by which this judgment is added after the analysis is complete.

Documentation

Because of the need for iteration and sustained effort over a long period of time, military experimentation takes long enough that particular projects need to be able to survive the departure of an involved officer and the arrival of his replacement. The best means of making the institution's memory longer than individuals' tenures is to have written reports. At the minimum, there should be an analysis report for each LTA, LOE, or AWE; LOEs and AWEs (if not LTAs) ought also to be the subject of *assessment* reports as described later. Ideally, there

will also be, from time to time, reports written on particular topics, as opposed to particular events: MCWL's reports, *Autonomous GPS-Guided Aerial Resupply Systems* and *Summary of Experimentation with Precision Targeting Systems* are examples of such reports.

Report-writing is addressed in a later section.

Obstacles to effective military experimentation

Military experimentation is difficult, as noted at the very beginning of this paper; one can readily imagine impediments to it a priori, and more can be imagined by considering the removal of the "characteristics of successful experimentation" recounted in the previous chapter, generating such obstacles as "absence of theory." But other obstacles can also stand in the way of conducting a successful military experiment. This chapter is devoted to some of them.

Experimenting on the surrogates

As mentioned in the discussion of surrogates, experimenters sometimes fall into the trap of experimenting *on* the surrogates, as if they were test articles to be evaluated.[51]

An example would be an experiment that sets out to find the utility of giving each Marine a small handheld radio for purposes of conducting limited-war "block two" operations in urban terrain. It would reasonable to do the experiment by obtaining a number of such radios commercially, and having the Marines then conduct a number of scenarios in a urban training area, using most or all of the "components of a military experiment" described in a previous chapter. Ideally, each scenario would be conducted twice, once with the radios and once without, the results compared, and the Marines debriefed and obliged to fill out questionnaires. The trouble arises if and when the focus shifts from "did having radios help?" to "were these good radios?"; the latter amounts to experimenting on the surrogates.

This trap becomes especially inviting when no clear statement has been made as to whether the system at hand is a surrogate or a prototype.

---

51. See also Karppi and McCue.

The issue is sometimes clouded by the need to measure how well the surrogates performed, either simply to be sure they were good surrogates, or because one person's surrogate is another person's system of interest, and the latter has loaned the surrogate to the former in the expectation of some analysis of how well it works.

The issue is also clouded by cases in which surrogates worked so well as to engender the recommendation that they be produced or bought en masse and given to operational units, as in fact happened with the handheld radios mentioned above.

Ignorance, or disregard, of previous work

The writing of reports is a necessary, but not sufficient, condition for the transfer of information across time. The reports must be accessible, and the staff must have an awareness of their duty to make themselves familiar with what has already taken place. All too often, MCWL personnel would "re-invent the wheel," or—worse—deny that the invention of the wheel had taken place, because of their failure to acquaint themselves with the Lab's body of knowledge as contained in its reports.

In addition to becoming acquainted with one's organization's previous work and then keeping up with any progress, one should also avoid being ignorant of relevant work done elsewhere. An analyst at MCWL recalled a maxim from her training as a laboratory scientist: "every experiment begins in the library."

Reliance on participants' opinions

Often, a shortcut analytic method is proposed: give the participants the piece of equipment in question (and some training on it), put them in a realistic situation with it, and then ask them if they liked it.

This method suffers from two separate difficulties.

First, they may not render a true opinion on whether they like it or not. They may believe that they are "supposed" to say that they like it,[52] or they may honestly not be able to tell whether or not they would like it, given only a short and somewhat artificial exposure. In some cases, Marines did not understand that they had been using surrogates, and they decried the devices on the basis of shortcomings—e.g., of ruggedness, or the type of battery used—that were really only shortcomings of the surrogate.

Second, whether or not they like it is not necessarily indicative of whether or not it is good. In one Urban Warrior LTA, the Marines said they liked an experimental gun sight—and maybe they did—but their scores were lower with it than with a conventional sight. The users of the original machinegun, for example, did not like it, but machineguns (even that original *mitrailleuse*, known to many Marines because one is displayed in the lobby of the Headquarters building at Camp Lejeune) later proved their worth.[53] American waistgunners in Second World War bombers liked tracer ammunition, despite considerable evidence that its use was actually counter-productive, in part because the tracer rounds were lighter than the other rounds and were correspondingly more deflected by the slipstream, and in part because the gunners would try to "whip" the bullet stream as if it were a long, flexible stick.[54] Second World War U.S. submarine skippers in the Pacific, to take a final example, did not like the air search radar when they were given it; with submariner's classic aversion to active systems, they were concerned that Japanese aircraft would somehow detect its emanations and home in on them, and their increased observation of aircraft when the radar was on seemed to confirm this

52. At least with Marines, however, this is not as much of a problem as one might suppose. Early in the present author's work with Marines, a general officer suggested to a group of young Marines that a particular piece of equipment would serve them well. The young Marines, most of whom had fewer stripes than the General had stars, respectfully replied, "No, Sir, that would not work for us," and the General accepted this for the valuable and honest input it was.

53. Brodie and Brodie, page 145.

54. Dr. J.J.G. McCue, personal communication.

concern. Analysis, however, showed that aircraft density in the vicinity of submarines did not increase, and that the increased number of sightings was explicable entirely on the basis that the radar was performing its function—detecting aircraft that would not be detected visually.[55]

These difficulties can be discerned in the stated reasons for disliking the experimental equipment, which are normally that it is too heavy or too delicate. But nearly everything is heavier or more delicate than one would like it to be, and improving it in one respect will worsen it in the other: the question is whether or not the additional weight and caution required by the new equipment are worthwhile.

Fear of failure

Conventional wisdom holds that we learn from our mistakes. In a sense, experimentation amounts to a formalization of this process. For a variety of institutional reasons, however, workers in military affairs—especially those in uniform—are intolerant of failure in themselves or others. These opposing attitudes collide in the case of military experimentation, whose practitioners seem therefore to need occasional re-assurance that not everything with which they experiment needs to succeed. Some even hold that if everything does succeed, that will be an indication of undue timidity in trying new ideas.

Sometimes this guidance is expressed succinctly, in intentional apposition to the usual mindset, “It’s OK to fail.”

Ignoring distinctions among types of failure

There are two possible ways in which an experiment may go badly.

1. It fails to produce data that support the hypothesis, or
2. It fails to produce data at all.

---

55. Morse and Kimball, pp 59-60.

The needed guidance, "It's OK to fail," is sometimes misinterpreted. The intent of the guidance is that when trying out new ideas or pieces of equipment, some of them will fail, and if one does not have some proportion of failures, one is not trying sufficiently new things. That is to say, it is acceptable to suffer failures of the first type cited above.

However, some have sometimes taken "It's OK to fail" in the wrong way, applying it to failures not related to experimental ideas or equipment, or to the still-developmental aspects of the art of military experimentation, but instead to failures of experimental surrogates, or in such mundane matters as frequency allocation, data collection, reservation of ranges, and the like. In these respects, i.e., in second sense cited above, it is not "OK to fail."[56]

One may discern, in these examples, the distinction between the types of failure that are acceptable and those that are not by applying, once again, the notion that *an experiment is not an exercise*: failure in experimentation is acceptable only in those parts of an experiment that would not be present if it were an exercise.

Pandering to the experimental unit

An experiment is not an exercise, but a unit participating in an experiment will usually, and justifiably, expect to receive some benefit in return for allowing itself to be used as a guinea pig. The most obvious benefit it can receive is training, and it is not unreasonable for the experimenters and the unit to negotiate their way to an experiment that is configured, in part, with a view to the training benefit it offers to the experimental unit.

However, two cautions are in order.

56. A Service or other entity could responsibly take the view that because military experimentation remains developmental, occasional failed *experiments* (as opposed to experimental ideas or pieces of equipment) are in fact to be tolerated. But this would constitute permission to take risks in experimentation, not permission to do a sloppy job.

First, the objective of providing training can, in most cases, be met readily enough, and little adjustment of the experiment will be required. If the experiment seems grossly deficient in terms of its training value (e.g., if the scenario sets up an engagement in which one side is almost guaranteed to be defeated immediately), then the experimenters probably need to reconsider its value as an experiment. In fact, the abandonment of traditional training-oriented artificialities (e.g., that leaders cannot be killed) can in itself create beneficial and novel training situations.[57] On numerous occasions, a Marine would approach a MCWL analyst after an experiment and say, "Sir, I know that this experiment wasn't for our training, but I just want to tell you that this was the best training that my Marines and I have ever had."

Second, any adjustments made to accommodate the unit can and should be made well in advance, when their effect on the experiment's main goals can be carefully considered.

False serendipity

A problem with serendipitous findings is that, precisely because they concern matters not contemplated when the experiment was designed, they may be spurious. Any apparent serendipitous finding needs to be subjected to careful examination to determine whether or not it might *stem from an artificiality of the experiment*, and not from an aspect of the experiment that mirrors the real world.

After one of the Urban Warrior LOEs, for example, an officer wrote, "helicopter-mounted rockets and machine guns proved remarkably

57. Sometimes there arises concern that experiments, necessarily embodying departures from reality (e.g., in equipment, TTPs, etc.), can be sources of "negative training." My great skepticism about such alleged negative training has no rigorous basis, but neither do the concerns: it is my belief that these concerns underestimate the discernment of the participants, and that any negative training that may exist is more than outweighed by the positive training benefit cited above—the elimination of the usual artificialities, even if it does come at the price of introducing some new ones.

effective." It was certainly the case that these weapons had killed more of the enemy than most people would reasonably have expected before the experiment, and it might even have been the case that expectations were low and that the experiment indicated that these weapons had a higher potential in urban warfare than the conventional wisdom had thought. But the main reason for the effectiveness of helicopter-borne rockets and machine guns was that the adjudication procedures did not provide for any way that they could miss: their high effectiveness was thus at least in part—and perhaps in very large part—an *artifact* of experimentation. Seeming serendipitous discoveries must always be checked for this kind of flaw: precisely because they concern matters not anticipated by the experiment's designers, there is no guarantee that they are valid.

Serendipity being by definition unanticipated, one ought not to rely on it to occur. One area in which there is a great temptation to rely on serendipity is that of Tactics, Techniques, and Procedures (TTPs). On multiple occasions, MCWL proposed the creation and validation of TTPs as an experimental objective, usually in response to the receipt of a new piece of gear. The progression became predictable: first, the intent to create multiple sets of TTPs, teach them to the ExFor, and experiment to see which worked best; then the intent to create a single set of TTPs, teach it to the ExFor, and experiment to see if it worked; and finally, to create no TTPs and provide no training, and instead just give the Marines the new piece of gear, watch them try to use it in an experiment, and record anything that worked as a TTP. This approach did not work: in the time available, the Marines were seldom able to discover any truly useful TTPs, and in some cases they recognized that this would be the case and didn't even use the new piece of gear.

Of course, one way to avoid falling into the trap of false serendipity is never to make any serendipitous findings at all. This cure, which has also been attempted, is probably worse than the disease.

Unwarranted generalization

Frequently, an experiment is billed as demonstrating a general capability, on the basis that it demonstrates particular instance of that

capability. For example, an experiment might entail the operation of a computerized system designed to manage interrupted SEAD missions, in which a mortar fires at the target before and after an airstrike, and that the system successfully does so, no mean feat. Are the experimenters then entitled to claim that they "have demonstrated a system that manages and deconflicts the application of artillery, naval surface fire support, close air support, and ground troops"?

They might say so, on the grounds that their system has performed an important task in that area of endeavor, but in fact there are not, because no meaningful sampling has taken place. This is the point of the "question for discussion" propounded in the Introduction, which asked why the townspeople hired the first applicant to be their doctor, and ran the second applicant out of town on a rail. The difference between the first applicant and second is that first took a test selected by the townspeople, but the second selected his test himself. Thus, even though the act performed in the test (the appendectomy of a dog) was the same in each case, the meaning is different: in the first case, it is a sample of a larger whole, but in the second it is not.

Similarly, the SEAD mission, while important, is only one aspect of the claimed larger set of capablilities, and it lacks the status of being a "sample" because it came first, and then the larger claim was built around it.

Absence of a surrounding knowledge-gaining enterprise

The conceptualization of worthwhile experiments, the development of means by which to carry them out—their execution, analysis, assessment, and documentation, and the subsequent use of that documentation—are all made easier if embedded in a *knowledge-gaining enterprise*. It is perhaps for this reason that the Marine Corps Warfighting Laboratory and other institutions engaged in military experimentation have been given the rubric, "laboratory."

Occasional failures of MCWL to foster worthwhile work can almost all be seen as incongruous behavior in something that is supposed to be a laboratory. Examples include disregarding previous work, failing to

document what one has done, or devoting great time and effort to non-research events.

The command mentality

While one would trust that a military command would not become fully pre-occupied with making itself look good, to the exclusion of accomplishing anything, it is certainly true that (at least in peacetime) military commands devote considerable effort to ensuring that they do nothing to make themselves look bad.

By definition, any document coming out of a command is signed by the commanding officer. It is his or her document, regardless of who actually wrote it, and it is read according to cultural precepts regarding the reading and writing of such documents. These precepts are incompatible with the frank reporting of an experiment: such reporting must recount any difficulties involved, yet in the culture of documents written by commands, statements regarding difficulties are often regarded as whining and excuse-making (especially if the difficulties were not surmounted), or attempting to put others in a bad light (especially if the difficulties were surmounted).

More generally, the command mentality can interfere with the creation of any report at all. Rightly or wrongly, the command mentality can dictate that the mere fact of reporting on a subject constitutes endorsement, and that therefore a report on something to which the command is unfavorably disposed ought not to be published, *even if the report confirms the unfavorable impression.*

Even more generally, commands are reluctant to use what they perceive as "loaded" terms, and the threshold for these can be suprisingly low. More than one command has balked at the term LOE, for example, because it contains the word "limited," and they don't want to be associated with anything that is limited.

Therefore an organization devoted to military experimentation might best not be a command, though there then arises the problem of a line of command for the forces involved in the experiment.

Throughout most of its history, the Marine Corps Warfighting Laboratory had, attached to it, a Special Purpose Marine Air-Ground Task Force (Experimental), the SPMAGTF(X). The intent behind establishing the SPMAGTF(X) was that it would consist only of a cadre of field-grade officers, a few company-grade officers, and a skeletal enlisted staff, temporarily augmented as necessary for performing experiments. The relationship between the SPMAGTF(X) and the rest of MCWL was almost always uneasy, however, and the right (or lack thereof) of the SPMAGTF(X) officers to propose and pursue their own lines of experimentation was never defined.

The "Stockholm Syndrome"

Psychologists have noted that the human tendency to bond with others, particularly if those others are responsible for meeting some of one's needs, is so strong that hostages even tend to bond with their takers, despite the fact that the latter may be threatening to kill the former. This phenomenon is called "the Stockholm Syndrome," after the hostage-taking event in which it was first documented, apparently by Strentz.[58]

In the case of military experimentation, the syndrome is wryly invoked as a convenient term for the tendency of observer/controllers, and others, to become overly sympathetic to those whom they are observing and controlling. This sympathy manifests itself in a variety of ways, all damaging to valid experimentation, such as an unwillingness to declare casualties from adjudicated fires. (Conversely, when the case for declaring casualties becomes overwhelming, the entire group is often declared "dead," the observer/controller not wishing to be in the position of choosing some to be dead and not others.)

Rotation of observer/controllers would seem to be an obvious cure, but the bonding may occur so fast that no reasonable rotation scheme could defeat it, and there are countervailing advantages to having observer/controllers stay with the same group of troops for a while—

58. See Strentz.

for example, they can keep better records once they know their troops' names.

The "Stockholm Syndrome" can strike at the highest levels, where it becomes difficult to distinguish from "emphasis on winning," described below. At a lower level, it is exemplified by the behavior of the observer/controllers in Urban Warrior's culminating AWE, who went ahead of the Blue units so as to find any tripwires.

Emphasis on winning

Everybody wants to be on a winning team, and experimentation benefits from this fact because it impels the participants to great efforts even though they and their loved ones are not in danger of death or imprisonment if they are defeated, as would be the case in a real war.

However, the desire to see the experimental side win or, after the experiment, to see it depicted in the analysis as having won, can readily overcome the desire to learn something from the experiment.

MCWL's treatment of fratricide illustrates this point. For weapons larger than the Squad Automatic Weapon (SAW), MCWL had no MILES gear, so fire had to be adjudicated. The O/C of the shooting unit would call ExCon and say, for example, "My guys are shooting at some guys in Building 19." ExCon would then contact the O/Cs of the other side and ask, "Have you got anybody in Building 19? You're taking fire and you should assess some casualties" It was pointed out that this procedure nearly ruled out fratricide, because the two O/C nets were separate, so in a fratricide incident the O/C of the targeted troops would not get ExCon's call. The reaction was that this was a needless concern, because fratricide is bad, so anything that reduces it must be good.

Turning data into power

Francis Bacon said, "knowledge is power."[59] A computer-age saying adds, "But information is not knowledge, and data are not information."[60] This chapter describes how:

- During the experiment, *observation* collects data; then
- *Reconstruction* turns data into information;
- *Analysis* turns information into knowledge; and
- *Assessment* turns knowledge into power.

The reconstruction, analysis, and assessment must all be turned into written, published reports, or else they are useless. *An experiment is not an exercise*, so considerations such as the benefit to the participants, or even the on-lookers, are of no lasting consequence: anything that is not written down in an organized way and made available to present *and future* users is a total waste in terms of experimentation, however valuable its side effects of training or public relations. Some have argued that the VIP onlookers represent the funding for the experiment, and that they need to see a "good show" or they will not provide funding in following years. My own observation has been that people of such importance are usually remarkably shrewd in discerning whether they are being shown a show or an experiment.

After the observation, reconstruction, analysis, and assessment steps are complete, and the report(s) written, any of a number of actions may take place. One frequent choice is the transfer of experimental gear or TTPs to an operational unit, for "experimental use": this chapter concludes with a discussion of this idea.

---

59. Cited widely, e.g., in the *New International Webster's Dictionary of the English Language.*

60. Ford.

Data collection

During and immediately after the experiment, data are collected by people and instruments.

The people (observer/controllers, analysts, ExCon and the participants themselves) and instruments (MILES, IGRS and the like) have been discussed already.

The data typically consist of:

- the task organization and orders of the units as of the beginning,
- the locations (as a function of time) of vehicles and troops, ideally collected by an instrumentation system, but possibly collected by observer/controllers,
- MILES shot, near miss, and hit data, and/or Simunition® hit data, and ammunition consumption
- accounts of engagements, given by witnesses (observer/controllers or analysts) and/or participants,
- logs, especially fires logs (often maintained by ExCon as part of adjudication), and
- accounts of decision-making, gathered after the fact in debriefs and questionnaires.

These data are the raw material from which the analysts produce the reconstruction.

Reconstruction

Leopold von Ranke (1795-1886), generally recognized as the "father of modern historicism," stated that his goal as a historian was that he would "merely tell how it really was."[61] This goal, strikingly modest by the standards of the historian-moralist-philosophers against whom Ranke was reacting, is considered by today's historians to be in fact quite difficult, if not impossible.

The goal of the reconstruction is to create an account of what happened, and, as in the study of history, the task is more difficult than it sounds. Based on the description above of the expected data, one might suppose that this task would be time-consuming, but not difficult—after all, the data are all there.

Such a supposition would be half right: the task is time-consuming (for a company-sized experiment that lasts a week, a half-dozen analysts could expect to spend ten days of individual effort, followed by five days of group effort, wrapped up by one analyst in a final week of work to accomplish the reconstruction), but it is also difficult.

The merely time-consuming part is the assembly of all the times and locations into tracks and engagements, and the creation of an account of the casualties.

The difficulties arise not only because the data are inevitably incomplete and mutually contradictory, but because "what happened" also includes the human element:

- To what were participants reacting when they took a certain action?
  - — What could they see? What couldn't they see?
  - — What had they heard on the radio?
- What went into a commander's decision?
  - — What did he know?
  - — What did he deduce or assume about what he didn't know?
  - — Why *wasn't* he aware of certain facts?

---

61. English translations of this widely-cited saying vary to more than the usual degree, apparently because the original ("*...wie es eigentlich gewesen*") is an unusually truncated turn of phrase in German. The source is clear, however: it is his "Critique Of Modern Historical Writing" ("Zur Kritik neuerer Geschichtschreiber") appended to his book, *History of the Latin and Teutonic Nations, 1494-1514* (*Geschichten der Romanischen und Germanischen Völker von 1494 bis 1514*), published in 1824.

- How did the participants come to be involved in a "Blue-on-Blue" fratricide event?
- Etc.

Some questions regarding such aspects prove to be unanswerable, but with a group of analysts, each having first assembled his or her own data and prepared an account to be given to the group, a surprising amount of information can be deduced by combining the different analysts' results.[62]

The end product is a complete, fact-based, time-synchronized, deconflicted, and *meaningful* account of what actually happened.

Analysis

Analysis takes the record of events, provided by the reconstruction, and seeks patterns in them.[63] It does so in a manner that is *objective*.

The seeking of the patterns is largely an attempt to determine which of the *outcomes* (identified during the design phase of the experiment, as described at the beginning of this document) actually came to pass. In some experiments, the outcome will be quantitative, e.g., the decrease, if any, in casualties as a result of the use of some supposedly casualty-reducing technology or tactic. In other experiments, the distinction will be qualitative, e.g., when maneuvering at night with tactical instrumentation, does the Marines' progress appear (on the IGRS replay) more orderly than when they move at night using conventional night-movement methods? It is important to notice that the pre-specification what to look for and what it will mean, as stated in the overview of experiments at the beginning of this document, goes a long way towards making such findings—qualitative though they may be—*objective*.

---

62. A method of doing so without unseemly acrimony, however, remains to be found.

63. Analysis is treated at greater length in the companion publication, *The Art of Military Experimentation*.

If serendipity, as described above, is to occur, it will usually occur during analysis. The analysts, informed by their personal observations during the events, may well notice a strong pattern that had not been pre-identified as a topic of interest. In MCWL's Urban Warrior experimentation, for example, analysts noticed (first during the event, and then when considering the reconstruction) that Marines were frequently "killed" at the point of preparing to enter a building. This tendency was traced to the Marines' training, and consideration of the Urban Warrior results eventually led to a successful effort to revise the syllabus. The revised syllabus was then tested with additional experimentation.

Much of the skeptical reaction evinced by military officers upon meeting civilian analysts is traceable to an unstated assumption that the analysts' stock-in-trade is the second-guessing of military decisions. So it is important to notice that neither the analysis step, nor any other, entails evaluation of the experiment's participants, or their performance.

Assessment

After the reconstruction is complete and the analysis has at least been drafted, MCWL finds it useful to conduct an "Assessment Conference." Recall that "assessment" is the step that turns knowledge into power.

Assessment addresses the *implications* of the experiment's findings. These are strongly sought after, and are in fact the whole reason for doing the experiment, and although the analysts may well be aware of them, they cannot make them part of the analysis per se, because they follow from the experiment's findings *and* a knowledge of the real world, including operational, political, and programmatic realities, and the analysts' assignment is to analyze only the experiment.

Military officers have greater latitude. Suppose, for example, that an experiment has tested a new radio mast for a submarine: the submarine or a surrogate has extended the mast above the water, the appropriate satellite has been re-oriented so as to cover the mast with the center of its main lobe, and signals have been received and their

strength measured. The analysts can determine that, off boresight, the signal margin would be inadequate, but it is not their place to say, "This system only worked because the satellite was aimed right at it, and operationally, nobody will ever do that." The military officers can say this.

In the Assessment Conference, therefore, the analysts brief knowledgeable officers and other subject matter experts on the findings of the experiment, and a discussion as to the implications of those findings ensues. The result is a report, written by somebody in uniform, on the implications of the experiment. This report usually contains recommendations as to which lines of inquiry should be developed, or dropped, in future experimentation.

Report writing

An experiment is not an exercise. Therefore its training value to the participants is only a welcome bonus, not a justification of the effort. The learning value to the experimenters is of transient value, at best. The only lasting value of the experiment is that contained in the resulting report(s).

These reports need to document not only the conclusions drawn from the experiment, but also most of the details:

- The question(s) that the experiment was supposed to answer, and why they were important;
- How and why the possible outcomes of the experiment were matched to the answer(s);
- Who and what were in the experiment, and where and how it was conducted;
- What happened, including a detailed reconstruction of each event
- Conclusion(s)—answer(s) to the question(s) around which the experiment was designed.
- Observations—other important discoveries arising from the experiment

- Recommendations for future experimentation, if any.

Many readers will not want to read so much, so there should be a summary stating just the conclusions.

Even if very few readers are interested in all the details, these details must be included in the report. One reason is that future workers will need to know them, either to perform further analysis, or to attempt to construct a comparable experiment of their own in continued investigation of the same topic. But another, and perhaps more important reason, is that without the presentation of all the details, the presentation of the conclusions will appear to be pure pontification. The presentation of the details provides solidity, setting the work apart from the great mass of pontification that is always available on military topics of interest.

As discussed above, an experiment may go badly either by failing to produce data that support the hypothesis, or by failing to produce data at all. In a traditional scientific experiment, the investigator is duty-bound to report the results, regardless of whether they support his or her hypothesis,[64] but is largely absolved of that responsibility if he or she has no results at all. In contrast, a military experiment will be expected to produce a report no matter what. This practice is arguably the more honest, but the reader—especially the accustomed to reading the conventional scientific reports—is likely to react badly *to the report*, thinking ill of it, when in fact the problem lay in the experiment. The difficulty of writing the report under such circumstances is increased by the fact that the analyst will usually attempt not to put any of the participants and planners in a bad light.

Publication of results

In addition to being written, a report must be published if it is to be useful. Publication has the obvious benefit of distributing the report to potential readers, some of whom might act on it in one way or another, but it has some important *side-effects* as well. These include:

64. Though in practice there is widespread sentiment, and even some empirical evidence, that negative results are under-reported.

- The fact that a report has been published indicates that somebody felt strongly enough about it to expend the resources to publish it. In this respect, one *can* actually "tell a book by its cover."
- Publication and widespread distribution increase the probability that copies will survive and be available to those who become interested at some future time.
- Publication of results constitutes an overt act on the part of the experimenting organization, which would otherwise be seen as simply spending money on exercises and public relations.

Publication on the Worldwide Web seems to have gained sufficient acceptance that it can be considered as an alternative to publication on paper, and it certainly has the effect of making the report available to potential readers, but before deciding to publish something in that way only, a researcher should consider the continuing (and understandable) skepticism regarding material found on the Internet, as well as on the degree to which electronic publication may not fully provide the positive side-effects listed above.

Giving equipment to operational units

After a successful experiment, there frequently arises the idea of giving the experimental equipment to an operational unit, usually one that has used it in an experiment.

This idea is fraught with difficulties, including that:

- The equipment in the experiment may have been a surrogate, able to perform some or all of the functions of the intended, eventual "real thing," but not sturdy, reliable, or otherwise suited to operational use.
- No support infrastructure of spare parts, trained technicians, or maintenance manuals exists to support the experimental equipment, even if it is a prototype and not a surrogate.

- Needed certifications, e.g., that the equipment is safe to carry aboard an aircraft, may be difficult or impossible to obtain, either for a prototype or a surrogate.
- If the equipment goes to a unit that did not use the equipment in an experiment, there is the added problem that the personnel in the unit have not had any experience or training with the equipment.

The likely result is that the operational unit is disappointed with the device's performance and becomes disenchanted without ever realizing that they are not working with the "real thing," and that the problems of fragility, maintenance, certification, and so on are largely or entirely the result of this fact.

Even apart from these problems, the operational unit will have trouble contributing to the experimental item's development in a meaningful way, because they will be unlikely to know what data to collect, and certainly will not have a dedicated person present to collect such data.[65] Therefore the reporting of their use of the equipment becomes an extra burden, which an operational unit is unlikely to want to bear.

Finally, it is unlikely that an operational unit would use experimental equipment in an operation. Therefore any use will be in deployed training, and probably not any more fruitful of insight than Stateside training (observable by analysts, et. al.) would be.

65. In the past, there was also the problem that the operational unit, especially if it was aboard ship, would have trouble passing its observations back, but the Internet has made an enormous difference in this regard.

Template for a military experiment

By way of review, this chapter offers a summary of the entire paper, cast in terms of a template for designing a military experiment.

The question

An experiment is a means of answering a question, so the planning of the experiment ought to start with the question. Finalization of the question at this stage may be premature, because the final form of the question may, realistically, have to be adapted to what experiment is feasible, but some effort should be devoted to refining the question at this stage. The parable of the blind men and the elephant applies—a group of workers may say they agree on the *topic* of the experiment, but when they actually sit down and try to formulate a definite question, they are likely to find that they have differences.

Thomas Edison said, "Genius is one percent inspiration, and ninety-nine percent perspiration."[66] This saying is often quoted to children, to emphasize the need for persevering with routine work. It ought also to be quoted to adults, to emphasize the need for aspiring to extraordinary thought: even assuming only a 40-hour week, one percent is 24 minutes, and few of us experience inspiration for 24 minutes of the average week.

Previous work

After a preliminary form of the question has been framed, and perhaps even before, it is important to find out what has been done already. This task requires some open-mindedness, because it is more than simply investigating to see if somebody has already done exactly

66. Cited widely, e.g., in the *New International Webster's Dictionary of the English Language.*

the experiment that is being proposed (which is highly unlikely): it is trying to find any and all written work that bears on the question. Such work would include not only previous experiments, but also "think pieces" written by strategists, historical articles, technical documents, training manuals, and living veterans who can be interviewed.

An experiment typically involves something (tangible, like a piece of gear, or intangible, like a tactic) new. But relevant previous work includes descriptions of what patent law refers to as "prior art"—what is being used or done now, before the new thing comes along.

The size and type of the experiment

The question, or even the general topic of the question, will suggest the type of experiment (wargame, LTA, LOE, or AWE) that is needed to investigate it. Typically, narrow questions are addressed in LTAs and larger questions are addressed in larger experiments, but there are important exceptions to this generalization.

Conceivably, a question could be very narrow, and yet require an LOE or AWE to provide the context: in that case, the experimenter must hope that the needed LOE or AWE is going to be done for other reasons, and that he or she can become involved, because nobody will be willing to do a large experiment to answer a narrow question. An example of this situation is provided by the precision targeting systems: numerous MCWL LTAs had refined their performance characteristics, but there remained the question of how much good they would do. To answer this question would require a large-sized mock battle, which was not likely to be put on merely to answer this question about PTSs, so determination of the degree to which PTSs would help in a company-sized action had to wait until a large-scale experiment (2001's Kernel Blitz) was going to be done for other reasons.

Conversely, some of the very broadest questions are addressable only by the least ambitious experimental effort—the seminar wargame.

Personnel

Knowing the topic and the size of the experiment, one can begin to estimate the numbers and types of personnel that will be needed.[67] Available units will probably be available precisely because they are in the work-up stage of the force deployment cycle, and will therefore have somewhat more than their share of inexperienced personnel. The use of an under-experienced unit is, however, probably advisable, if only in that it immunized the experiment's results against the skeptics' assertion that a hand-picked unit had been used so as to lead to an unrealistically favorable result. Of course, the use of an under-experienced unit as the OpFor would invite a less-rebuttable critique.

The personnel—experimental unit, OpFor, roleplayers, observer/controllers, firewalkers, ExCon, and all—need to be trained prior to the beginning of the experiment's events. The experimental unit has to learn to do the experimental tactics, use the experimental equipment, or to do whatever unusual thing the experiment is to address, and they need to have reached a plateau in this knowledge before the experiment begins. All personnel need to learn some experiment-unique skills, such as the adjudication procedures for weapons not represented by MILES, and how to behave when declared a casualty. The observer/controllers and firewalkers need to know how to respond to instructions from ExCon and how to keep records of the progress of the experiment. The firewalkers need to learn how to use their flash-bang artillery simulators and God-guns. ExCon needs to

67. This notion illustrates, once again, the contrast between an experiment and an exercise. The author once attended a meeting that was the first to address an upcoming experiment. One participant took the view that the first item on the agenda ought to be the articulation of the experiment's goals. Another took the view that the most fundamental aspect of the experiment was the number of people who would be involved, and that philosophical discussions, such as that regarding the experiment's goals, could wait until after the important questions had been answered. These individuals ended up in a shouting match, which was won (on the basis of rank, as well as shouting ability) by he who advocated starting with the number of people. After the experiment was complete, the price of poorly-articulated goals was paid in full.

learn to use its equipment, and how to create any records that it is expected to keep.

Equipment

If the experiment is designed to test a particular piece of equipment, that piece of equipment needs to be available, ready, and working *prior to* the beginning of the experiment so that the participants can receive training on it. There is considerable evidence that when the project managers in charge of developing experimental equipment find out that it is to be used in an experiment, they conclude that the experiment will be the test of the equipment, and that therefore they are absolved of testing it prior to delivery. This attitude must be detected in advance of the experiment and corrected.

If a piece of experimental equipment doesn't work in the experiment, one can at least report the fact. But if the experiment is designed to test a concept, then it is likely that one or more future pieces of equipment will be represented by surrogates, and these surrogates have to work or there will be no experiment. The saying, "it's OK to fail" applies to the creation of prototypes, not to the creation of surrogates.

Finally, any instrumentation must be guaranteed to work, because without it, data will be lost and the value of the experiment reduced, possibly to zero.

Method

This large category includes everything from the details of how to adjudicate non-MILES weapon shots to what statistical approach will be taken when analyzing numerical data produced by the experiment. Some of these topics have been addressed elsewhere in this document, and/or in the companion piece, *The Art of Military Experimentation*.

At the "template" level, the important idea is that the designers of the experiment must ensure that their methods are matched to their goals. Points that figure prominently in the hypothesis must be represented with fidelity in the experiment's model—be it a computer model, an exercise-like mock combat with real troops and MILES, or anything else. Conversely, weak points in the simulation (e.g., the MILES weapons' inability to shoot through walls) must be assessed for the potential to produce distortions in the experiment's outcome. If a quantitative result is expected, an analyst should be consulted to provide assurance that the amount of experimentation (in effect, the "sample size") is sufficient to support the desired level of quantitative accuracy.

Refinement

The previous section mandates what amounts to a *methodology audit*, which may well result in decisions to revisit nearly every aspect of the whole experiment. This refinement is a healthy step, not to be confused with wasted motion.

Another source of requirements for refinement is outside influences, which may impose limitations on what experiment can be done; these limitations can change, and then part or all of the experiment must be reconsidered in light of the new limitations.

After these discussions of refinements and limitations, there may ensue a discussion of whether the experiment is still worth doing. In this regard, exercises set the lower limit: an experiment cannot be worthwhile if it gives the analysts less information than they could get by observing a regularly-scheduled exercise. Note that this standard sets only a lower limit: possibly there are experiments that satisfy this lower-limit criterion, but are nonetheless not worth doing.

Conduct of the experiment

During the conduct of the experiment, revisions may again be necessitated by changing, unforeseen, or inadequately appreciated circumstances. These can include weather, or the restrictions imposed by outside entities. "Inadequately appreciated circumstances" can, on

rare occasion, also include the substance of the experiment itself: it can get underway, only to have the experimenters realize that the nature of the experiment differs from what they had expected.

Care must be exercised when making adaptations during the experiment. In an experiment with a baseline case and an experimental case, alteration of either one can necessitate alteration of the other. Also, one must avoid the appearance (and the reality!) of revising the experiment so as to obtain a preferred outcome.

Report writing—analysis and assessment

Report-writing on the part of analysts is treated at some length in the companion piece, *The Art of Military Experimentation*, as well as in a previous section of the present paper.

Here, it will suffice to repeat that if the experiment is to be of any worth, an analysis report must be written, signed out, published, and distributed if it—and the experiment as a whole—is to be of any worth.

Having read the analyst's report (or, more likely, a final draft thereof), the military members of the experimentation team, assisted by other uniformed subject matter experts as needed, ought to convene to make an assessment of the experiment. In all likelihood, they will want to begin by hearing the analyst(s) give a briefing based on the analysis report.

Based on this briefing, and on their reading of the (draft) report, the uniformed military people can draw conclusions as to the *meaning* and *implications* of the report. In large measure, the role of the military people is not to draw conclusions that the analysts *couldn't* draw, it is to draw conclusions that the analysts *wouldn't* draw. For example, suppose that an LTA results in the finding that a new bomb-aiming system as a CEP of 3 inches. Everybody knows that this is a great improvement over the existing technology, but an analyst would hesitate to say so unless the LTA had included a baseline case. Analysts will also be reluctant to make judgments regarding risk to life and limb, understandably feeling that it is not their place to do so. Finally,

the analysts' recommendations for future work and those of the military people are based on such different perspectives that each must be presented.

The assessment effort should result in a report of its own, separate from that of the analysis effort. This report, too, must be written, signed out, published, and distributed if it—and the experiment as a whole—is to be of any worth.

Iteration

Even if the experiment turns out to be a success—perhaps *especially* if it does—there may well be reason to repeat it, or to do something very similar to it. Again, this is not a sign of waste or weakness if it causes a worthwhile increase in knowledge, greater than could be had by observing a training exercise.

Closure

At MCWL, an administrator noticed that experimentation with ACASS (which eventually became a MCWL success story) seemed to be going on indefinitely. "How will you know when you're finished?" he asked, and although the question arose from frustration and apprehension as much as puzzlement, it is a good one.

A good answer is, "When we know what works, and have documented it." This answer can be applied to tactics as well as to hardware: the Lab's Project Metropolis, for example, set out to develop improved urban tactics, and was finished with that project when the improved tactics had been developed, codified, taught to Marines, and shown in a final round of experimentation to lead to fewer casualties than did the urban tactics being taught theretofore.

Bad answers would include, "When the person who thought of it leaves," "When the new General comes," and "When people get tired of it."

Of course, there also needs to be room for an answer of "When we've decided it was a bad idea after all," but this decision needs to be reached carefully, and not as a proxy for any of the bad answers.

The time from start to closure is almost certain to be longer than any uniformed person's tenure in the experimenting organization, leading to the need for a good turnover process, for written reporting, and for constancy of purpose not only at the project level, but also at the level of the experimenting organization as a whole.

Glossary

Analysis	The process by which the ground-truth-level resulting from reconstruction is turned into knowledge, especially knowledge regarding the question(s) around which the experiment is structured.
AWE	Advanced Warfighting Experiment—
Assessment	A written product resulting from military officers' discussion of an experiment's analysis report.
Base case	That part of an experiment in which the equipment, TTPs, or other experimental variables are adjusted to correspond to present conditions, or some other conditions that are taken for granted.
Battlecruiser	An illfated innovation in which the armor customarily associated with battleships was sacrificed in the interest of speed.
BluFor	In an LOE or AWE, those forces representing American forces.
CEP	Circular Error Probable—in a situation involving some form of shooting, with all shots directed at the same target, the CEP is the radius of the circle in which half the impact points are expected to appear. (Cf. *DoD Dictionary of Military and Associated Terms*: "the radius of a circle within which half of a missile's projectiles are expected to fall.") Note that this definition does not entail an assumption that the pattern of errors has circular symmetry. See also McCue, 2002.
Data	Plural of datum, an atom of information.
Dönitz	Admiral of German submarines during WW II.
Demonstration	Degenerate case of an experiment in which the experimental event can have only own outcome, and thus can point to only one answer.

Eclectronic	An electronic assembly composed of components drawn from multiple sources
Exercise	"A military maneuver or simulated wartime operation involving planning, preparation, and execution. It is carried out for the purpose of training and evaluation"—*DoD Dictionary of Military and Associated Terms*.
ExFor	In an LOE or AWE, those forces principally benefiting from the experimental equipment, tactics, or concept of operations. Normally the same as the Blufor.
Experiment	The unification of a question (to which multiple answers are possible), an event (that can result in different outcomes) and a matching of the outcomes and the answers.
Experimental Case	That part of an experiment in which the equipment, TTPs, or other experimental variables are intentionally adjusted to a counterfactual state that is to be compared to the normal state.
Fires	Fire support: "Fires that directly support land, maritime, amphibious, and special operation forces to engage enemy forces, combat formations, and facilities in pursuit of tactical and operational objectives"—*DoD Dictionary of Military and Associated Terms*.
God gun	Handheld MILES master controller.
GPS	Global Positioning System—"A satellite constellation that provides highly accurate position, velocity, and time navigation information to users"—*DoD Dictionary of Military and Associated Terms*.
Hypothesis	An idea put forward for comparison against real-world data, especially those gleaned from a future experiment.
Hotwash	(Sometimes conflated with "hogwash.") An all-hands meeting immediately following an experiment, to capture and compare first-hand first-impressions.

Hunter Warrior	CWL's first major project (and eponymous March 1997 AWE), exploring a concept of expeditionary operations in which small teams fought the enemy entirely through the use of supporting fires, applied using several items of information technology. These were so advanced as to require representation by surrogates. Though ill-received in many quarters, the Hunter Warrior concept of operations strongly resembled that used in 2002 by US forces in Afghanistan.
Model	A mental, physical, and/or computational construct for exploring the unreal.
Midway	U.S. v. Japan aero-naval battle in early June 1942, the dramatic turning point in WW II's Pacific campaign.
MILES	(Modular Integrated Laser Engagement System)—A system that provides surrogate small arms fire via a laser attached to the user's service weapon, and vest bearing photocells.
LOE	Limited Objective Experiment—A mid-size experiment, including ExFor, an OpFor, and a considerable level of free play on at least one side.
LTA	Limited Technical Assessement—a single-purpose experiment, somewhat similar to a field test, but more flexibly conducted.
O/C	Observer/Controller—a member of an experiment's staff who performs both data collection and experiment control functions, usually at a low level and focusing on one small group of participants.
Operations Research	"The analytical study of military problems undertaken to provide responsible commanders and staff agencies with a scientific basis for decision on action to improve military operations. Also called operational research; operations analysis"—*DoD Dictionary of Military and Associated Terms*.

OpFor	Opposing Forces—those forces in an LOE or AWE that oppose the ExFor.
Prototype	"A model suitable for evaluation of design, performance, and production potential"—*DoD Dictionary of Military and Associated Terms*.
PTS	Precision Targeting System—a device combining rangefinder, compass, GPS receiver, and computer, capable of measuring the location of a visible target.
Reconstruction	A complete, fact-based, time-synchronized, deconflicted, and meaningful account of what actually happened.
Roleplayers	Experiment participants other than the ExFor and the OpFor; these often represent bystanders, insurgents, refugees, hostages, or other civilians.
Schema	A diagram that explains an idea.
SCUD	NATO codename for a widely proliferated, Soviet made short-range ballistic missile, numbered SS-1 by NATO.
Serendipity	The unexpected discovery of a pleasant fact.
Simulation	A model that produces a time sequence of states.
Simunitions®	Dye-filled 9mm rounds, fired from a modified service weapon, used in conjunction with eyewear and other protection to create a non-injurious means of simulating firefights for purposes of training or experimentation.
Surrogate	A model not suitable for evaluation of design, performance, and production potential.
Test	(noun) A tightly controlled experiment, usually on a piece of equipment, that seeks to measure performance in one particular dimension, or in a small set of well-defined dimensions.

Theory	"Systematically organized knowledge applicable in a wide variety of circumstances, especially a system of assumptions, accepted principles, and rules of procedure devised to analyze, predict, or otherwise explain"—*Webster's II New Riverside University Dictionary.*
Thought Experiment	A mental act in which an experimental situation is envisioned, with no intent to carry it out, and the implications of each possible outcome are contemplated in turn.
U-boat	WW II German submarine
Urban Warrior	MCWL's major project (and eponymous March 1999 AWE) following Hunter Warrior,
Wargame	"A simulation, by whatever means, of a military operation involving two or more opposing forces using rules, data, and procedures designed to depict an actual or assumed real life situation"—*DoD Dictionary of Military and Associated Terms.*
.	
Wolf pack	Group of submarines coordinated by a commander ashore.
Wotan	The Norse god of wisdom and logic, latterly associated with war and battle. His name survives in our word, "Wednesday."

List of acronyms

(See also Glossary)

ACASS	Advanced Close Air Support System
ARMVAL	Advanced Antiarmor Vehicle Evaluation
AWE	Advanced Warfighting Experiment
CEP	Circular Error Probable
CNA	Center for Naval Analyses
CNAC	CNA Corporation
CWL	Commandant's Warfighting Laboratory; the original incarnation MCWL
ExCon	Experiment Control
ExFor	Experimental Force
FO	Forward Observer
GPS	Global Positioning System
IGRS	Integrated GPS Radio System
JCATS	Joint Conflict and Tactical Simulation
LOE	Limited Objective Experiment
LTA	Limited Technical Assessment
M&S	Modeling and Simulation
MCWL	Marine Corps Warfighting Laboratory
MILES	Modular Integrated Laser Engagement System
MOUT	Military Operations on Urbanized Terrain
NATO	North Atlantic Treaty Organization

O/C	Observer Controller
OK	[origin obscure]
OODA [loop]	Observe-Orient-Decide-Act [loop]
OpFor	Opposing Force
pH	Potential of Hydrogen
Ph.D.	Philosophiae Doctoris
PTS	Precision Targeting System
RM	Royal Marines
SAW	Squad Automatic Weapon
SCUD	See Glossary; SCUD is not an acronym.
SIMNET	Simulation Network, precursor of the Close Combat Tactical Trainer
SMAW	Shoulder-Launched Multipurpose Assault Weapon
SPMAGTF(X)	Special Purpose Marine Air-Ground Task Force (Experimental)
SRI	Stanford Research Institute, former name of the company known as SRI
TEWT	Tactical Exercise Without Troops
TTP	Tactics, Techniques, and Procedures
UAV	Unmanned Air Vehicle
UCATS	Universal Combined Arms Targeting System
USMC	United States Marine Corps
VIP	Very Important Person
WMD	Weapon of Mass Destruction
WW II	World War II

Bibliography

(See also separate bibliography of MCWL analysis reports.)

David S. Alberts and Richard E. Hayes, *Code of Best Practice for Experimentation*, Washington DC, DoD Command and Control Research Program, 2001.

Thomas B. Allen, *War Games*, New York, McGraw-Hill, 1987, Berkley paperback reprint 1989.

Percy Bridgman, *The Logic of Modern Physics*, New York, Macmillan, 1922.

Bernard Brodie, *Sea Power in the Machine Age*, Princeton, Princeton University Press, 1941.

Bernard Brodie, *A Guide to Naval Strategy*, Princeton, Princeton University Press, 1944.

Bernard Brodie and Fawn M. Brodie, *From Crossbow to H-Bomb* (revised and enlarged edition), Bloomington, Indiana University Press, 1973.

James S. Corum, *The Roots of blitzkrieg*. Lawrence, The University Press of Kansas, 1992.

Department of Defense, *DoD Dictionary of Military and Associated Terms*, Joint Publication 1-02, formerly JCS Publication 01, now available on the World Wide Web at http://www.dtic.mil/doctrine/jel/doddict/index.html .

Michael DiMercurio, *Voyage of the Devilfish* (novel), New York, Onyx, 1993.

David Hackett Fischer, *Historians' Fallacies: Toward a Logic of Historical Thought*, New York, Harper Torchbooks, 1970.

Bradley A. Fiske, *The Navy as a Fighting Machine*, 1916, second edition 1918, reprint (cited here) Annapolis, Naval Institute Press, 1988. Introduction and notes by Wayne P. Hughes, Jr.

John M. Ford, "Fugue State" (story), *Under the Wheel: Alien Stars Volume III*, Baen 1987.

Christopher R. Gabel, *The U.S. Army GHQ Maneuvers of 1941*. Washington, The Center for Military History, United States Army, 1991 (U.S. Government Printing Office).

George Francis Gause, *Vérifications expérimentales de la théorie mathématique de la lutte pour la vie*. Paris, Hermann, 1935. Translated as, *The Struggle for Existence*, Dover reprint, 1971.

William H. Honan, *Visions of Infamy*, New York, St. Martin's, 1991.

Jeter A. Isely and Philip A Crowl, *The U.S. Marines and the Art of Amphibious War*. Princeton, Princeton University Press, 1951.

Herman Kahn, *On Thermonuclear War*, Princeton, Princeton University Press, 1961.

Helen Karppi and Brian McCue, "Experiments at the Marine Corps Warfighting Laboratory," *Marine Corps Gazette*, Volume 85 number 9 (September, 2001).

Richard Kass, "Joint Warfighting Experiments: Exploring Cause and Effect in Concept Development," *Phalanx*, Volume 35 number 1 (March, 2002).

John Keegan, *The Price of Admiralty*, New York, Viking, 1988.

Bernard Osgood Koopman, *Search and Screening*, New York, Pergamon, 1980, an updated version of the OEG Report #56, of the same title, published in 1946 by CNA's predecessor organization.

Marine Corps Warfighting Laboratory, *Tactical Instrumentation*, X-File 3-35.13, part of the series on Military Operations in Urban Terrain, 1999.

Brian McCue, "A Chessboard Model of the Battle of the Atlantic," *Naval Research Logistics*, forthcoming.

Brian McCue, *Wotan's Workshop: Military Experiments Before the Second World War*, Center for Naval Analyses, Occasional Paper, October 2002.

Brian McCue, *Estimation of the Circular Error Probable (CEP)*, CIM D0005820.A1, Center for Naval Analyses, April 2002.

Philip M. Morse and George E. Kimball, *Methods of Operations Research*, Operations Evaluation Group Report 54, Washington, 1946, available from the Center for Naval Analyses. The 1951 edition, differing slightly in content and pagination, has recently been reprinted by the Military Operations Research Society.

Peter P. Perla, *The Art of Wargaming*. Annapolis, United States Naval Institute, 1990.

Project Metropolis, *Military Operations on Urbanized Terrain (MOUT) Battalion Level Experiments, Experiment After Action Report*, Quantico, Marine Corps Warfighting Laboratory, 2001.

Richard Rhodes, *The Making of the Atomic Bomb*, New York, Simon & Schuster, 1986.

Donald H. Rumsfeld, *Quadrennial Defense Review Report*, September 2001.

Michael Schrage, *Serious Play*, Boston, Harvard Business School Press, 2000.

Roy A. Sorenson, *Thought Experiments*, Oxford, Oxford University Press, 1992.

R. H. S. Stolfi, *Hitler's Panzers East: World War II Reinterpreted*, University of Oklahoma, 1992.

T. Strentz, "The Stockholm syndrome: law enforcement policy and ego defenses of the hostage," pp. 137-50 in *Forensic Psychology and Psychiatry*, eds. F. Wright *et al.*, New York 1980.

Sun Tzu Wu, *The Art of War*, tr. Samuel B. Griffith, Oxford, Oxford University Press, 1963.

Robert H. Thompson, "Lessons Learned from ARMVAL," *Marine Corps Gazette*, July, 1983.

Linda D. Voss, *A Revolution in Simulation: Distributed Interaction in the '90s and Beyond*, Arlington, Pasha Publications, 1993.

Webster's II New Riverside Dictionary, Boston, Houghton Mifflin, 1984.

Andrew Wilson, *The Bomb and the Computer: Wargaming From Ancient Chinese Mapboard To Atomic Computer*. New York, Dellacorte, 1968.

E. Bright Wilson Jr., *An Introduction to Scientific Research*, Mineola, Dover, 1990, reprint of McGraw-Hill second edition, 1952.

Roberta Wohlstetter, *Pearl Harbor: Warning and Decision*, Stanford, Stanford University Press, 1962.

MCWL (and CWL) Analysis and Assessment Reports, 1997-2002

(Chronological Order; undated documents listed by the dates of the events they describe)

Dwight Lyons, Victor Dawson, Mike Johnson, Helen Karppi, Ken Lamon, Duncan Love, Fred McConnell, Bill Morgan, Tom O'Neill, John Reynolds, Janette Stanford-Nance, Robert Sullivan, LDCR Dave Cashin (USN), Maj Bill Cabrera USMC, Maj Robin Gentry USMC, Capt Wade Foster USMC, Capt Nick Slavik USMC. *Hunter Warrior Advanced Warfighting Experiment Reconstruction and Operations Training Analysis Report,* Aug 1997

John Reynolds. *Limited Technical Assessment 120 mm Armored Mortar System*, 21-28 September 1997

Brian McCue. *Urban Warrior Limited Technical Assessment; Rigid Foams,* 30-31 Dec 1997

Hunter Warrior Workshop Assessment, 27 Mar 1998, 15-16 April 1998

Dwight Lyons, John Goetke, Ken Lamon, Fred McConnell, Brian McCue, John Reynolds, Doug Turnbull, Capt Kevin Brown USMC. *Urban Warrior Limited Objective Experiment 1 Reconstruction and Operations Analysis Report*, 8 May 1998

Dwight Lyons, Maj Kevin Brown, USMC, John Goetke, Ken Lamon, Fred McConnell, Brian McCue, Tom O'Neill, John Reynolds, Doug Turnbull, and Will Williamson, *Urban Warrior Limited Objective Experiment 2 Reconstruction and Operations Analysis Report,* 15 Nov 1998

Brian McCue. *Blackhawk Down Non-Lethals Experiment,* Dec 1998

UK LOE CSS Riverine Experiment, May 1998

Brian McCue, *1st Responder LTA,* Jun 1998

MSgt. Theodore Bogosh USMC, Mike Brennan and Maj. Wade Johnson USMC. *Capable Warrior Limited Objective 3 Experiment Report*, Feb 1999

Capt Mark A. Sullo. *Assessment Report for Combat Decision Range*, 20 May 1999

Dwight Lyons, Brian McCue, Fred McConnell, Capt. Wade Foster USMC, John Reynolds, John Goetke and Will Williamson. *Urban Warrior Limited Objective Experiment 3 17-20 July 1998*, 7 Jul 1999

Dwight Lyons, Michael Brennan, Julius Chang, John Goetke, Renee Guadagnini, Mike Johnson, Brian McCue, Fred McConnell, John Reynolds, Gary Phillips, Doug Turnbull, Maj. Paul Curran USMC, and Capt Wade Foster USMC. *Urban Warrior Culminating Phase I Experiment (CPE) 12-19 Sept 1998 Reconstruction and Operations Analysis Report*, 7 Jul 1999

Dwight Lyons, Michael Brennan, Julius Chang, Yana Ginburg, John Goetke, Helen Karppi, Janet Magwood, Brian McCue, Fred McConnell, Joe Mickiewicz, John Reynolds, Doug Turnbull, Maj Bill Inserra USMC, Maj Dave Kunzman USMC, Capt Wade Foster USMC and Capt Chris Goodhart USMC. *Urban Warrior Advanced Warfighting Experiment 12-18 March 1999 Analysis Report*, 23 Jul 1999

Brian McCue and John Goetke. *Urban Warrior Aviation Limited Technical Assessment*, 20 Dec 1999

Helen Karppi, R. Lavar Huntzinger, Janet Magwood, Brian McCue and John Reynolds. *Capable Warrior Limited Objective Experiment 4 Analysis Report*, 9 Dec 1999

John Goetke, R. Lavar Huntzinger, Helen Karppi, Janet Magwood, Brian McCue. *Capable Warrior Limited Objective Experiment 6 Analysis Report*, 28 February—18 March 2000

Brian McCue. *Advanced Close Air Support System Analysis Report*, Aug 2000

Brian McCue. *Capable Warrior 120 mm Mortar and Laser Targeting Device. Jun 2000*, 18 Jan 2001

Lavar Huntzinger, R*eport of Analysis Findings for Project Tun Tavern*, 30 Jan 2001

Brian McCue. *Advanced Close Air Support System Analysis Report Aug 2000*, 30 Jan 2001

Project Metropolis Military Operations on Urbanized Terrain (MOUT) Battalion Level Experiments, Experiment After-Action Report, Feb 2001

R. Lavar Huntzinger, John Goetke, Helen Karppi, Brian McCue. *Millennium Dragon Advanced Warfighting Experiment 8-11 September 2000 Analysis Report*, 14 Feb 2001

Brian McCue. *Sensor-to-Shooter LTA Report Dec 2000*, 27 Feb 2001

Project Metropolis Phase I Report, 29 Mar 2001

Brian McCue. *Lightweight Personnel Detection Device & Park Watch Limited Technical Assessment & Limited Objective Experiment Report*, Apr 2001

Brian McCue, *GPADS LTA Quicklook Report*, Apr 2001

Brian McCue. *RSTA Wargame Report*, May 2001

Brian McCue. *Summary of Experimentation with Precision Targeting Systems*, Jul 2001

Brian McCue. *ACASS and MELIOS (Advanced Close Air Support System and Mini Eyesafe Laser Infrared Optical Set) LTA*, Aug, 2001

Brian McCue. *Zinc-Air Battery Gunshot Damage Limited Technical Assessment*, Sep 2001

Yana Ginburg. *Zinc Air Battery Assessment Twenty-nine Palms, CA; Performance During CAX 9 & 10*, Oct 2001

Helen Karppi, Jennifer Ezring, John Goetke, Richard Laverty, Dwight Lyons, Brian McCue and Wendy Trickey. *Kernal Blitz (Experimental)/Capable Warrior AWE Analysis Report*, 1 Oct 2001

Marian E. Greenspan. *Report on the Multi-Function Rifle Scope Demonstration*, 27 Nov 2001

John Goetke. *Compilation of MCWL ISR Experimentation Findings*, Nov 2001

Brian McCue. *Autonomous GPS-Guided Aerial Resupply Systems: A Review of MCWL Limited Technical Assessments of BURRO and GPADS*, Dec 2001

Helen T. Karppi. *Mortar Ballistic Computer Limited Technical Assessment Analysis Report*, 3 Dec 2001

Yana Ginburg. *Dragon Runner LTA Report*, 19 Dec 2001

Precision Target Acquisition, Mobile Limited Technical Assessment Analysis Report (Initial Testing), 14 Dec 2001

Capt. James Stone USMC, Robert Bickel, John Goetke & SSgt. Kevin Ashley USMC, *Bulk Fuel Company, 2nd Engineer Support Battalion Limited Objective Experiment*, 7 Feb 2002

Helen T. Karppi. *Universal Combined Arms Targeting System (UCATS) and Precision Target Acquisition Mobile (PTAM) Digital Communications Limited Technical Assessment Analysis Report, Fort Bragg 2002*, Apr 2002

Capt. James Stone USMC, Robert Bickel & John Goetke, *Battle Griffin 02 Limited Objective Experiment Quicklook Report*, 10 Apr 2002

Randy Bickel and John Goetke, *High Speed Vessel Interim Experimentation Report: Period of 18 October 2001—17 March 2002*, May 2002

Wendy R. Trickey and Brian McCue. *ACASS and MELIOS DPI Limited Technical Assessment*, May 2002

Lt Col. N.J. Cusack RM, Capt. J.J. Falcone USMC, Helen T. Karppi. *Urban Blockage: Line of Sight Angles to Airborne and GEO-Stationary Satellite Platforms in the Urban Environment*, 6 May 2002

Marian E. Greenspan. *Mobile Fire Support System Limited Technical Assessment August 27-29, 2001*, 17 May 2002

Marian E. Greenspan. *Multi-Function Rifle Scope Limited Technical Assessment*, 17 May 2002

Major James Stone, Capt Michele Kane, Robert Bickel & John Goetke. *Millennium Challenge 02 Limited Objective Experiment: Joint High Speed Vessel Analysis Report*, 16 Aug 2002

Yana Ginburg. *Urban Eyes North Little Rock Urban Ground Reconnaissance Experiment 9-22 February 2002,* Sep 2002

Jennifer Ezring. *Individual Combat Identification System (ICIDS) Limited Technical Assessment (LTA) Report,* MCWL Analysis Report 02-10, October 2002

Brian McCue. *Initial Investigation of an RDO STOM TO, TE, and CONOPs: Olympic Dragon 04 Wargame 1 and Modeling and Simulation Analysis Report,* November, 2002. *Assessment Report, Olympic Dragon 04 Wargame 1 and Modeling and Simulation,* Nov 2002

List of figures