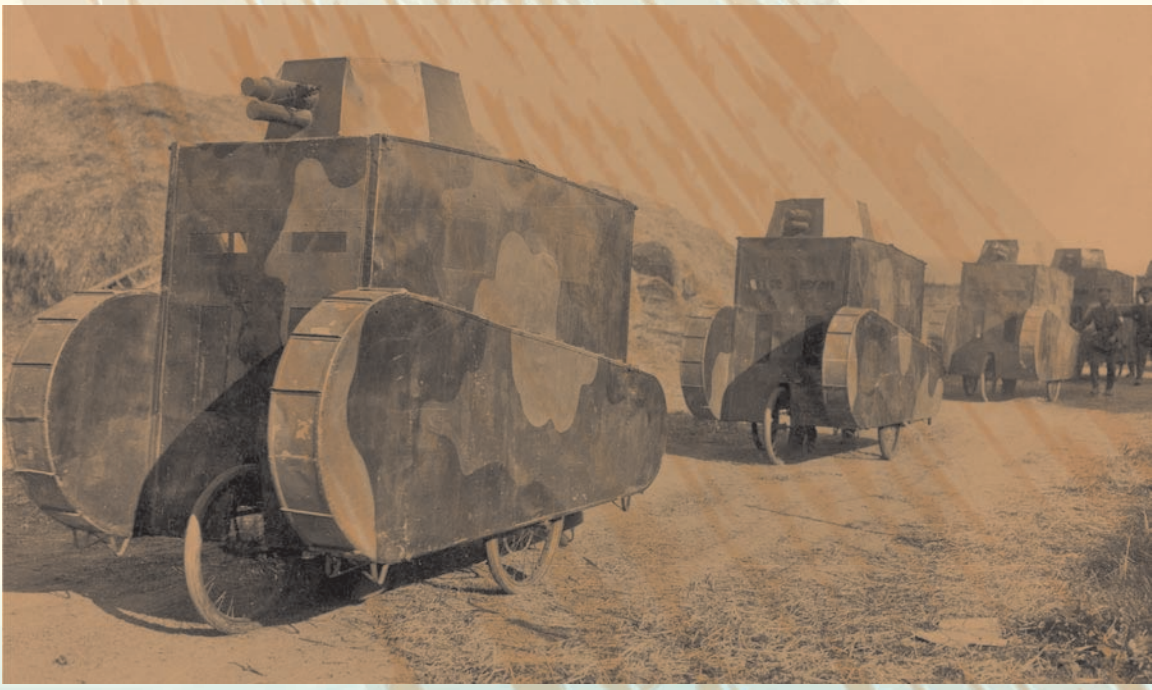
Wotan's Workshop

Military Experiments before World War II



Brian McCue

Cover: *Mock tanks constructed of cardboard were used during maneuvers on 9 September 1928, at Uckermark, the German training base.* (Getty Images)

Wotan's Workshop

Military Experiments before World War II

by
Brian McCue

The views expressed in this book do not necessarily represent the opinions of the U.S. Department of the Navy or the U.S. Marine Corps.

CNA Corporation
4825 Mark Center Drive
Alexandria, Virginia 22311
www.cna.org

Marine Corps University Press
3078 Upshur Avenue
Quantico, Virginia 22134
www.tecom.usmc.mil/mcupress

Wotan's Workshop: Military Experiments before World War II
© 2013 by CNA Corporation. All rights reserved.
Published by Marine Corps University Press, Quantico, VA.

About the title: Wotan, the Norse god of wisdom and logic, was also latterly associated with war and battle. His name survives in our word *Wednesday*.

Title page photo: *A column of mock tanks—with wood and canvas armor shapes mounted on light vehicles—participated in German military maneuvers in 1928.* (Imperial War Museum/Q71388)

Contents

Introduction.....1

Themes.....3

The Structure of Experimentation
Models, Modeling, and a Paradox
Surrogates
Artificialities
Theory, Hypothesis, and Serendipity
"All's fair in love and war"—What about in Experimentation?

Three Idiosyncratic Nonservice Efforts.....11

G. F. Gause's Bugs
Hector Bywater's Wind-Up Ship Models
Fletcher Pratt's Naval War Game
Observations on the Idiosyncratic Efforts

The U.S. Prepares for World War II.....19

The Fleet Problems, 1923–40
The Experiments of Billy Mitchell
Major Pete Ellis and USMC Interwar Experimentation
The Army's Louisiana Maneuvers
Pacific Fleet Fighter Direction Officer School
Observations on the U.S. Experiments

Germany Prepares for World War II.....41

The German Army's Experiments with Blitzkrieg
The German Navy's Experiments with Wolf Packs

A Limited Technical Assessment
Observations on the German Experiments

Overall Observations.....53

Recapitulating the Themes
The Experiments' Points of Similarity
The Paradox of Modelling, Resolved

Notes 59

Bibliography.....63

Introduction

Over the last few years, military experimentation has attained unprecedented salience. The secretary of defense, the chairman of the Joint Chiefs of Staff, and Congress all called for increased efforts in military experimentation in the early 2000s.[1] Because this support is relatively new, military experimentation is in the anomalous position of being popular, yet unfamiliar. The resulting lack of understanding of the nature of military experimentation has acted to the detriment of the various efforts now ongoing at the service and joint levels. The outward resemblance of military experiments to the more familiar exercises and field tests, and the outward resemblance of the experiments' *technology surrogates* to prototypes, have only served to deepen the misunderstanding.

An attempt to better understand military experimentation by detailed examination of some of today's efforts would be hampered by the need for a considerable background in the technologies that the experiments address. There is also room for concern that discussion of present-day efforts would be seen primarily as praise or criticism of the particular efforts, and thereby rendered useless as a vehicle for discussion of experimentation itself.

An alternative way to strive for a better understanding of military experiments is through a set of historical examples. Today's impetus for military experimentation has arisen largely because we have experienced a large amount of technological change during a protracted period of peace. The period from the demise of the Soviet Union to the present is longer than two decades, and they have been two decades of dramatic technological progress in sensors, materials, and communications—all areas of undoubted military potential—and above all in computation, whose potential applicability to warfare remains a topic of heated discussion. For these reasons, the present period is often compared to the 1918–39 interwar period, whose dramatic developments in

such areas as radio and aviation and considerable evolutionary improvements in submarines and tanks led to great speculation and argument as to the future of warfare and the applicability of technology to it. Then as now, the military resorted to experimentation as a means of discovering the implications of the new technologies.

To increase understanding of the process of military experimentation, this work examines some of these experiments; today, the technologies need no introduction, and the Second World War provides hindsight through which the experiments and their findings can be viewed. The efforts described here have been chosen as an instructive set, not an exhaustive one.

This being a work of analysis, rather than history, secondary sources—as well as primary sources in the form of some participants' memoirs—have been used freely. This paper was originally part of CNA's project on military experimentation.

Themes

A number of themes will run through the cases examined here. These themes are important in any consideration of military experimentation, and in fact a major goal of this work is to illustrate them through the use of the pre–World War II examples.

The Structure of Experimentation

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*, usually prestated, between the outcomes of the event and the answers to the question.

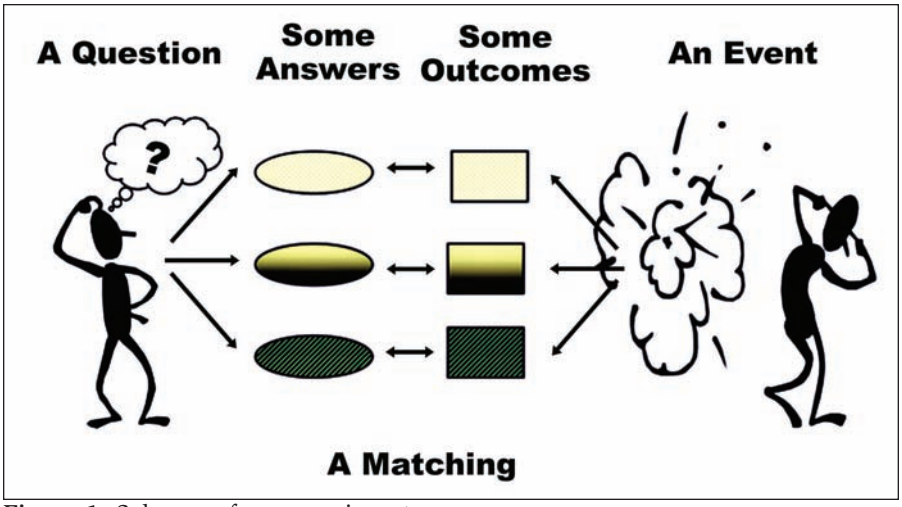


Figure 1: Schema of an experiment

A familiar example is the use of litmus paper to test the pH (acidity) 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 *prestated matching* is that the color blue indicates an acid whereas the color red 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, or any of the other 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.

Elsewhere I have referred to military experimentation as an "art"—if it is so scientific, why is it an art?

The reason is that in military experimentation* 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.
- The matching between event outcomes and answers to the question is preassigned.

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*, not an experiment. Demonstrations and explorations can be of value, but they are not experiments.

Models, Modeling, and a Paradox

Today, models of warfare are automatically assumed to be computer models. Many people, especially those in uniform, additionally assume computer models of warfare to be of questionable validity and tend to reject findings based on them. To them, field

*And, probably, in all other kinds as well, perhaps excluding only the most "scientific" and well-funded.

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.

Models are an apparent violation of the computer-age principle, "Garbage in, garbage out." A related aphorism holds that if you add a tablespoon of fine wine to a barrel of garbage, the barrel still contains garbage, but if you add a tablespoon of garbage to a barrel of fine wine, a conversion does occur, and the barrel becomes a barrel of garbage. A model being, by definition, something other than that which is modeled and therefore imperfect in some respect, this logic would suggest that in no case can the results of a model be trusted—there is always the possibility that some important feature of the problem has been omitted, and thus that the model and its output are garbage. This line of reasoning would apply to military experiments in the field or at sea just as much as it would apply to computer models, and it can be a compelling counsel of despair.

Yet military experiments, including most of those described in this paper, have a record of success. Each for his or her own reason, the analyst and the military officer may well balk at the notion that the whole can be more correct than the parts. To the analyst, very possibly trained in physical science, the idea seems to run counter to the whole notion of reductionism that has powered science since the Age of Enlightenment. To the military officer, the idea seems to run counter to the hierarchical notion embodied in the military structure itself, and more recently in such constructs as the Universal Military Task List, that the way to get the whole job done correctly is to build it up correctly from correctly done subtasks. One of the goals of this paper, to which we shall return in the "Overall Observations" section at the end, is to resolve this apparent paradox.

Surrogates[2]

In field or fleet experiments, as in exercises, technological *surrogates* often replace the genuine items of equipment that would be used in battle. In some instances, the motivation for using a surrogate is that the real thing would be expensive to use, or dangerous. A typical example would be the use of flour-filled bags as hand grenades—they mark recipients as casualties without causing injury. In other instances, a surrogate must be used because the real thing is not yet available. Examples would include the use of automobiles as tanks, or the use of destroyers as submarines, in a country that does not yet have any tanks or submarines but is working on building some and would like to get a head start on studying their employment.

Artificialities

Many considerations (such as cost, danger, the limits of the physically possible, and shortcomings of any surrogates) will constrain a military experiment from being a complete faithful representation of the combat situation under study. The points of difference are generally known as *artificialities*, and it is the responsibility of the experimenter to limit the degree to which these cause the wrong conclusion to be drawn. This responsibility is much greater than any to limit the artificialities themselves. Indeed, any amount of artificiality can, theoretically, be withstood as long as there is a correct matching of outcomes (however artificial they may be) to answers. Some of the efforts considered in this paper feature rather high levels of artificiality, but succeeded nonetheless.

How can one tell that one is successfully coping with artificiality? An initial impression, one way or the other, is not enough. As we will see, Major General William "Billy" Mitchell's famous ship bombing experiment, realistic although it appeared through its use of a real ship and real airplanes, may have suffered so badly from just one or two artificialities that an incorrect conclu-

sion was drawn. Conversely, most people's initial reaction to the Pacific Fleet Fighter Direction Officer School's tricycles—functioning as airplanes—was probably mirth, but the school was able to handle its panoply of artificialities and become a source not merely of training, but of new knowledge regarding the emerging topic of fighter direction.

The answer lies in the possession and use of a *theory*.

Theory, Hypothesis, and Serendipity

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,"[3] as in "music theory" or "game theory."

Especially in the military, the widespread derogatory use of the term in the first two senses has detracted not only from its use in the third sense, but may even have deterred some people from the activity described therein. 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),[4] or both.[5]

Yet, as Kurt Lewin* said, "There is nothing so practical as a good theory."

There have been a few successful theories of warfare that fit

*Psychologist and inventor of the now-commonplace term "group dynamics."

the third definition above. Frederick W. Lanchester's theory of attrition warfare, expressible in a set of coupled differential equations, has probably seen at least as much misuse as use, but it has some explanatory power and in any case is better than the set of vacuous platitudes that Lanchester was trying to displace. CNA's predecessor organizations, the wartime ASWORG (Anti-Submarine Warfare Operations Research Group) and the postwar OEG (Operations Evaluation Group), developed a theory of "search and screening."[6]

Such formal theories as these, or even rules of thumb such as the "3:1 ratio of offense to defense," allow formulation of an experiment in a way that takes artificialities into account and restricts the harm they can do. The instances of pre–World War II experimentation shown in this paper will provide examples.

It is useful to distinguish between a *theory*, the intellectual framework of understanding in which the inquiry takes place, and a *hypothesis*, the proposition whose truth is to be tested by the experiment. As mentioned previously, there are forms of experimentation other than hypothesis testing (measurement, for example), but the test of a hypothesis is a particularly simple experimental design because there are only two answers—the hypothesis is either true or false. Therefore only two outcomes need to be distinguished—outcomes proving the hypothesis, and outcomes disproving it.*

In some cases, the hypothesis being tested is really a statement that the underlying theory is true. In that case, the experiment may be called a *crucial experiment*.

Serendipity refers to unexpected, and welcome, discovery. Especially when entering a particularly new area of inquiry, an experiment may result in an unforeseen outcome. Being unforeseen, this outcome is not matched to any answer to the experiment's

*A philosopher of science would rightly interject at this point that one cannot arrive at definitive proof of a hypothesis, so that really there are outcomes that disprove the hypothesis and outcomes that fail to do so.

question and therefore probably points to a defect in the theory underlying the experiment. Several major discoveries of mainstream science (e.g., Ernest Rutherford's discovery of the atomic nucleus) have been made in this manner. Perhaps because of the weakness of military theories, serendipitous discoveries often emerge from military experiments. Yet it would be a mistake to just go out and undertake some activity in the hope that a serendipitous discovery would occur.

"All's fair in love and war"[7]—What about in Experimentation?

Like the exercises they resemble, military experiments are governed by various rules, as if in a game. The participants, possessed of desirable traits such as competitiveness and desire to win, often bend or break these rules.

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.[8] 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, this notion is not among the many points of similarity between exercises and experiments. 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. In warfare, there is no need to keep the other side safe (in fact, quite the contrary), perhaps even a reduced need to avoid risk to one's own side, and no external constraints on the time or place of combat. In experiments, these considerations arise, and cannot be ignored or wished away.

Actual combat has even less in common with experiments than do exercises. Sometimes, during the process of defining the experiment's question and its answers, and how these will relate to

the experiment's event and its possible outcomes, those involved perceive that their experiment is really a test (e.g., of a piece of hardware, or a doctrine), and they will object that the proposed test is, one way or the other, unfair. Almost invariably, the desire to make the experiment a fair test becomes conflated with a desire to make the combat (in the experiment) a fair fight. The correct rejection of the latter usually comes at the expense of losing the former as well.

Three Idiosyncratic Nonservice Efforts

The three trials described in this section were undertaken by individuals, not nations or military services. These individuals were civilians, and while G. F. Gause clearly saw some relationship between his work and war, he was not in any sense trying to foresee the shape of the coming Second World War.

Yet these experiments make a good beginning for our survey. One reason is to provide instances of *wargaming*, an important aspect of simulation not addressed in the other sections. Another reason is simply to stretch the reader's imagination: the combat models these experiments incorporate are neither the computer models in widespread use today nor the exercise-like models used in the experiments with which the balance of the paper is concerned.

G. F. Gause's Bugs[9]

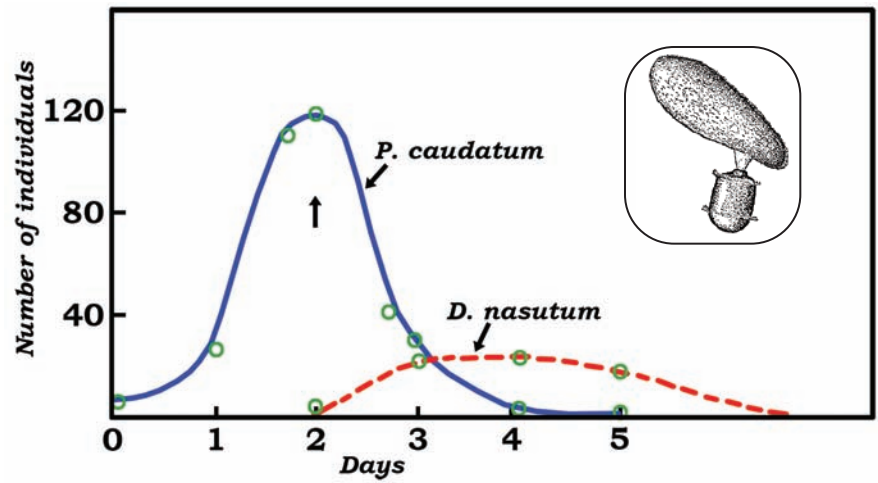
The naturalist Georgii Frantsevich Gause (not to be confused with the mathematician Carl Friedrich Gauss) experimented with colonies of beetles and bacteria that competed with one another either indirectly, by drawing upon the same resources, or directly—by predation.

One of his goals, notable to us for being cast in starkly military terms, was to "give a clear answer to Darwin's question: why has one species been victorious over another in the great battle of life?"

Another goal was much more definite: to see if the animals' populations followed the equations formulated earlier by Ronald Ross, Alfred J. Lotka, and Vito Volterra. Mathematically, these are quite similar to the equations independently formulated by Lanchester (and others) to describe force-on-force attrition warfare:[10]

$$R(t) = -b \times B(t)$$
$$B(t) = -r \times R(t)$$

where R(t) and B(t) are the sizes of the two forces Red and Blue,



In an illustration from Gause's book The Struggle for Existence*, the smaller voracious* Didinium nasutum*, which requires at least one* paramecium *every three hours, is depicted devouring a* Paramecium caudatum*. This graph depicts the basic interaction of the predator-prey relationship between the two. It demonstrates the precipitous decline of the prey (*paramecium*) immediately after the introduction of the predator (*didinium*). It also shows the decline in the predator population with that of the prey.*

and r and b are positive numbers indicative of the quality of the fighters on each side. The naturalists' version of the equations differs somewhat because it lacks the element of mutual predation; rather, although it may include one-way direct predation as above, it includes exponential growth by breeding, limited by the two populations *each* acting to limit their *joint* size, by competing over a finite food supply, which is itself a third population.

Given the presence of a physical refuge in which a population of the prey could survive while the predator starves outside, these equations predict oscillations in the populations' size: the predator grows numerous, kills nearly all the prey, and then starves almost to extinction, whereupon the remnant population of the prey (having survived in the refuge) multiplies almost unchecked and becomes numerous, once again providing plenty of food for the predator.

Gause performed the same experiment multiple times, starting

with identical populations of the predator and prey, and experienced some serendipity when it turned out that starving predators continue to reproduce at or near the usual rate, the starvation manifesting itself in a reduction of each individual's size, not in a reduction of the population count. Switching over to measuring population in terms of mass, not number of individuals, he observed the expected oscillations, but also noticed that eventually one species or the other would die off completely.

Hector Bywater's Wind-Up Ship Models

Published in 1925, Hector Bywater's *The Great Pacific War* foretells World War II in the Pacific with quite a degree of accuracy: the surprise attack by the Japanese, the invasions of Guam and the Philippines, and the eventual victory of the United States based on the island-hopping strategy. In the book, the Americans force the Japanese to a decisive Battle of Yap Island in a way that is completely parallel to the way the Japanese planned to force the American fleet to a decisive battle in the real world's later Battle of Midway.

Bywater, a civilian, was a journalist and commentator on naval affairs. His biographer, William Honan, makes the case that the resemblance between the war eventuated and the way Bywater presented it was neither coincidence nor prescience: it was that the participants read Bywater's book and liked the ideas in it better than their own.

Bywater, in turn, discovered his ideas by experimentation. He proceeded from tabletop wargaming (like that of Fletcher Pratt, described below) to experiments using self-propelled ship models on a pond. Most builders of ship models rejoice in the painstaking reproduction of minute details, but Bywater eschewed these, holding

> a multiplicity of 'gadgets' to be a nuisance in a working model, . . . they should therefore be reduced to the minimum consistent with the character of the ship.[11]

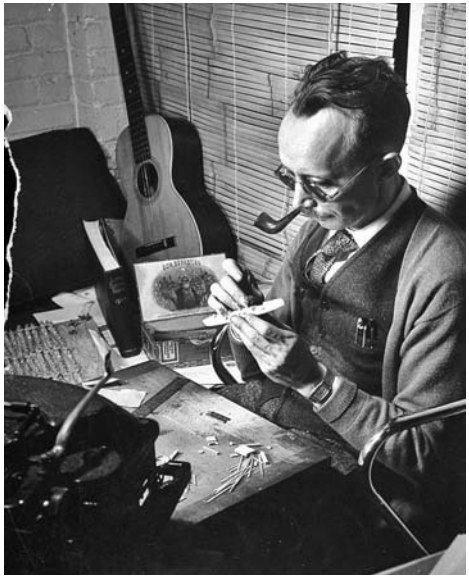
Radio-controlled ship models did not yet exist, so Bywater and his assistants had to *program* the maneuver and fire of the ships when they launched them from the pond's perimeter. Perhaps this limitation actually offered a realism beyond that of the tabletop games, in which players are notoriously able to put ships through intricate evolutions that would be beyond the command and control capabilities of real ships. The firing was, in any case, entirely cosmetic, leaving maneuver as the important variable.

The central problem of naval strategy at that time was, "How can the superior fleet bring the inferior fleet to battle?" As at Jutland, a force that thought it might lose could readily flee in most situations. Bywater, apparently through the maneuvering of the mechanical ship models on the pond, came to an answer: threaten to invade an island within the enemy's defensive perimeter—ideally, one that he cannot afford to lose. In his resulting book, *The Great Pacific War*, the Americans did this at Yap, in the Caroline Islands.

This answer ran counter to the planning wisdom of the time, both American and Japanese. Opposed landings were held to be difficult or impossible, and U.S. planning centered on the "through ticket" from Hawaii to the Philippines. Fifteen months after the appearance of Bywater's book, the United States' War Plan Orange changed from the "through ticket" to island hopping, via the Marshalls and Carolines. After the war began, the Japanese used the island-invasion ploy at Midway, getting the major battle they had sought, but losing it. As every teacher knows, the best way to detect copying is in the similarity of wrong answers, not correct ones: Admiral Isoroku Yamamoto, like Bywater's Americans, inexplicably positioned his carriers ahead of his battleships.

Thus far, this account has followed that of Bywater and his work given by Honan.[12] He, much less Bywater himself, gives no description of how the operation of the ship models on the pond produced the insights that it is said to have produced. One is left to think that the models and their collective motions somehow

fertilized Bywater's capacity to imagine naval actions, and thus led him to reach conclusions that could not have been reached unaided. This explanation is very possibly correct, but there is an intriguing alternative: Bywater and his helpers may have in fact *failed* to make the wind-up ships behave in the desired fashion. Indeed, it is difficult to imagine how they could have made mechanical ships stay in formation for long, or execute turns with any degree of simultaneity. Precisely because clockwork ships cannot be made to perform clockwork operations, they illustrate the point that a formation, and *a fortiori* a large operation, can stay in synchrony only by constant adaptive adjustment. A Japanese weakness in the ensuing war was to create complicated plans that looked good on paper, but were unduly difficult to execute—all the more so because of the Japanese weakness in communication technology. The Americans were better served by their simpler plans and better communications technology.



Hansel Mieth/Time Life Pictures/Getty Images

Fletcher Pratt works at his home on a model of a Chinese gunboat in preparation for the next round of his war games.

Fletcher Pratt's Naval War Game[13]

In the late 1930s, writer and polymath Fletcher Pratt perfected a naval wargame similar in spirit to the games conducted on the famous checkerboard deck of Sims Hall at the U.S. Naval War College.[14] Although he was professionally involved with naval matters as a journalist, and some of those who played his game

were officers (naval and otherwise), his game was in large part a social affair and the players were a co-ed group of writers, socialites, and artistes. If, as some have suggested,[15] there is value in having a diverse group of players, Pratt had it. Given the existence, and fame, of the War College game, the Pratt game is noteworthy because of the following reasons:

- Its modest level of effort. Pratt and his group of New York City intellectuals played it for fun, in a ballroom, with homemade ship models. The rulebook is slim and newcomers could learn to play in 15 minutes. The contrast with today's computer-based simulations is astounding.
- Its encapsulation of the value of surface combatants via a single simple formula taking into account tonnage, the thickness of armor, the size and number of guns, etc. This formula, and therefore the game on which it was based, represented a *theory* of naval combat.*

Pratt also conceived of an intriguing solution to a bothersome problem. Previous wargaming systems had used dice-rolling to determine which shots, if any, hit their targets. These games' dice-based determination of hits and misses had difficulty capturing many of the considerations that bear on whether or not naval gunfire hits its mark, including the ability of the gun director to correct on the basis of the splashes of shots that miss; the increased difficulty of hitting if the target is moving faster, or maneuvering; or if the shooting ship is doing so (and, conversely, the increased ease of hitting when shooting at a calmly moving target, or from a calmly moving ship); and the dependence of probability-of-hit on target aspect—bow-on and stern-on ships are easier to hit than broadside ships because the shooter's error

*Allen notes (on page 119) that one of Pratt's Army officer players was Trevor N. Dupuy, who later created the "Quantified Judgment Model" of land combat, which features a formula that attempts to summarize all manner of force characteristics as a single number.

is greater in range than in track. Rather than create a complicated mathematical model of this situation, Pratt simply obliged his players to lie down on the floor (i.e., deck) of the ballroom and direct their fire by aiming, as best they could, a pointer at the target and noting down their estimate of the range (in inches). An umpire would then use a tape measure to find the point of aim from this information, and hits were scored if the model target ship (or another ship!) was at the point of aim.

Pratt can be considered to have been experimenting when, in advance of the Battle of the River Plate, he set up a game pitting the German "pocket battleship" *Admiral Graf Spee* against several smaller ships. Military wisdom of the time—including that of Pratt—held that the larger ship should win, but the game predicted a victory for the multiple small ships. Such of course, was the outcome later, when the *Graf Spee* was defeated by the British heavy cruiser *Exeter* and the light cruisers *Ajax* and *Achilles*.

Observations on the Idiosyncratic Efforts

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."[16] The same could certainly be said with *combat* replacing *competition*. Gause recognized that the mathematical theories might contain inadequacies and in fact saw his experiments as exploring a split between observation and theory: "these processes are extremely complicated and [their trends] often do not harmonize with the predictions of the relatively simple mathematical theory."[17] In our terms, Gause's work was a *crucial experiment* because it was a test of the very theory on which it was based. To an extent, it was therefore also a crucial experiment regarding the related Lanchester equations. Gause's predation examples ended in some markedly different endstates, and he missed an opportunity to make an observation that would be extremely relevant to the Lanchester case as well: in a case with predation, the endgame is an unstable process with too few

creatures for the law of large numbers to apply. It would be interesting to see how much of the observed variation in outcomes (which species exterminates the other, and how many of the victors there are when this happens) is predicted by the equations alone.

Gause's work is also notable for being a test-tube version of today's computer-based experiments in which intelligent agents battle one another.[18]

Bywater's work is nearly at the opposite end of the spectrum from that of Gause: he had no mathematical formulas to guide him, or to be tested, and his work did not even produce any statistics regarding numbers of ships "surviving" or the like, because he had no means of imputing damage to the ships. One interpretation of his work (and another will be presented in the section entitled "Overall Observations," at the end of this book) is that it stands as an example of work that is *objective*, yet not at all *quantitative*, despite common usage that treats these terms as equivalent. Unfortunately, his work is also an example of work in which the conclusions are documented, but the method of reaching those conclusions is not. As such, it is a warning to those who would do likewise: we can't tell whether his work was, to use terms defined earlier on, an *experiment* or an *exploration*.

Pratt's work lies in between. The contradiction, by Pratt's own game, of Pratt's (and others') views on the outcome of the *Graf Spee* scenario, is especially notable for its disproof of such platitudes as "you can't get anything out of a model that you didn't put into it," "a model only embodies the prejudices of its designer," and so on. This disproof could arise because of Pratt's intellectual honesty, aided by the formal structure of his game. His belief in his facts regarding ships's speeds, guns, and armor, and in the way he had structured them into a game, suffice to make him believe in the output of the game, even though he would not have accepted it as a bare assertion.

The U.S. Prepares for World War II

This section looks at five sets of experiments undertaken in the United States with a view toward an eventual (or, in the later cases, imminent) Second World War. These cases are the U.S. Navy's Fleet Problems, at-sea exercises with a considerable experimental component; the famous joint ship bombing experiment of General William "Billy" Mitchell; the amphibious invasion experiments inspired by Major Earl H. "Pete" Ellis of the Marine Corps; the Army's Louisiana Maneuvers, which were field exercises with an experimental component; and the experiments with fighter direction conducted as an aspect of training at the Pacific Fleet Fighter Direction Officer School.

*The Fleet Problems, 1923–40**

Apart from their doubtlessly considerable training value, the U.S. Navy's Fleet Problems I–XXI are generally considered to have been pivotal in the development of U.S. carrier doctrine.[19] Regrettably, some are less well documented than others, and historians who have done archival research on the Fleet Problems have invariably expressed dismay at the large amount of information regarding them that seems to have been lost, or never written down in the first place. This, in itself, is a lesson for present and future military experimentation efforts.

Experiments I–IV, taking place in 1923–24, are therefore notable for their lack of actual aircraft carriers: USS *Langley* (CV 1), had been commissioned in 1922 (following conversion from its earlier incarnation as a collier, *Jupiter*), but was not available for use in the Fleet Problems. Instead, other types of ships were used as surrogates for carriers. In Fleet Problem I, for example, the attacking Black fleet had as its carriers the battleships *New York* (BB 34) and *Oklahoma* (BB 37), whose flight operations were represented by

*Since this section was originally written, an important new source on this topic has been published: Albert Nofi, *To Train the Fleet for War: The U.S. Navy's Fleet Problems, 1923–1940* (Newport, RI Naval War College, 2010).

J.M.F. Haase Collection/San Diego Aero Space Museum via Tailhook

The USS Langley *(CV 1) steams off the southern California coast with its aircraft during one of the Fleet Problems of the mid-1920s.*

their catapult-launched spotting planes. The scenario was a Black attack on the Blue-defended Panama Canal, and a single Black airplane, dropping miniature bombs, was ruled to have destroyed the Gatun spillway, which would lower the level of Lake Gatun, rendering the canal impassable to ships.

USS *Langley* at last participated in Fleet Problem V in 1925. In this scenario, an attack on the Hawaiian Islands, the Black aggressor was again given the carrier, as well as two seaplane tenders and aviation-capable battleships and cruisers. The Blue defenders had no carriers, and the airplanes on board their battleship, USS *Wyoming* (BB 32), could not be launched for want of a catapult.

The *Langley's* performance was sufficiently impressive as to re-

sult in a recommendation that the *Lexington* (CV 2) and *Saratoga* (CV 3) be completed quickly.

Fleet Problems VI (1926) and VII (1927) again addressed attacks on the Panama Canal. Before Fleet Problem VII, the fleet participated in what was really a joint problem because the U.S. Army was involved as well. In it, the *Langley* again appeared in the aggressor's fleet, and its aircraft were used to defend against attacks from land-based Army aircraft, the first instance of what would later be termed combat air patrol (CAP). Even with the participation of the *Langley*, the attacks on the Miraflores Locks (at the Pacific end of the Panama Canal) were executed by single airplanes, acting as surrogates for larger forces. A commentator noted,

> In later problems when carriers were available from which attacks in force could be launched and great reality could be introduced into the maneuvers, the vital necessity for air defense of the Canal was to become even more apparent.[20]

In Fleet Problem VII, the *Langley* joined the Blue side and was to protect a slow convoy crossing the Caribbean Sea. The carrier again provided CAP, but had already recovered its planes when, after the end of the exercise, Black land-based aircraft attacked the convoy. This Fleet Problem was taken to indicate the need of carriers for freedom of maneuver, and freedom of action in employing their aircraft.

Fleet Problem VIII (1928), in Hawaiian waters, suffered from bad weather.

Fleet Problem IX (1929) marked a major advance—the introduction of the large, fast carriers *Lexington* and *Saratoga*. The venue was again the Panama Canal, and the goal was to validate with actual carriers and full-size air complements the conclusions drawn earlier from the exercises conducted using surrogates.

The Blue defenders got the *Lexington* while the Black attackers had the *Saratoga* and USS *Aroostook* (CM 3), a converted

minelayer seaplane tender that was to represent the *Langley*. The *Lexington* ran into Black's battleship division, and would surely have been "sunk" by gunfire, but was ruled to be only "damaged" for the sake of preserving the remainder of the problem. This "damage" took the form of decreased speed.

Serendipity arose:

> The climax of this exercise came when *Saratoga* left the main force of battleships and, accompanied by one light cruiser, made a high-speed run from the west. . . . This strike, however, was not part of the original plan for the exercise and seems to have come about simply because the destroyer screen for *Saratoga's* battleship escort did not have the fuel to stay with it. Nevertheless, after [Fleet Problem IX], carriers "were accepted as fleet units."[21]

The Blue light cruiser USS *Detroit* (CL 8) had sighted the *Saratoga* and was able to track it on into the night, but the crippled *Lexington* lacked the speed to give chase until the next morning, when its "damage" was ruled to have been repaired. *Saratoga* launched 70 airplanes from a position some 145 miles away from the canal. In an incident parallel to that which had befallen the *Lexington*, *Saratoga* then encountered the Blue battleship division, as well as an enemy submarine. It was ruled to have been sunk by each, despite which it managed to launch another 13 airplanes.

The *Aroostook* launched its one plane, representing *Langley's* aircraft, on a one-way mission to the Atlantic end of the canal, to bomb the locks and spillway at Gatun, and the port of Coco Solo. The Blue defenders, however, were unaware that this seaplane represented the entire *Langley* complement, and took little notice of it until it alighted at the Atlantic end of the canal and the pilot "surrendered," explaining what he was supposed to be considered to have done.

Saratoga was recovering its airplanes (which had successfully attacked the Miraflores and Pedro Miguel Locks, as well as two

Army airfields, all at or near the Pacific end of the canal) when an airstrike from the *Lexington* arrived and "sank" *Saratoga* for the third time. A fourth "sinking" of the *Saratoga* was attempted by Blue land-based naval aircraft the following day, but the two sister ships were operating in proximity—so close, it was reported, as to have engaged each other with their 8-inch guns—and they attacked the *Lexington* instead.

Fleet Problems X and XI took place in 1930. In the former, Blue got the *Saratoga* and *Langley*, while the enemy coalition Black had the *Lexington*. The contest was nothing less than a struggle for control of the Caribbean Sea. Neither side had a good idea of where the other's ships were, and bad weather lessened the amount of air search that either could accomplish. After some days, aircraft from the *Lexington* spotted the *Saratoga*, and successive strikes put it, and then the *Langley*, and finally a number of Blue surface ships, out of action. These operations resembled, in detail as well as in the large, later operations in the war with Japan, and they showed the primacy of the offensive in carrier warfare. Fleet Problem XI similarly pointed to the importance of scouting, and the after-action report apparently offered a counsel of perfection:

> After the game, it was recommended that scouting squadrons be increased to 18 planes and that a more suitable scouting plane be developed. It was felt that better flotation was needed for amphibians and that a greatly increased range for carrier-based scouts, as well as the ability to take off from a short run, were necessary. Among desirable secondary characteristics were small size, folding wings, and high speed, even at the cost of ceiling and armament.[22]

The willingness to sacrifice armament must be considered in light of the observation, made after Fleet Problem X, deploring the scouts' inability to bomb carriers when they found them. The report on Fleet Problem XI also recommended the creation of bat-

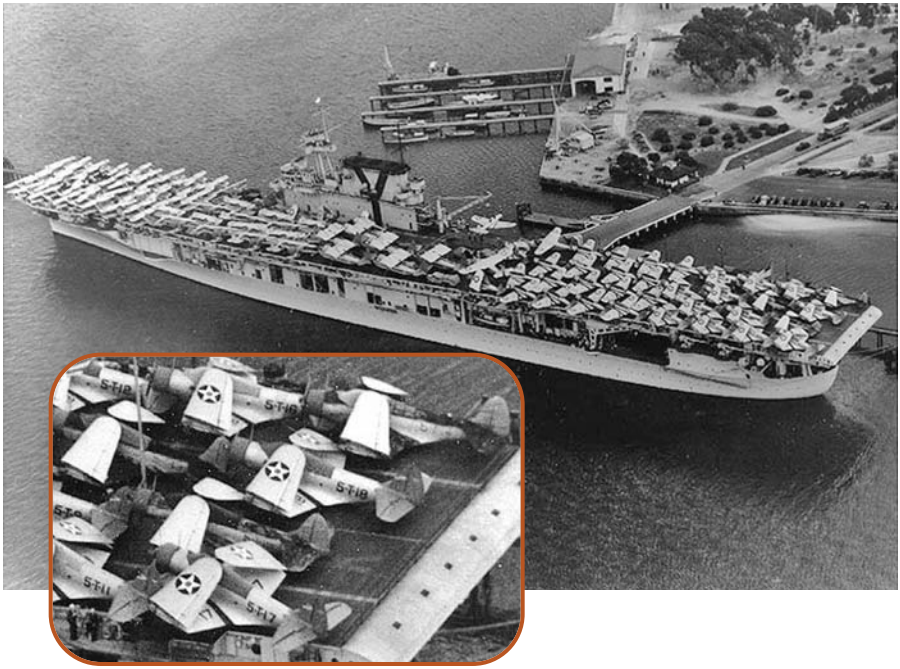
tlegroups as we know them today: cruisers and destroyers assigned to escort a carrier, all training together as a team.

Fleet Problem XII, conducted in 1931, included scouting and at-sea refueling, but the major theme was combat between a battleship-intensive fleet and a carrier-intensive fleet. It was found that the commander of the latter had best be located on the carrier itself, an arrangement that we take for granted today, but that was novel at the time.

Fleet Problem XIII, held in 1932, dealt with air search for submarines, and it found that the submarines were clearly vulnerable to airborne detection and attack. Despite a very different setup from that which had faced the commanders in Fleet Problem X, the commanders in Fleet Problem XIII made the same choice: a

USS Yorktown *(CV 5), at Naval Air Station, North Island, San Diego, California, in June 1940: Three Torpedo Squadron Five (VT-5) Douglas TBD-1 Devastators at the after end of the flight deck are painted in experimental camouflage schemes tested during Fleet Problem XXI.*

Official U.S. Navy photograph, U.S. National Archives



top-priority effort to find and defeat the enemy carrier. The commander of the Blue aircraft noted this and saw a corollary: the side with more carriers would have a great advantage.

In 1933's Fleet Problem XIV, the Blue force was to protect the West Coast against Black raids. Black's *Lexington* again stumbled into contact with Blue battleships, getting caught between two and ruled "sunk." *Saratoga* conducted successful attacks on Long Beach, Venice, and El Segundo, and then moved north to hit San Francisco. While launching its strike there, it was attacked by aircraft from the *Richmond* (CL 9) and *Langley*. Aircraft from the *Saratoga* and *Langley* proceeded to damage each other's ships, again underscoring the primacy of the offensive in carrier warfare, and hence the need for high-quality attack airplanes and weapons. This lesson was drawn again from the ensuing Fleet Problem XV in 1934.

In 1935, USS *Ranger* (CV 4) joined the series for Fleet Problem XVI, actually a disconnected set of joint problems conducted in cooperation with the Army and the Coast Guard. Fleet Problem XVII, in the following year, seems to have made only technical contributions, such as the introduction of the automatic pilot.

Fleet Problem XVIII (1937) returned to the business of refining carrier doctrine, this time addressing the question of whether carriers should operate with the main body of the fleet (defined, of course, by the presence of battleships) or separately. No strong conclusion seems to have been reached.

Fleet Problem XIX included a now-famous attack on Pearl Harbor, foreshadowing in 1938 the Japanese attack of 1941. The *Lexington* was eliminated early, by a long-range flight from San Diego, but aircraft from the *Saratoga* flew attacks quite similar in detail to those flown later by the Japanese. The defenders' difficulties were analyzed as stemming from the mobility of the fast carriers.

Fleet Problem XX (1939), in the Caribbean, featured the carriers *Yorktown* and *Enterprise* (CV 5 and CV 6, respectively), but not the *Langley* or *Saratoga*.

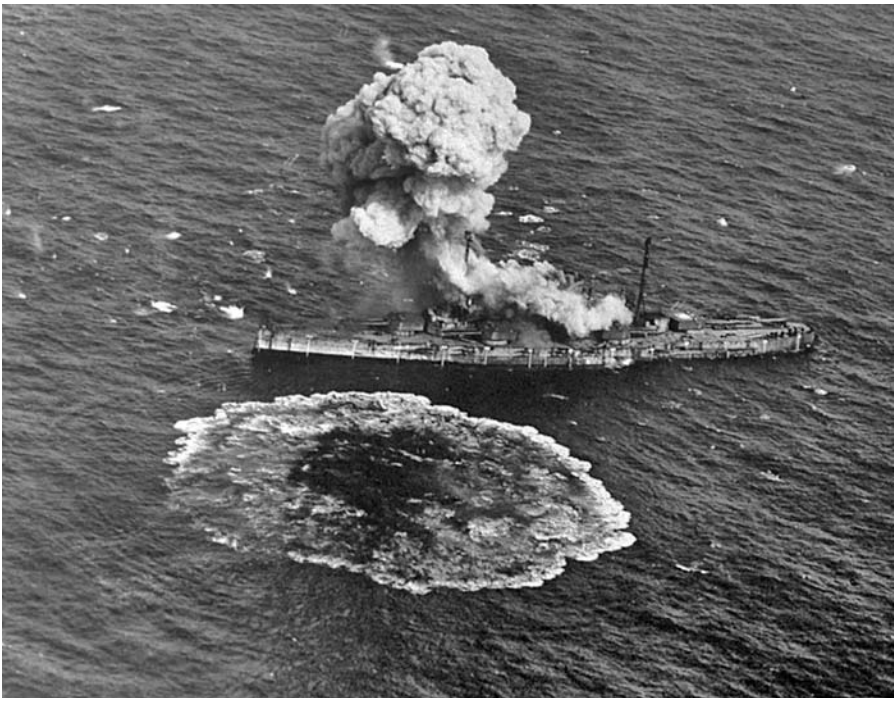
The last Fleet Problem, XXI, took place in 1940, in the Hawaii area. It consisted of two exercises devoted to refining such points of carrier warfare as coordination and planning of scouting and screening.

The Experiments of Billy Mitchell

In 1921, the decommissioned ex-German battleship *Ostfriesland* was bombed by U.S. Navy and Army airplanes in an experiment made famous by its connection to then–Brigadier General William "Billy" Mitchell. The ship sank, Mitchell declared battle-

Over 13–21 July 1921, the 49th and 96th Bombardment Squadrons assigned to Brigadier General William "Billy" Mitchell's 1st Provisional Air Brigade at Langley Field, Virginia, conducted tests to determine the efficacy of aircraft against warships. The group's Martin MB-2 and Handley-Page HP 0/400 aircraft successfully bombed and sank three ex-German warships, including the formidable 22,437-ton battleship Ostfriesland*, shown here, off the coast of Virginia.*

U.S. Naval Institute Photo Archive



ships to have been made obsolete by airpower, and the impression stuck.[23]

The experiment actually began with a search by the bombers for the decommissioned battleship ex-*Iowa* (BB 4, not the later BB 61), which was operating under radio control. It was moving at six knots and the searchers knew only that it was offshore somewhere between the mouth of the Delaware River and that of Chesapeake Bay, a distance of about 100 miles. To conserve airplanes, the search was conducted with dirigibles.[24] These having found it, the ship was subjected to bombardment with sand-filled dummy bombs.[25] "Seacraft are not only very easy to find, but their type and character are also as easy to determine from the air," wrote Mitchell.[26]

In the first round of heavy bombing against the *Ostfriesland*, Navy and Marine aircraft, making multiple passes, dropped 33 medium-sized (230 lb.) bombs (out of 36 carried), hitting with nine—but only two exploded, and with low-order explosions at that. The next wave of Army and Navy aircraft dropped 600-lb. and 550-lb. bombs, respectively. Nineteen were dropped (out of 24 carried), resulting in five hits and a near miss. Of these, only one hit and the near miss—both Army-dropped—detonated. Inspection showed that the duds, while doing some damage through sheer kinetic energy, did not compromise the watertightness of the hull, but either the exploding hit or the water hammer effect of the near miss had started several major leaks.[27]

The next day, the Army dropped five 1,000-lb. bombs (out of 12 carried by six airplanes), scoring three hits and two near misses. In a pause, Navy inspectors boarded the target ship and found that the hits had done great damage, but not to the watertightness of the ship. The Army then loaded 2,000-lb. bombs onto its aircraft, dropping six and hitting (or just barely missing, and applying the water hammer effect) with three. The ship sank stern-first relatively promptly, with a final 2,000 pounder administered to the bow as it went under.[28]

Major Pete Ellis and USMC Interwar Experimentation[29]

U.S. Marine Corps History Division

Marine Major Earl H. "Pete" Ellis's vision of Marine Corps amphibious operations in the Pacific islands precipitated a series of Navy and Marine Corps landing exercises in the 1920s. These revealed deficiencies in almost every aspect of the operation, which prepared the services for their actions two decades later.

In the 1920s, after the disastrous experience of the Gallipoli landing in World War I, conventional wisdom held that opposed amphibious landings—especially in daylight—were impossible operations. Yet the United States' important Pacific possessions, Guam and the Philippines, lay on the far side of the vast Japanese Mandate—the former German islands entrusted to Japan in its capacity as one of the victorious powers of the First World War.

Major Earl H. "Pete" Ellis, USMC, who took a great interest in these islands of the Pacific (and was to die among them), saw the need to prepare for their conquest. He envisioned Marine Corps amphibious operations as the means by which the United States would retake the islands, and higher-ups in the Corps (perhaps having interservice politics in view) agreed. Ellis met his mysterious end in 1923, but the Navy and Marine Corps conducted a series of landing exercises in the 1920s. Although these—like some coeval USMC (and Army) refights of Civil War battles using 20th-century equipment—were actually exercises rather than experiments, they revealed deficiencies in almost every aspect of the operation, including fires, logistics, and the vessels and vehicles needed to accomplish the actual transition from sea to land.

The Corps was slow to react, but when it finally did—in the

1930s—it did so decisively, establishing the Fleet Marine Force as an entity with a charter to engage in expeditionary warfare, and setting about the creation of a usable manual of landing operations.

Absent any authorities on the topic, the latter effort was accomplished by dint of what would, today, be termed *brainstorming*. Quantico's schools for officers were devoted entirely to the project, in which each Marine prepared a chronological account of a landing operation. These lists were then subjected to a multistage winnowing process at the hands of more senior officers. This process is remarkable for its bootstrap nature. One of the Marines wrote that the group

> approached its subject . . . about the same as every other committee, with a lantern in one hand and a candle in the other—but neither of these seemed to throw much light on the subject, so we wound up by hiding our lights under a bushel and using the imagination that God gave us to use for this particular purpose.[30]

The result was the famous *Tentative Manual for Landing Operations*, published in 1934. As its title showed, it was the first word on amphibious landings, not the last. Therefore, it needed to be tested. Accordingly, the Fleet Marine Force participated in seven annual fleet landing exercises, conducted in concert with the Navy at the islands of Culebra, in the Caribbean, and San Clemente, off San Diego. These exercises, despite numerous artificialities, helped to test the *Tentative Manual*—particularly with regard to fires, close air support, and logistics—and also gave some experience with new pieces of equipment. A final exercise occurred in June 1941, at New River, North Carolina. Simultaneously, work of an academic nature, based upon history and attempting to advance theory, continued, as did limited technical experiments with particular devices. The landing craft as we know it did not emerge as a solution until the very end of this

process, with the first test of the Higgins boat—that eventually became the landing craft, vehicle, personnel (LCVP)—occurring in Fleet Landing Exercise Number 6, in 1940.

The Army's Louisiana Maneuvers[31]

The U.S. Army's 1941 experiments, usually known as the Louisiana Maneuvers, were quite large in scope. They actually took place in two separate venues, one on the Louisiana-Texas border and one on the border separating the Carolinas, each encompassing thousands of square miles. Hundreds of thousands of soldiers took part in force-on-force battles adjudicated by thousands of umpires on the basis of a rulebook.

During the Carolina phase of the maneuvers near Jefferson, South Carolina, on 26 November 1941, members of the 101st Coast Artillery attached to the IV Corps "Reds" used a wooden beam to simulate a 37mm antitank gun. The gunners called it a "wood-o-meter" and claimed it could fire 400 splinters and 15 drops of turpentine per minute.

Charles Gorry/Associated Press

Surrogates represented unavailable equipment, including especially the antitank guns whose abundant deployment was one of the topics of experimentation.

Although the maneuvers were seen primarily as a tool for training (at all levels), the Louisiana portion contained elements of serendipity and outright experimentation. The former was exemplified by the "Battle of Shreveport," which showed "the decisive influence of destroyed bridges," a lesson that would of course be confirmed in later Allied operations in Europe during the Second World War, and by various discoveries pertaining to the importance of close air support on the one hand, and to its difficulty on the other.

The explicit experimentation took the form of hypothesis-testing:

- The first phase of the Louisiana portion pitted a large, traditional force against a smaller, but more mobile one, to test what we would now call the maneuver warfare hypothesis that a small, agile force can prevail against a larger, but clumsier, one. The action was a meeting engagement, in which each side had orders to attack.
- The second phase of the Louisiana portion was even less symmetric, testing the hypothesis that the smaller and more mobile force could successfully *defend* against the larger force. It was in this phase that the destroyed bridges figured so prominently.

In the first phase, the antitank weapons of the traditional force blunted the armored attacks of the mobile force, and ultimately the traditional force prevailed. Arguably, the adjudication procedures of the experiment—the rules of the game—portrayed antitank weapons as unrealistically effective, and unrealistically invulnerable, and thus skewed the outcome.* Time ran out before the second phase reached a conclusion.

*More fundamentally, some weapons—such as the .50-caliber machine gun—were portrayed as being antitank weapons, despite the fact that tank armor had progressed to the point of invulnerability against these weapons.

The Carolina phases were planned after the Louisiana phases had ended. The maneuvers' organizer, Lieutenant General Lesley J. McNair, was a strong believer in antitank weapons, and no steps were taken to correct the rules' possible bias against tanks. On the contrary, hand grenades, represented by small bags of flour, were ruled capable of destroying tanks, notwithstanding an existing rule that tanks' very presence would neutralize any infantry within 100 yards.

Then-Colonel Dwight D. Eisenhower related an amusing incident regarding surrogates—and cheating.

> An umpire decided that a bridge had been destroyed by an enemy attack and flagged it accordingly. From then on, it was not to be used by men or vehicles. Shortly, a Corporal brought his squad up to the bridge, looked at the flag, and hesitated for a moment; then resolutely marched his men across it. The umpire yelled at him:
>
> "Hey, don't you see that that bridge is destroyed?"
>
> The Corporal answered, "Of course I see that it's destroyed. Can't you see we're swimming?"[32]

It is easy to see how this cavalier approach to the rules, if widespread, could have undercut what turned out to be a key finding—mentioned above—of the maneuvers, the importance of bridges.

The first Carolina event was a more extreme version of the first Louisiana event, with even greater disparities in size and mobility between the large traditional force on the one hand and the small armored force on the other. The traditional force prevailed, despite some spectacular local successes on the part of the mobile force, in part because of some cheating such as starting early, deploying out of area, and using the local commercial telephone system, in part because of the bias in the rules regarding antitank weapons, but also in part because of some extremely aggressive and creative play on the part of the traditional force, and a tendency on the part of the armored force to dissipate itself in piecemeal attacks.

The second Carolina event resembled the second Louisiana event, in which a small armored force attempted a mobile defense against a larger traditional force. Like the second Louisiana event, the second Carolina event was ended by the calendar, before a conclusion had been reached. Considerable serendipitous learning had taken place on the armor-heavy side, including the lesson that tanks cannot function without infantry support.

By comparison to other experiments with land forces, the Army's maneuvers in 1941 are remarkable on several counts, notably:

- There was no bias in favor of innovation; on the contrary, the deck was very arguably stacked in favor of the traditional force and against the armored force (e.g., by the toleration of considerable cheating on the part of the former, and by the skewedness, noted previously, of rules involving antitank weapons).
- Although the traditional force was dubbed "Blue" and the untraditional armored force "Red," there was not a clear identification of one side or the other as "The Americans."
- Contrary to the visions of British and European tank theorists, the experience in the Spanish Civil War, and the experience in the Third Reich's blitzkrieg conquests (all of which had taken place by the end of the Army's experiments), the 1941 maneuvers highlighted the mobile defense as a potential mission for the armored forces. Looked at differently, they highlighted attacks against armor by traditional forces.
- From the maneuvers, the Army very clearly drew the lesson that tanks need to be mixed with infantry and used in a combined-arms manner.[33] All participants in the Second World War, with the possible exception of the Russians, eventually reached

this conclusion. The discovery of a successful combined-arms doctrine is rendered all the more difficult by its similarity to a doctrine emphasizing the use of tanks in support of infantry. The latter, taken to its conclusion, leads to the development of "infantry tanks," designed to accompany infantry and accordingly capable of moving no faster than the infantry can walk. Experience in France and North Africa was to reveal great weaknesses in this form of tank and tank warfare.

- The organizers did not flinch from the idea that commanders might err during the maneuvers. "My God, Senator, that's the reason I do it," responded General George C. Marshall in response to a senator's question on this point. "I want the mistake [made] down in Louisiana, not over in Europe, and the only way to do this thing is to try it out, and if it doesn't work, find out what we need to make it work."[34]

To the degree that these salient features form any kind of coherent picture—and it might be a mistake to read too much into them—it is a picture of an experiment whose not-so-hidden agenda is to debunk the claims of the innovators and show the continuing supremacy of a force organized along traditional infantry lines, immunized against tanks by some extra antitank weapons and an aggressive approach to their use.

Pacific Fleet Fighter Direction Officer School[35]

British experience gained in the Battle of Britain showed that in an air battle, the defending aircraft needed the central control of fighter directors if they were to defend against bombers. This experience reached the U.S. Navy, via the Royal Canadian Air Force Radio School, prior to Pearl Harbor, and was immediately deemed applicable to the defense of carrier task forces against the massed attacks of Japanese dive- and torpedo-bombers. Radio

Navigating and Direction Officers' Association

Similar to the Pacific Fleet Fighter Direction Officer School, the British Fighter Direction Training Centre also used surrogates in their training. The well-known (in Great Britain) Wall's ice cream tricycles were used as both enemy aircraft and friendly fighters at Speckington Hall, less than two miles from the Yeovilton Naval Air Station.

would allow the fighter directors to communicate with their fighters, and radar would allow them to perceive the enemy attackers, but to succeed, they would also need tactical proficiency. The Pacific Fleet Fighter Direction Officer School was created to teach them how to fight the battle.

The commander of the school, Lieutenant Commander John "Jack" Hook Griffin, had been assigned to Royal Air Force and Royal Navy fighter squadrons, for whom the war had already started. To teach his freshly-minted ensigns how to be fighter director officers (FDOs), Griffin amassed a collection of Navy publications and manuals, but he also undertook to create a laboratory for the simulation of fighter battles. Having been given an aircraft hangar in which to conduct his school, he set about recreating the air battle on the deck of the hangar.

The result must have been quite a sight. One foot on the deck

represented a nautical mile on the sea surface. Tricycles, modified to accommodate a grown-up, geared down to move more slowly, and provided with hoods so that the "pilots" could only see for a few feet in any direction, were the fighters. Telephones, with long wires to each tricycle, represented the radios—later, these telephones were replaced with actual radios so as to eliminate the trailing wires. The tricycles also had compasses and clocks, and were equipped with a table informing the driver of how fast to pedal to simulate various airspeeds. Trainees on a catwalk thus had a view of the deck that simulated a view of a radar screen: viewing it from high above, the trainees saw the tricycles as blips, with a painted circle at the center of the hangar representing the task force itself and concentric circles surrounding it, as on a plan-position indicator radar screen. These trainees then, as would real-life radar operators, communicated their observations by telephone to a separate plotting room.

This set-up allowed the trainees to plan and test defensive tactics, and to learn, apply, and validate points of British fighter doctrine, such as a teaching that every group of incoming bombers should meet at least token opposition. However, the trainees also found some useful contraventions of British doctrine regarding, for example, how far out a raid ought to be met, and the utility of holding fighters in reserve, to contend with enemy breakthroughs. These discoveries reflect the use of the FDO school set-up for experimentation, as distinct from training.

The school also acted as the first step in what became a larger Navy experiment, the attempt to determine what kind of man would make a good FDO. This experiment came to include the assessment of performance in actual fleet duty, not just in the hangar at the school, and the result was surprising. High potential as an FDO did not correlate with flying experience, or a background in science or engineering; it correlated with prewar income level. Former teachers were the only outlier—they tended to do well as FDOs, despite their lack of income.

Observations on the U.S. Experiments

The early Fleet Problems are notable for their intensive use of surrogates—battleships standing in for carriers, *Langley* standing in for larger carriers, *Aroostook* standing in for *Langley*, and single planes standing in for formations of a squadron or more—to conduct experiments in carrier warfare while using few, if any, actual carriers. These formulations show great imagination and clear-sightedness and should be noted by experimenters of today, despite occasional mix-ups such as the failure to inform the defenders at Gatun that a seaplane from *Aroostook* would replace *Langley*'s aircraft, and lapses such as the apparent nonrecognition that a one-plane surrogate attack favors the attacker, because of the reduced probability that the attack will be detected.

In the Fleet Problems, the "Blue" side was clearly identifiable as the U.S. side, if only because of its missions. In the Louisiana Maneuvers, as noted above, the identification was less clear, but surely a considerable hint of Americanness inhered to the Blue side. In neither case, however, was the Blue side endowed with the modern (or even futuristic, especially in the early Fleet Problems) equipment—tanks or carriers. This set-up, quite different from today's, may have had the advantage that the controllers' natural bias in favor of the American side and any bias they may have harbored in favor of new technology (including any tendency to overestimate the effectiveness of coming "wonder weapons") would work in opposite directions.

Mitchell's results, of course, stand at odds with the later experience of World War II, in which aircraft certainly did attack ships successfully, but not by doing what his aircraft did to the *Ostfriesland*—pass over in level flight and drop heavy bombs. Closer consideration of the German battleship experiment suggests various reasons that level bombing was so successful in the 1921 trial.[36] For example, the ship was dead in the water and thus presumptively easier to hit than a moving target would be, and since it was unmanned, there was no damage control. The latter point

was probably quite important inasmuch as the bombing took place over two days, and leaks started on the first day admitted water unchecked all through the night, leaving the target quite low in the water as of the beginning of the second day.

These reasons are good examples of *artificialities*. It would have been difficult to make the ship a moving target (though not impossible; *Iowa* had been operated under radio control earlier in the trials), and, of course, safety considerations would preclude the presence of a damage control party during the bombing. The amplification of the first day's hits by the lack of damage control can be seen as pointing, unnoticed, to a lesson learned only after war broke out—the great importance of active damage control once a hit has been taken.

The search for the *Iowa* also illustrates the difference between a fair test and a fair fight. For a fair fight (i.e., a contest that could equally well be won by either side), the region in which *Iowa* might be found would have to be (as it seemingly was) rather limited. For a fair test, on the other hand, the region in which *Iowa* might be found would have to be as large as might be the case if *Iowa* were a hostile warship in a real-world operation. The two are entirely different.

More puzzling is the disjunction between the great difficulty encountered by scout aircraft in the Fleet Problems, and the ease with which Mitchell's dirigibles found the *Iowa* in the search phase of his experiment. Later wartime experience confirms the exercise result. Very possibly, the area in which the *Iowa* was known to steam was not so great as Mitchell makes it sound—100 miles may seem sizeable when considered as the length of the Maryland coastline, but it is not a great distance by maritime standards and, moreover, Mitchell does not say how far out to sea the search region extended.

Partial or total damage to ships figured strongly in several of the Fleet Problems (e.g., in the crippling of *Lexington*, or the various "sinkings" of *Saratoga)*. Yet—especially considering that

Mitchell's experiment may have been "rigged" and that the reaction of the Navy witnesses to *Ostfriesland*'s sinking may have been exaggerated or misinterpreted in some accounts[37]—nobody at the time really knew very much about how bombs damage ships. In this light, the relative success of the Fleet Problems seems paradoxical: absent an understanding of the effects of bombs on ships, how is it that the problems did such a good job of highlighting the major issues of carrier warfare? A similar question could be asked regarding the adjudication of combats in the Louisiana Maneuvers—these were handled by on-scene umpires using rules of thumb embodied in a slim manual, and the results can be considered to be an approximation, at best.

The story of interwar Marine Corps experimentation with amphibious operations, summarized above, is often told. In its command briefing and otherwise, the present-day Marine Corps Warfighting Laboratory, for example, makes frequent reference to it. These references typically point to the success of the effort, to the fact that much of the learning came from aspects of the exercises that the participants would have considered to be instances of failure, and in some cases to the fact that the principal product of the experimentation was knowledge, not the Higgins Boat (which was produced elsewhere, and brought in), or any other piece of gear. These are all important points. But the following seldom-noted points are important, too:

- The incorporation of academic effort, at the Quantico schools;
- The "bootstrap" nature of the knowledge-gaining process; and
- The lengthy timescale on which the events took place.

The process by which the Marine Corps students at Quantico developed the *Tentative Manual for Landing Operations* is a more extreme version of the same phenomenon. A group of people, seemingly drawing only on common sense and a knowledge of

what had *not* worked at Gallipoli, were able to create a successful set of instructions for doing something that nobody had previously known how to do. However, it is important to notice the next step—experimentation to test the contents of the manual.

That final point deserves some emphasis. Major Ellis directed the Corps' attention to the Pacific Islands before the First World War was even over; the Commandant approved his plan in 1921; a landing exercise was held at Culebra in 1922, and again in Panama in 1924; the Fleet Marine Force was formed in 1933; the *Tentative Manual* came out in 1934; experimentation began in 1935 and continued until just after America's entry into the Second World War. From the standpoint of today, the story moves right along and these events seem to occur in rapid succession. But they span more than 20 years. Today's efforts at experimentation come under heavy bureaucratic fire if they do not produce something in two years.

Commander Griffin's FDO school work, although experimental only in part, deserves attention because of his careful creation of a qualitatively and quantitatively well-designed surrogate that had no physical elements in common with the real thing. Doubtless many found it risible on their initial exposure, if not on a continuing basis. But its treatment of speeds and distances had been thought through logically and resulted in realistic engagements and a realistic flow of events over time, because of the correctness of the relationship between the distances on the hangar's deck and the speeds at which the tricycles were moving. In this way, the FDO school could in some respects represent an airstrike more accurately than had the Fleet Problems, even though the former used a tricycle and the latter used *Aroostook's* seaplane.

Germany Prepares for World War II

This section examines two prewar German efforts in military experimentation—the one that produced the tanks-and-planes blitzkrieg, and the one that yielded the U-boats' wolf packs.

The attempt to draw lessons from the German experience is sometimes critiqued on the grounds that the Germans lost. Yet while the United States eventually realized that it might be drawn into a second world war, Germany was planning on starting one, and accordingly set about preparations carefully. Nor can the blitzkrieg or the wolf packs be considered ineffective in themselves. Therefore, these preparations, if not others, merit some study.

The German Army's Experiments with Blitzkrieg

Tanks were introduced by Allied armies in the First World War

In its early between-the-wars experiments with combined infantry and armor operations, the German Army used simulated tanks—armor shapes over conventional road cars.

Associated Press

in an attempt to break the stalemated trench warfare. The German side captured some tanks and later made a few of their own. The war ended before large-scale armored warfare could occur, but not before it could be envisioned by officers on both sides.

During the period between the world wars, the German Army reinvented itself along the lines envisioned by General Hans von Seeckt, who believed in the creation of a small, elite army operating on precepts that became known as blitzkrieg, and are known in the present day as maneuver warfare—quite a change from the trench warfare of the First World War. These ideas, formulated by von Seeckt even before World War I had ended, dovetailed with circumstance when the Treaty of Versailles limited the size of the German Army to 100,000 men and mandated a long period of service (to prevent the training of a large population by cycling them through). Although von Seeckt was not among interwar Germany's many proponents of armored warfare, he agreed with them that the tank would have a part in the mobile force he envisioned, and he had tank surrogates produced for participation in maneuvers. These surrogates, made of wood and canvas and mounted on cars or even on bicycles, were used in force-on-force maneuvers that tested the new concepts of mobile warfare.[38] When the surrogate tanks' movement bogged down because of the poor crosscountry mobility of the automobiles and bicycles on which they were based, the reaction was not to conclude that tanks wouldn't work in mobile warfare, but instead that the surrogates needed to be made more realistic.[39]

Communications with the tanks proved to be especially important, both because of the fluid and fast-moving nature of the tank battle, and because one of their intended roles was for reconnaissance. In 1928, a communications exercise was held, with all Reichswehr divisions and groups participating.[40]

In 1931, the German Army conducted six exercises combining infantry and dummy tanks, notably at the Grafenwöhr training area, which is still used for that purpose today. Although their

primary purpose was training, they had an experimental aspect as well, because they were used to help create requirements for the real tanks, then being designed.[41]

> By 1 July 1934 the experimental work had attained such dimensions that it became necessary to set up a special Command of the Tank Forces. . . . The task of the new command was to continue the experiments with the mechanized forces and explore and test the tactical structures that might put these formations to the most effective use. In the autumn of 1935 the various cogitations and practical exercises culminated in large experimental maneuvers at Münsterlager, of which the most important result was the decision to establish three panzer (armored) divisions.[42]

This passage is especially notable for its explicit reference to "cogitations."

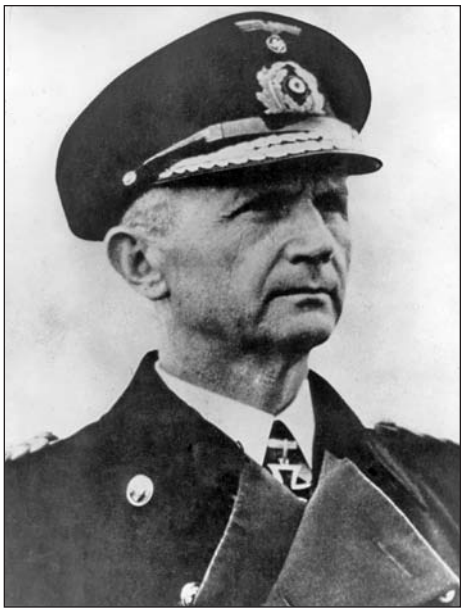
Blitzkrieg being a combined ground-air doctrine, the air arm needed to be developed along with the tank arm. In the 1920s, "simulated air attacks and aerial observation were developed by the military air staff as a normal part of command exercises and divisional maneuvers,"[43] sometimes using small balloons as surrogates for airplanes.[44] A more detailed surrogate was also sometimes used, as the American military attaché related after observing maneuvers in 1924:

> An officer, specially marked and often an ex-aviator, was permitted to ride on a motorcycle unmolested through and around the opponent's line. Returning, he reported in writing to the umpire designated, the result of his supposed aerial flight—the umpire permitting so much or all of the report as would be in keeping with an actual aerial flight to be transmitted to the commander sending out the aviator.[45]

In teaching air-to-air combat, the Germans used gun cameras

to record the aim of the "shot" for evaluation.[46]

Interestingly, Heinz Guderian's invention of the Panzer division, Germany's embodiment of the winning combined-arms doctrine of employing tanks, came after the era of experimentation, during which attention focused on avoiding the doomed "infantry tank" idea, and the related spreading of the tank force too thin.[47]



U.S. Naval Institute Photo Archive

In the 1930s, Germany's Admiral Karl Dönitz used readily available S-boats to conduct at-sea experiments to validate his concept of wolf pack tactics for U-boats.

For better or worse, the German Army's experiments, more than any others under study here, resembled some efforts of today in that they came closest to integrating tactical experimentation with ongoing hardware development, although even in this case, the two efforts belonged to separate organizations within the army.

The German Navy's Experiments with Wolf Packs

In the 1930s, Germany's Admiral Karl Dönitz conducted some at-sea experimentation to validate his concept of "wolf pack" tactics for U-boats, that is, German submarines.

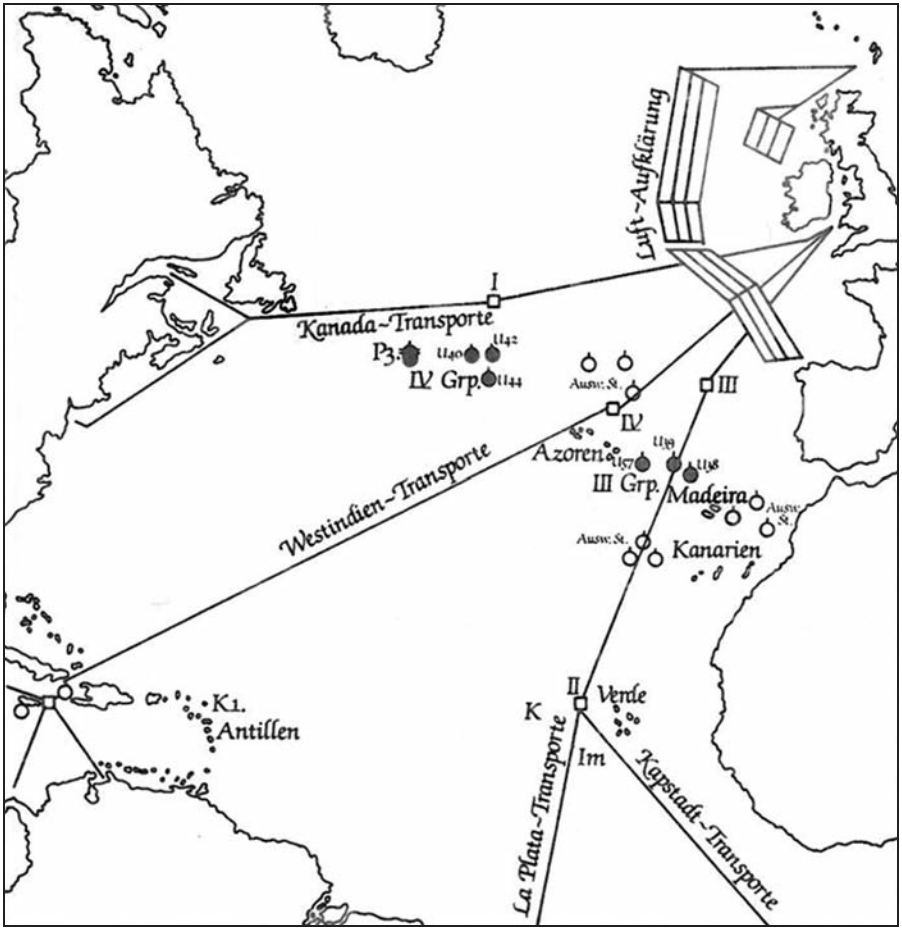
His idea was—and had been for some time[48]—that the submarines of the day were not so much undersea ships as submersible ships, and that earlier doctrine's view of them as individual raiders had failed to take into account not only their near-total lack of mobility and vision when submerged, but the possibility that the other side would form its ships into convoys.

Dönitz seems to have appreciated that even without escorts,

the convoy is a countermeasure to the individual submarine because the close grouping of the merchant vessels reduces the size of the region from which any are spotted. Because the submarine is limited in its ability to attack, the convoying side is more than willing to trade a large chance that one vessel will be sighted for a slim chance that many vessels will be sighted all at once. Escorts, if present, provide the added advantage of repulsing the attack of a single submarine, or at the very least preventing a

In the winter of 1938–39, Dönitz undertook a chart-based war game—this is an example—to explore actions on a scale too large to portray in his at-sea experiments.

Padfield. *Dönitz: The Last Führer*, 172



reattack. Dönitz also noted an underutilized asset of submarines, their respectable surface speed, and that surface operation actually represented a countermeasure to the antisubmarine sonars of the day.

Based on this inventory of submarines' strengths and weaknesses, and anticipation of the adoption of convoy as a countermeasure, he conceived of wolf pack tactics: a dozen or more submarines would form a long line at right angles to the expected track of the convoy, spread out as far as possible without creating a gap through which the convoy might pass. When a submarine saw the convoy, it would send a signal to higher headquarters, which would mastermind the convergence of the submarines, at a point farther along on the convoy's route, where they would submerge and lie in wait, and attack the large number of ships with a large number of submarines. Any escorts would be overwhelmed by the submarines' tactic of attacking all at once—on the surface, and at night, if possible.[49]

Dönitz had several questions about his idea. He later elaborated on them in the following way:

> a) *The exercise of control.* How far is it possible to exercise command over a number of U-boats? Is it possible during the actual attack, or only as far as to ensure coordinated action before the attack? What is the ideal balance between the exercise of overall command and giving the U-boat its independence of action? Must command be exercised by a person actually at sea? In a U-boat? Or in a surface vessel? Is it, anyway, possible to exercise command from a U-boat? Can command be exercised wholly or partially from land? . . .
>
> b) *Communications.* How can a U-boat be contacted when it is surfaced, when it is at periscope depth, when it is completely submerged, from another U-boat, from a surface ship, and from a land

> station? . . . The whole question of transmitting, receiving, and reporting beacon signals. . . .
>
> c) *Tactical.* How should the U-boats, operating together, act? . . .[50]

To answer these and other questions, he resorted to experimentation.

Sources disagree as to when this experimentation began. Some date it as early as the first part of the 1920s, when Germany had not yet violated the Versailles Treaty ban on submarines, arguing that torpedo boat exercises in tactical development, undertaken in that period, were in fact exercises in submarine tactical development, with the torpedo boats being used as surrogates.[51] This is certainly possible. The above-cited questions all refer to the part of the plan during which the submarines would be on the surface. Dönitz was in a torpedo boat flotilla at the time.

In 1935, Dönitz was given command of the Third Reich's first U-boat flotilla, and immediately began work on wolf-pack tactics. Whether or not the torpedo boat evolutions had been intended as submarine experiments, they were used as a source of insight into future U-boat operations. One of Dönitz's subordinates wrote:

> The end of 1935, then, saw the birth of those wolf-pack tactics which were later to be perfected in so masterly a manner. But between anticipation and perfection there were many stages. For reconnaissance and screening duties we adopted the old torpedo-boat tactics as our god-parent.[52]

Later, in 1937, Dönitz began to experiment with actual submarines: a wolf pack of some 20 submarines located and successfully "attacked" a convoy of armed transports sailing from East Prussia to Swinemunde, in the Baltic Sea, with Dönitz exercising command by radio from a surface ship at Kiel. Subsequent experiments in the Baltic and elsewhere were supplemented by real-world experience in the Spanish Civil War.[53]

Dönitz then shifted his attention to the high seas, undertaking a chart-based war game to explore actions on a scale too large to portray in his at-sea experiments. Such games tend to be cumbersome, and this one seems to have taken quite a while.

> In the winter of 1938–39 I held a war game to examine, with special reference to operations in the open Atlantic, the whole question of group tactics—command and organization, location of enemy convoys, and the massing of further U-boats for the final attack. No restrictions were placed on either side and the officer in charge of convoys had the whole Atlantic at his disposal and was at liberty to select the courses followed by his various convoys.

The points that emerged from this war game can be summarized as follows:

> 1. If, as I presumed, the enemy organized his merchantmen in escorted convoys, we should require at least 300 operational U-boats in order to successfully wage war against his shipping. . . .
>
> 2. Complete control of the U-boats in the theatre of operations and the conduct of their joint operations by the Officer Commanding U-boats from his command post ashore did not seem feasible. Furthermore, I felt that his "on-the-spot" knowledge particularly as regards the degree of enemy resistance and the wind and weather conditions prevailing would be altogether too meagre. I accordingly came to the conclusion that the broad operational and tactical organization of the U-boats in their search for convoys should be directed by the Officer Commanding U-boats, but that the command of the actual operation should be delegated to a subordinate commander in a U-boat situated at some distance from the enemy and remaining as far as possible on

> the surface. I therefore insisted that a certain number of U-boats under construction should be equipped with particularly efficient means of communication which would enable them to be used as command boats.
>
> 3. [The programmed force of U-boats would be inadequate.][54]

Dönitz recounts that his belief that his adversaries would use convoys "was not generally held."[55]

In May 1939, Dönitz performed a wolf pack versus convoy experiment in the Atlantic, based on a scenario of war with Great Britain, even though his superiors insisted that such a war could not possibly occur. In this experiment, the convoy had escorts. In July, Dönitz's submarines practiced against the German surface fleet and its supply vessels as they made a training cruise, despite the protestations of his commanding officer that a war with Great Britain could not occur under any circumstances.[56]

A Limited Technical Assessment

Kahn describes[57] a test of a German antiaircraft weapon, perhaps the famous 88mm Flieger abwehr kanone (FLaK) gun. In today's terminology of military experimentation, this test would be a limited technical assessment (LTA). Theory had predicted that each shot from the gun would have a 25-percent chance of hitting an airplane, and a proving-ground test showed the same result. Of course, wartime experience would demonstrate that the true probability of hitting an airplane with a given shot was about a thousand times less than this.

> • The gun and its concept of operation made great demands on the crew, both physically and mentally. The LTA used "select crews, [who] might well have been described as athletes with Ph.D.'s in physics."
>
> • Only one target aircraft was presented at a time,

but "the battery was a complicated affair requiring the split-second coordination of six people. One man found the height of the plane, another man found its speed, a third man estimated the range, . . . as soon as there are two planes in the sky simultaneously one man will find the height of one plane while the second man is finding the speed of the other."

- The test used timing fuzes, a critical component, that were specially made and of quality greater than could be mass produced.
- The targets always flew at an altitude at which the gun worked well. Kahn observes that during the war, Allied aircraft flew twice as high to disadvantage the antiaircraft guns. One might add that by always presenting aircraft at the same altitude, the test greatly simplified the important and difficult task of estimating the aircraft's altitude.

Perhaps there is room in a development program for a test such as that described by Kahn; the Germans' mistake lay in considering the results of the test to be the last word, rather than the first.*

Observations on the German Experiments

It is interesting to note that the World War II ideas of the blitzkrieg and the wolf pack were both initially conceived even before the First World War had ended. Moreover, those involved in the interwar experimentation—and in the subsequent Second World War application of the experiments' results—had personal combat experience in World War I.

The Germans seem to have benefited from a systematic and serious approach, especially with regard to surrogates. From an experiment in which the surrogate tanks bogged down, some experimenters might have concluded that tanks can only advance

*Or second, counting the theoretical calculation.

in favorable terrain. In contrast, the Germans kept in mind that their tanks were only surrogates, and that one experiments *with* surrogates rather than *on* them. They retained the concept of the tank as an all-terrain vehicle, and worked on improving the surrogates. The seriousness with which experimentation was taken at all levels can be seen in the story of the aerial observer on his surrogate airplane, a motorcycle. In a less serious effort, the troops among whom he was moving would likely have taken him prisoner or at least obstructed his progress, and—on the other side of the coin—he would have been allowed to report all he saw and heard, not just what could have been observed had he been airborne.

Dönitz's clear-sightedness is remarkable. Not only did he correctly anticipate a war with Great Britain and turn to the radically new wolf pack, he also perceived that the correct countermeasure to the massed U-boats would be the escorted convoy* and therefore tested his tactics against convoys, and eventually against convoys with escorts. In the torpedo boat, he saw enough parallels to U-boats to base submarine tactics on those of the surface boats, and perhaps even use the torpedo boat as a surrogate U-boat in experiments, despite its lack of the submarine's most obvious trait—the ability to submerge.

At the end, he was somehow able to articulate a force level requirement, 300 submarines.[58] This is not the place for a thoroughgoing reanalysis of the U-boat war, but it is worthwhile to note that Dönitz had separately estimated that in a war with Britain, his submarines would have to sink two-thirds of a million tons of shipping per month. Germany started the Atlantic phase of the war in 1939 with slightly fewer than 60 oceangoing submarines, some not yet operational, and sank an average of about one-sixth of a million tons of shipping per month through the end of 1941, when the United States entered the war and everything changed. Only at the very end of this period did new construction (and the

*Note that Dönitz saw the value of convoying before the war, whereas Allied navies were skeptical even after the war had begun.

training pipeline) provide more operational boats than were being lost. Thus a very rough calculation suggests that one U-boat would sink a long-run average of (1/6 million)/60, or about 3,000 tons per month (including the substantial fraction of its time spent in port), and so to sink two-thirds of a million tons of shipping per month, 240 submarines would be needed:

$$(2/3 \times 10^6)/[(1/6 \times 10^6)/60] = 240$$

Which is not at all far off from Dönitz's estimate of 300.* Thus Dönitz's experience, perspicacity, and experimentation led him to a substantially correct appreciation of submarine operations in the war to come.

*This calculation addresses only Dönitz's ability to predict, based in part upon experimentation, the force level he would need in order to inflict the level of shipping losses that he sought. Whether or not that level of shipping loss would have led to the defeat of Great Britain is a separate issue.

Overall Observations

The accounts of particular instances of military experimentation have concluded with some commentary on the individual case at hand. This section makes observations that apply more broadly.

Recapitulating the Themes

We began by defining an experiment in terms of a question, with a set of possible answers; an event, with a set of possible outcomes; and a prestated matching between the outcomes and the answers. Several themes were then stated regarding models and modeling (and a paradox about them), surrogates, artificialities, theory, hypothesis, serendipity, cheating, and the difference between a fair experiment and a fair fight.

The idiosyncratic efforts of civilians included Gause, whose biological experimentation was *crucial* in that its outcome would either validate or falsify the very theory on which it was being conducted. Bywater's methods are obscure, but he clearly derived some benefit—not wholly serendipitous, because he set out to find a doctrine for naval victory and he found one—from watching the movements of his wind-up surrogates that it could not derive merely from the workings of his formidable imagination. Of all the efforts examined here, Pratt's work comes the closest of all to the creation of a combat model as the term is normally understood. As such, his game's artificialities and surrogates (if not susceptability to cheating!) were doubtless closely examined, and because his games were largely social events, his is the only case in which the fair fight was an acceptable goal.

The U.S. services' experiments made wide use of surrogates—the Fleet Problems' various stand-ins for carriers and the Louisiana Maneuvers' dummy tanks and guns—but do not seem to have always benefited from a clear understanding of the artificialities these introduced. The effectiveness of the one-plane airstrikes and the various oversold antitank weapons were exag-

gerated, and perhaps it is no accident that if these efforts had agendas, they were to further carrier aviation on the one hand, and to dampen enthusiasm for tanks on the other. In the latter regard, it is interesting to note that the Louisiana exercises seem to stand as the only example of experimentation undertaken with a mindset of debunking innovation rather than fostering it. The anecdote of the squad "swimming" across the bridge is a veritable parable of military experimentation, illustrating at once the ease with which cheating can be rationalized, and the way in which it creates artificialities that can short-circuit the entire benefit of the experiment.

Mitchell's bombing of the *Ostfriesland* was well thought out in terms of its question (can bombs sink battleships?), its event (bombing a battleship), and the possible outcomes (it sinks or not), but the answer was clouded by artificialities (notably the lack of damage control) that verged on unfair experimentation (if not cheating), as well as by interservice friction that Mitchell seems to have made no effort to reduce, so his results were not fully accepted at the time. Nor did wartime events bear him out; of the experiments considered here, his are the only ones (counting the ship-finding and ship-bombing as two separate experiments) whose results do not hold up under the weight of subsequent World War II experience. The FDO school's simulation of carrier air battles is probably the best thought-out *physical* surrogate considered here, in that the tricycles' speeds and fields of view were derived mathematically so as to result in engagement timelines congruent to those arising in actual air attacks on carriers. The Fleet Problems, on the other hand, are notable for being structured so that—in an instance of the "modeling paradox"—the general lack of mathematical correctness did not adversely influence that of the conclusions. The Marine Corps' experimentation constituted a relatively straightforward test of the validity of the *Tentative Manual for Landing Operations*, made the least use of surrogates, and was therefore the least at risk from debilitating artificialities.

The German experiments, too, used surrogates, and were exemplary with regard to controlling the considerable artificialities arising, or threatening to arise, from their surrogates. The torpedo boat was (either in experimentation, or simply as an aid to Dönitz's thinking, whichever the case may have been) a brilliant choice of surrogate for the wolf-pack experiments, in that its major artificiality (the inability to submerge) was moot in the context of the operational concept under test. Other U-boat concepts of operation could not have been tested with torpedo boats, but Dönitz's could be—and may have been. The experiments of the German Army suffered from numerous artificialities, but it was conscious of them and appears to have dealt with them well. Certainly the troops and umpires in the field seem to have followed the rules. It would be interesting to try to trace the major failing of Reichswehr experimentation (its nondiscovery of the effectiveness of combined arms, without which it can be said to have only partially developed the blitzkrieg) to faults in experimental technique, but it is also possible that the discovery of combined arms is an advanced finding toward which the experiments had to work their way through a sequence of more basic findings that consumed the time available.

Although almost all the efforts benefited from serendipity, *all* those examined here started out trying to prove or disprove a theory, or to test a proposed course of action or piece of equipment. Nobody made progress by just going out and trying something to see what would happen; the U.S. Marine Corps' mock battles staged on Civil War battlefields were instances of this, separate from the effort to learn about amphibious warfare, and they don't seem to have contributed to knowledge at all, although of course they were probably good training for the troops.

The Experiments' Points of Similarity

An important point held in common by almost all of the efforts is that they were trying to assess the effect of major changes, or—

more generally—they were addressing big questions. The various inadequacies of their experimental methods did not matter because they were trying to address large issues, not details. The exceptions are Fletcher Pratt and General Billy Mitchell, who asked, respectively, if a pocket battleship could be defeated by three cruisers, or if a true battleship could be defeated by bombers. Pratt's wargame, informal though it was, was based on a huge body of empirical knowledge of the interaction of naval guns and armor. The game integrated that knowledge and served as a method for exploring its consequences. Mitchell's experiment was not conceptual at all, and thus completely dependent upon success in getting the details correct.

It is interesting that so many of the experiments address issues of command and control: C_2 (later C_3, and still later C_3I, etc.) is often considered to be a Cold War or even post–Cold War preoccupation, but it figures as prominently in the experiments of Dönitz and the Pacific Fleet Fighter Direction Officer School as it does in the experiments of the present-day Marine Corps Warfighting Laboratory, and it was an important secondary aspect of most of the other experiments.

The Paradox of Modelling, Resolved

We have seen that several efforts at experimentation have gotten useful and correct answers to big questions, without the benefit of a solid basis in fact. For example:

- The Fleet Problems discovered, and solved, many problems in carrier warfare despite the lack of a good understanding of the effects of bombs on large ships, and despite the defect that their one-plane "squadrons" were unduly hard for the defenders to detect.
- Bywater conceptualized a successful philosophy of naval operations by playing with wind-up boats on a pond.

- Marine Corps junior officers drew the *Tentative Manual for Landing Operations* out of thin air, thickened only by an iterative review process.

As articulated earlier in the "Themes" section, by-the-bootstraps efforts to acquire knowledge (of which all the experiments described here are examples, some to a lesser degree than the three cited above) seem to be a violation of the precept, "Garbage in, garbage out."

The resolution of this paradox is that although the experimental set-up may have been wrong about particulars, the questions under study would have the same answers regardless of the choice of particulars. Continuing with the Fleet Problems as an example, the lessons learned in them (e.g., the need for carriers to operate independently of battleships, the utility of combat air patrol, and above all the primacy of the offensive in carrier warfare) would hold under a very wide range of technological implementations.

This line of reasoning should, in fact, not be alien to either the analyst or the military officer. The military officer is likely to have been taught that one can learn military art from the study of history, despite the fact that the weapons of the historical people worked differently from the weapons of today. The analyst's background in physical science should include the observation that the content of most major subfields (e.g., fluid dynamics) is based on premises that are really only working fictions. As Francis Bacon, often cited as the inventor of the "scientific method," said, "Truth arises more readily from error than from confusion." It is for this reason that cheating is destructive of experiments; it replaces the rules, simplistic though they may be, with confusion.

The process of modeling has been compared to the art of cartooning, on the basis that the skill lies in knowing what to leave out. A more radical simile would be impressionist painting. The whole picture evokes the whole subject, even though—when viewed closely—no part of the picture resembles any part of the subject.

In some cases, the paradox does not really exist:

- Sometimes the purpose of the experiment was to demonstrate an "existence theorem" (i.e., a statement that something could possibly exist). Fleet Problem XIII's successful air attacks on submarines may have suffered from a variety of aspects of unrealism, but they at least showed that it could possibly be the case that aircraft could threaten submarines, a proposition that Dönitz, so prescient in other areas, had famously denied: "The aircraft can no more eliminate the submarine than the crow can fight a mole."[59]
- Sometimes the fact that the experiment may have given the wrong answer is not important, because the goal was to find questions, not answers.

The most radical version of this view would see experimentation as an extension of the Socratic Method: the participants can realize the truth for themselves if only they can be stimulated to think about it correctly, and the experiment exists to provide the stimulus.[60] Bywater's activity with the wind-up boats is a perfect example of this approach. As a template for experimentation (vice training), however, this view requires that the participants, after the revelations afforded them by the experiment, record the truth as they have learned it, or at least make it known to somebody who can record it. A mere account of what happened will certainly not suffice, because the actual events were shaped by all the artificialities and shortcomings of the experiment, and very possibly also by the prerevelation mistakes of the participants.

Notes

1. See Donald Rumsfeld, *Quadrennial Defense Review*, chapter V; Henry Shelton, *Joint Vision 2020*; William S. Cohen, *Annual Report*, chapter 11. Congress recently mandated the gigantic Millennium Challenge 2002 experiment.

2. See also Karppi and McCue.

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

4. Dupuy, 1987.

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

6. B. O. Koopman, *Search and Screening*.

7. Various sources are given for this quote, which seems to have originated with the little-known Francis Edward Smedley.

8. Typical applications of this phrase appear on pages 35 and 40 in *The Voyage of the* Devilfish, the novel by DiMercurio.

9. See Gause, Epstein, and McCue.

10. See Lanchester.

11. Bywater, quoted by Honan, 116.

12. See Honan.

13. See Pratt, and also Allen, 116–19.

14. See Vlahos.

15. Wilson, 75ff.

16. Gause, 7.

17. Gause, vii.

18. EINSTein Project: An Artificial-Life "Laboratory" for Exploring Self-Organized Emergence in Combat, documenation and program available at EINSTein web site, http://www.cna.org/isaac.

19. Except as otherwise noted, material in this section is based on MacDonald.

20. James M. Grimes, quoted in MacDonald.

21. Watts and Murray, 402n. Quote within the quote is from Friedman, Hone, and Mandeles, "The Introduction of Carrier Aviation into the U.S. Navy and the Royal Navy."

22. MacDonald, 35.

23. Layman.

24. Mitchell, 64.

25. Zimmerman.

26. Mitchell, 64.
27. Zimmerman.
28. Zimmerman.
29. This section drawn from Isely and Crowl.
30. Isely and Crowl, 36.
31. See Gabel.
32. Eisenhower, quoted in Gabel, 47–48.
33. Gable.
34. Gabel, 64.
35. Boslaugh, 35ff.
36. See Zimmerman.
37. Layman.
38. These vehicles are also mentioned by Guderian himself, in a section entitled, "The Era of the Dummy Tanks," 160ff.
39. Corum, 186.
40. Corum, 187.
41. Guderian, 162.
42. Guderian, 162–63.
43. Corum, 149.
44. Corum, photograph, 162.
45. Quoted in Corum, 165–66.
46. Corum, 164.
47. Corum, 202ff.
48. See Dönitz, chapter 3 ("Wolf-Pack Tactics"), which makes it clear that the use of U-boats in groups had been thought of by the Germans during World War I, but never put into practice.
49. These ideas are discussed by Kuenne, Gardner, and McCue (2005). Their attribution to Dönitz is more problematic; several sources (e.g., Frank, Keegan, and many others) describe Dönitz as thinking these thoughts, often in the context of his loss, in the First World War, of the submarine *U-68*, but do not provide solid references. Keegan (224) provides a reference, but it dates from 1939. He mentions (223) and even quotes pack-oriented doctrine statements applied to destroyers, but Dönitz seems to have been largely a recipient of this doctrine rather than a source.
50. Dönitz, 20.
51. Padfield says there is "even the possbility that some of the exercises were actually designed to study the problem of U-boat surfaced attack. No direct

evidence to support this has appeared, but . . .” and then goes on to list quite a number of pieces of circumstantial evidence (101).

52. Dönitz, 19.

53. This paragraph drawn from Frank, 23, and Dönitz, 21.

54. Dönitz, 33.

55 Dönitz, 34.

56. This paragraph drawn from Frank, 25. See also Dönitz, 21, and Keegan, 224. The latter gives the incorrect impression that the May 1939 experiment was the first test of U-boat wolf-pack tactics.

57. Kahn.

58. Frank, 25; Dönitz, 33.

59. Dönitz so wrote in his war diary in August 1942. Cited by Price, 87.

60. This view is close to that of Schrage.

Bibliography

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

Boslaugh, David L. *When Computers Went to Sea*. Los Alamitos, CA: IEEE Computer Society and the Society of Marine Architects and Engineers, 1999.

Bywater, Hector. *The Great Pacific War: A History of the American-Japanese Campaign of 1931–33* (novel). Boston: Houghton Mifflin, 1925. Reprinted variously since, most recently Applewood, 2002.

Cohen, William S., Secretary of Defense. *Annual Report to the President and Congress.* 2001.

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

Davis, Paul K., and Donald Blumenthal. *The Base of Sand Problem: A White Paper on the State of Military Combat Modeling*. Santa Monica, CA: RAND Corporation, 1991.

DiMercurio, Michael. *The Voyage of the* Devilfish. New York: Onyx, 1993.

Dönitz, Karl, and R. H. Stevens, translator. *Memoirs: Ten Years and Twenty Days*. London: Weidenfeld and Nicholson, 1959. Originally published 1958.

Dupuy, Trevor N. *Numbers, Prediction, and War*. New York: Bobbs-Merrill, 1979, and other editions.

———. *Understanding War*. New York: Paragon House Publishers, 1987.

Epstein, Joshua. *Nonlinear Dynamics, Mathematical Biology, and Social Science*. Reading, PA: Addison-Wesley, 1997.

Felker, Craig C. *Testing American Sea Power: U.S. Navy Strategic Exercises, 1923–1940*. College Station: Texas A&M University Press, 2007.

Frank, Wolfgang. *The Sea Wolves*. New York: Ballantine, 1959.

Friedman, Norman, Thomas C. Hone, and Mark D. Mandeles. "The Introduction of Carrier Aviation into the U.S. Navy and the Royal Navy," draft report prepared for the Director, Net Assessment, Office of the Secretary of Defense, 12 May 1994, cited in Watts and Murray.

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

Gardner, W. J. R. *Decoding History: The Battle of the Atlantic and ULTRA*. Annapolis, MD: Naval Institute Press, 1999.

Gause, G[eorge] F[rancis]. *The Struggle for Existence*. Baltimore: Williams and Witkins, 1934. Dover reprint, 1971.

Guderian, Heinz, and Christopher Duffy, translator. *Achtung—Panzer!* London: Arms & Armour, 1992. Original edition, 1937.

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

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

Karppi, Helen, and Brian McCue. "Experiments at the Marine Corps Warfighting Laboratory." *Marine Corps Gazette* 85, no. 9 (September 2001).

Keegan, John. *The Price of Admiralty*. New York: Viking Press, 1998.

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

Koopman, Bernard Osgood. *Search and Screening.* OEG Report 56. Washington, DC: Operations Evaluation Group, 1946. (Available from the Center for Naval Analyses.) Major revision, New York: Pergamon, 1980.

Kuenne, Robert E. *The Attack Submarine: A Study in Strategy*. New Haven, CT: Yale University Press, 1965.

Lanchester, Frederick W. *Aircraft in Warfare: The Dawn of the Fourth Arm*. New York: D. Appleton & Co., 1916. Nearly the whole of the small portion that relates to attrition modeling may be found in: James R. Newman, editor. *The World of Mathematics*, volume 4, 2138–57. New York: Simon and Schuster, 1956, with a commentary on 2136–37.

Layman, R. D. "The Day the Admirals Wept." *Warship* XIX (1995): 75ff.

MacDonald, Scot. "Evolution of Aircraft Carriers." *Naval Aviation News*; Feb, Mar, May, Jun, Aug, and Sep, 1962.

McCue, Brian. "Subsurface Combat In A Test Tube: Building upon the Work of G. F. Gause," *Military Operations Research* 17, no. 1 (2012): 45–57.

———. "Subsurface Combat In A Test Tube: Re-analyzing the Work of G. F. Gause," *MORS Phalanx* 43, no. 3 (September 2010): 22–24.

———. "A Chessboard Model of the U-Boat War in the Atlantic, with Applications to Signals Intelligence." *Naval Research Logistics* 52, no. 2 (March 2005): 109–136.

Mitchell, William. *Winged Defense*. New York: G. P. Putnam's Sons, 1925. Dover Publications, 1988 reprint.

Morse, Philip M., and George E. Kimball. *Methods of Operations Research*. Washington, DC: Operations Evaluation Group, 1946. 1951 edition recently reprinted by the Military Operations Research Society.

Murray, Williamson, and Allan R. Millett, editors. *Military Innovation in the Interwar Period*. Cambridge: Cambridge University Press, 1996. Paperback, 1998.

Nofi, Albert. *To Train the Fleet for War: The U.S. Navy's Fleet Problems, 1923–1940*. Newport, RI: United States Naval War College, 2010.

Padfield, Peter. *Dönitz: The Last Führer*. New York: Harper and Row, 1984.

Pratt, Fletcher. *Fletcher Pratt's Naval War Game*. New York: Harrison-Hilton Books, 1940. Republished, Milwaukee: Z & M Publishing Enterprises, 1973.

Price, Alfred. *Aircraft Versus Submarine*. New York: Jane's Information Group, 1980.

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

Schrage, Michael. *Serious Play*. Boston: Harvard Business School Press, 2002.

Shelton, Henry H. Chairman of the Joint Chiefs of Staff, *Joint Vision 2020.*

Vlahos, Michael. "Wargaming, an Enforcer of Strategic Realism: 1919–1942." *Naval War College Review* 39 (Mar–Apr 1986): 7–22.

Watts, Barry, and Williamson Murray. "Military Innovation in Peacetime," 369ff. in Murray and Millett.

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

Zimmerman, Gene T. "More Fiction than Fact—The Sinking of the *Ostfriesland." Warship International* 12, no. 2 (1975): 142–54.

MARINE CORPS
UNIVERSITY PRESS

CNA
ANALYSIS & SOLUTIONS