1982

Did We Learn Anything From That Exercise? Could We?

Frederick Thompson

Recommended Citation

Thompson, Frederick (1982) "Did We Learn Anything From That Exercise? Could We?," *Naval War College Review*: Vol. 35 : No. 4 , Article 5.
Available at: https://digital-commons.usnwc.edu/nwc-review/vol35/iss4/5

Thompson: Did We Learn Anything From That Exercise? Could We?

25

Did We Learn Anything From That Exercise? Could We?

by

Frederick Thompson

Exercises are a source of information on tactics, force capabilities, scenario outcomes, and hardware systems effectiveness. But they distort battle operations in ways which prevent the immediate application of their results to real world situations. For example, because they are artificial, the force capabilities demonstrated need not exactly portray capabilities in actual battle. Further, our analysis process is imperfect. Our data can be incomplete or erroneous, the judgments we make during reconstruction and data refinement may be contentious, and our arguments linking the evidence to our conclusions may be incorrect. Still, exercises are among the most realistic operations we conduct. Our investigations of what really happened in an exercise yield valuable insights into problems, into solutions, and into promising tactical ideas.

The Nature of Exercises

How do naval exercises differ from real battles? Clearly the purpose of each is different. Exercises are opportunities for military forces to learn how to win real battles. During exercises, emphasis is on training people and units in various aspects of warfare, practicing tactics and procedures, and coordinating different force elements in complex operations. Ideally, the exercise operations and experiences would be very much like participating in a real battle. For obvious reasons, exercises fall short of this ideal, and thereby distort battle operations. These distortions are called "exercise artificialities." An understanding of the different kinds of exercise artificialities is essential to understanding exercise analysis results. The exercise artificialities fall loosely into three classes: those which stem from the process of simulating battle engagements; those which stem from pursuit of a specific exercise goal; and those stemming from gamemanship by the players.

Engagement Simulation. Obviously real ordnance is not used in an exercise. As a result, judging the accuracy of weapon delivery and targeting, force attrition, and damage assessment become problems in an exercise. If a real SAM is fired at an incoming air target, the target is either destroyed or it is not. There is no corresponding easy solution to the engagement in an exercise. Somehow, the accuracy of the fire control solution must be judged, and an umpire must determine whether the warhead detonates and the degree of destruction it causes.

What is the impact of this simulation on battle realism? Suppose the SAM is judged a hit and the incoming target destroyed. The incoming target will *not* disappear from

radar screens. It may, in fact, continue to fly its profile (since it won't know it's been destroyed). So radar operators will continue to track it and the target will continue to clutter the air picture. A cluttered air picture naturally consumes more time of operators and decision makers. Now suppose the SAM misses the incoming target. If time permitted, the SAM ship would fire again, thereby depleting SAM inventories. However, the judgment process is not quick enough to give the SAM ship feedback to make a realistic second firing. In fact, AAW engagement resolution may not occur until the post-exercise analysis.

Now suppose the SAM misses the incoming missile, but the missile hits a surface combatant. Then the problem is to figure out how much damage was done to the combatant. An umpire will usually role dice to probablistically determine damage; a real explosion wreaks destruction instantaneously. As a result, there will be some delay in determining damage and even then that damage may be unrealistic.

It is easy to see how the flow of exercise events may become distorted, given the delay between engagement and engagement resolution during an exercise. Other examples of distortion abound. For example, it may happen that a tactical air strike is launched to take out an opposing surface group armed with long-range antiship missiles, but only after those missiles have already dealt a crippling blow to the CV from which the air strike comes. In another case, aircraft will recover on board CVs with simulated damage as well as those CVs still fully operational. In general, it has so far been impossible to effect in exercises the real-time force attrition of an actual battle so that battle flow continues to be realistic after the first shots are fired.

Such artificialities make some aspects of the exercise battle problem more difficult than in a real battle; others make it less difficult. Because destroyed air targets don't disappear from radar screens, the air picture becomes more complicated. On the other hand, a SAM ship will seldom expend more than two SAMs on a single target and therefore with a given inventory can engage more incoming missiles than she would be able to in reality. Further, the entire AAW force remains intact during a raid (as does the raid usually) as opposed to suffering progressive attrition and thereby having to fight with less as the raid progresses. It is unclear exactly what the net effect of these artificialities is on important matters like the fraction of incoming targets effectively engaged.

Safety restrictions also distort exercise operations and events. For example, the separation of opposing submarines into nonoverlapping depth bands affects both active and passive sonar detection ranges, especially in submarine vs. submarine engagements. Here, realism is sacrificed for a reduced probability of a collision. For the same sorts of reasons, aircraft simulating antiship missiles stay above a fixed altitude in the vicinity of the CV, unless they have prior approval from the CV air controller, which distorts the fidelity of missile profiles. In other examples, surface surveillance aircraft may not use flares to aid night time identification. Tactical air strike ranges may be reduced to give the strike aircraft an extra margin of safety in fuel load. The degree of battle group emission control, especially with regard to CV air control communications and navigation radars, is determined partially by safety consideration. Quietness is often sacrificed in favor of safety.

The point is that safety is always a concern in an exercise, whereas in an actual battle, the operators would probably push their platforms to the prudent limits of their capabilities. These safety restrictions impart another artificiality to exercise

Thompson: Did We Learn Anything From That Exercise? Could We?

Did We Learn Anything? 27

operations. By constraining the range of Blue operational options, his problem becomes more difficult than the real world battle. By constraining Orange operational options, Orange's problem becomes harder, and hence Blue's problem easier, than in the real world.

Another source of distortion is the use of our own forces as the opposition. US naval ships, aircraft, and submarines differ from those of potential enemies. It is probable that enemy antiship missiles can be launched from further away, fly faster, and present a more difficult profile than can be simulated by manned aircraft in an exercise. The simulated antiship missile in an exercise thus presents an easier target in this regard. Customarily, Orange surveillance has fewer platforms with less on-station time than do some potential enemies, say the Soviets. In ASW, there may be differences between US submarine noise levels and potential enemy submarine noise levels. All these differences in sensors and weapon systems distort detection, identification, and engagement in exercises and thereby make aspects of exercise operations artificial.

A more subtle distortion occurs when US military officers are cast in the role of enemy decision makers. The US officers are steeped in US naval doctrine, tactics, and operating procedures. It is no doubt difficult to set aside these mind-sets and operate according to enemy doctrine, tactics and procedures. Add to this the fact that one has only a perception of enemy doctrine, tactics, and procedures to work from, and the operating differences between an actual enemy force and a simulated enemy force become more disparate. With this sort of distortion, it is difficult to identify exactly how the exercise operations will be different from those they try to simulate. But the distortions are real, and are at work throughout the exercises.

Exercise Scenarios. The goal of an exercise can drive the nature of the exercise operations; this is a familiar occurrence in all fleets. The degree of distortion depends upon the nature of the goal. Consider two examples.

First, consider an exercise conducted for the express purpose of examining a particular system's performance in coordinated operations. It is likely to involve a small patch of ocean, repeated trials in a carefully designed scenario, dictated tactics, and most importantly a problem significantly simpler than that encountered in a real battle. At best, the battle problem in this controlled exercise will be a subtask from a larger battle problem. Participants know the bounds of the problem and they can concentrate all of their attention and resources on solving it. Now such exercises are extremely valuable, both in providing a training opportunity to participants and in discovering more about the system in question. But the exercise results apply only in the limited scenario which was employed; in this sense the goal of the exercise distorts the nature of the operations. Exercise operations in these small, canned, controlled exercises are artificially simple as compared to those in a real battle.

Next consider a large multithreat free-play exercise which is conducted partially for training, perhaps the most realistic type of exercise conducted. The exercise area will still have boundaries but will probably include a vast part of the ocean. Commercial aircraft traffic and shipping may well be heavier than would be the case in a hot war environment. As the exercise unfolds there will be a tendency for the controlling authority to orchestrate interactions. By doing this, the options are constrained unrealistically for both sides. Blue or Orange may not be able to pick the time and place to fight the battle. Both sides know that a simulated battle will be

fought, and higher authority may hasten the interaction so that the participants can fight longer. Clearly this is a case where trade-offs must be made, and it is important to understand this when exercise results are being interpreted.

In both kinds of exercises, artificialities are necessary if the goals are to be met. Partly as a result, the operations are not exact duplicates of those likely to occur in the same scenario in a real battle. Aside from recognizing that forced interactions distort an exercise battle, little work has been done to learn more about how these distortions affect the resulting operations.

Gamesmanship and Information. A separate class of artificialities arise when exercise participants are able to exploit the rules of play. Consider a transiting battle group. It may be possible to sail so close to the exercise area boundary that from some directions the opponent could attack only from outside the area, and that is prohibited. Thus, the battle group would reduce the potential threat axes and could concentrate its forces only along .xes within the operating area. Clearly, the tactical reasoning which leads to such a decision is valuable training for the participants, and exploiting natural phenomena such as water depth, island masking, and so on are valid tactics. But exploiting an exercise boundary to make the tactical problem easier distorts operations in the scenario and is a kind of gamesmanship.

Consider another situation. In exercises, both sides usually know the opposition's exact order-of-battle. So participants have more information than they are likely to have in a real battle, and that information is known to be reliable. Blue also knows the operating capabilities of the ships simulating the enemy, and may be able to deduce associated operating constraints from them. For example, he knows more about US submarine noise levels and operating procedures than he does about likely opposition submarines. He also knows how many Orange submarines are actually pariticpating in the exercise, and as he engages them, he may be able to estimate the size of the remaining threat by examining time and distance factors.

Classes of Exercise Artificiality

   I. Battle Simulation Artificiality
   - No Real Ordnance
   - Safety Restrictions on Operations
   - Simulate Opposition Platforms
   - Imperfect Portrayal of Enemy Doctrine, Tactics, and Procedures

   II. Scenario Artificiality
   - Forced Interaction
   - Focus on Small Piece of Battle Problem

   III. Gamesmanship and Information
   - Exact Knowledge of Enemy OOB
   - Exact Knowledge of Enemy Platform Capabilities
   - Exploitation of Exercise Rules
   - Tactical Information Feedback Imperfections

Thompson: Did We Learn Anything From That Exercise? Could We?

Did We Learn Anything? 29

A final exercise artificiality is the poor information flow from battle feedback. With real ordnance, undetected antiship missiles hit targets, explode, and thereby let the force know it is under attack. This does not occur in exercises. A force may never know it has been located and has been under attack. As a result, the force may continue to restrict air search radar usage when in a real battle, all radars would have been lit off shortly after the first explosion. The force may never be able to establish a maximum readiness posture. In a real battle, there would have been plenty of tactical information to cue the force that it is time to relax emission control. This kind of exercise artificiality affects both the engagement results and the flow of battle events.

In spite of these artificialities, exercises still provide perhaps the only source of operational information from an environment which even attempts to approximate reality. Though artificial in many ways, exercises on the whole are about as realistic as one can make them, short of staging a real war. This is especially true in the case of large, multithreat, free-play fleet exercises. The only time a battle group ever operates against opposition may be during these exercises. So for lack of something better, exercises become a most important source of information.

The Nature of the Analysis Process

The analytical conclusions drawn from examining exercise operations are the output of a sequence of activities which collectively are called the exercise analysis process. While there is only one real analytical step in the process, it has become common to refer to the entire sequence as an analysis process. The individual steps themselves are (1) data collection, (2) reconstruction, (3) data reduction and organization, (4) analysis, and (5) reporting. It is of immense value to understand how the body of evidence supporting exercise results and conclusions is developed. We will examine the activities which go on in each step of a typical large, multithreat, free-play fleet exercise, and end with some comments to make clear how the analyses of other kinds of exercises may be different.

Data collection. The first step is to collect data on the events that occur during an exercise. Think of the exercise itself as a process. All the people in the exercise make decisions and operate equipment and so create a sequence of events. The data which are collected are like measurements of certain characteristics of the process, taken at many different times during the exercise. The data are of several different types.

One type is keyed to particular events which occur during the exercise: a detection of an incoming missile, the order to take a target under fire, the act of raising a periscope on a submarine, a change in course, and so on. This sort of data is usually recorded in a log along with the time of its occurrence. Another kind of data is the perceptions of various participants during the exercise. These data are one person's view of the state of affairs at one point in time. The individual could be an OTC, a pilot, or a civilian analyst observer. Another type of data is the evaluative report of a major participant, usually filed after the exercise is over. These provide the opinions of key participants on the exercise, on a particular operation and what went wrong, on deficiencies, etc. Finally, the memories of participants and observers also are a source of data. Their recollections of what went on during a particularly important period of the exercise may often be valuable.

There are two kinds of imperfections attendent to all this. The first is imperfections in the data collected: they don't reflect accurately what they were

intended to reflect. That is, some data elements are erroneous. The second imperfection stems from having taken measurements only at discrete points in time, and having only partial control over the points in time for which there will be data. A commander in the press of fighting a battle may not have the time to record an event, or his rationale for a crucial decision. An observer may likewise miss an important oral exchange of information or an important order. After the exercise is over, memories may fade and recollections become hazy. So the record of what went on during the exercise, the raw data, is imperfect.

Once most of the raw recorded data are gathered in one place, reconstruction begins. In general, gross reconstruction provides two products: geographical tracks of ships and aircraft over time, and a chronology of important exercise events: time and place of air raids, submarine attacks, force disposition changes, deception plan executions and so forth. Tentative identification of important time periods is made at this time. These periods may become the object of finer grained reconstruction later as new questions emerge which the gross reconstruction is unable to answer. The table below lists the primary products of gross reconstruction. The major event chronology includes the main tactical decisions such as shifts in operating procedures, shifts in courses of action, executions of planned courses of action, and all others which might have affected what went on.

Reconstruction is arguably the most important step in the exercise analysis. Many judgments are made at this level of detail which affect both the overall picture of what went on during the exercise as well as the validity of the results and conclusions. It is much the same kind of laboratory problem scientists face in trying to construct a data base from a long, costly series of experiments. The basic judgments concern resolving conflicts among the data, identifying errors in data entries, and interpreting incomplete data. Judging each small case seems minor. However, the enormous *number* of small judgments collectively have a profound effect on the picture of exercise operations which emerges. The meticulous sifting which is required demands knowledgeable people in each area of naval operations as well as people possessed of a healthy measure of common sense. Judgments made during reconstruction permeate the remainder of the exercise analysis process. These judgments constitute yet another way for errors to enter the process.

Reconstruction Products

Data Source	Product
Ship's Position Logs, Course & Speed Logs, Fixes, DRT, DRAI.	Ship Tracks
A/C position info (range/bearing/time)	A/C Tracks
Ship and A/C Tracks, Narrative Logs, Ships Weapons Use Logs, Sensor Use Logs, Operations Plans and Revisions, Message Files.	Major Event History (ASW action, submarine attacks, surface engagements, WAS strikes, AAW incidents, major tactical decisions)

Thompson: Did We Learn Anything From That Exercise? Could We?

Did We Learn Anything? 31

Data Reduction and Organization. The line between reconstruction and data reduction and organization is blurred. At some point, most of the reconstruction is done and summary tables of information begin to emerge. In antiair warfare for example, tables will show the time of the air raid, raid composition, number deteeted, percent effectively engaged, and by whom. An antisubmarine warfare summary table might show contacts, by whom detected, validity of detection, attack criteria achievement, and validity of any attacks conducted. Other products based upon the reconstruction are tailored to the specific analysis objective or the specific question of interest. For example, in command and control, a detailed history of the flow of particular bits of information from their inception to their receipt by a weapon system operator might be constructed. In surface warfare, the exact sequence of detections and weapon firings are other examples.

Two important acts occur during this step. First certain data are selected as being more useful than other data and then the individual bits are aggregated. Second, the aggregate data are organized into summary presentations (in the form of tables, figures, graphs, and so on) so that relations among the data can be examined. Obviously, the way in which data is aggregated involves judgments as to what data to include and what to exclude. These choices and the selection of the form of the presentation itself involve important judgments. As before, the judgments comprise another potential source of error.

Analysis. Analysis is the activity of testing hypotheses against the body of evidence, constructing new hypotheses, and eventually rejecting some and accepting others according to the rules of logic. While reconstructing, reducing, and organizing data, analysts begin to identify problem areas, speculate upon where answers to questions might lie, and formulate a first set of hypotheses concerning exercise operations. It is now time to examine systematically the body of evidence to ascertain whether the problems are real, whether answers to questions can indeed be constructed, and whether the evidence confirms or refutes the hypotheses. Arguments must be constructed from the evidence, i.e., from the summary presentations already completed, from others especially designed for the hypothesis in question, or from the raw data itself. The construction of such logical arguments is the most time-consuming step in the process and the most profitable. Yet the pressure from consumers for quick results, a justifiable desire, may severely cut down on the time available. In such situations, hypotheses may emerge from this step as apparently proven results and conclusions, without the benefit of close scrutiny. This is an all too common occurrence.

One kind of shortcut is to examine only evidence which tends to confirm a hypothesis. The analyst uses the time he has to construct as convincing an argument as he can in support of a contention. Given additional time, an equally persuasive argument refuting the contention might have been developed. Errors may also enter the analysis in the course of judging the relative strength of opposing bodies of evidence. Where such judgments are made, conventional wisdom would have both bodies of evidence appear along with an argument why one body seems stronger. In these ways the analysis step may introduce additional uncertainty into the analysis process.

Reporting. The final step in the analysis process is reporting. It is during this step that analysts record the fruits of their analytical labors. There are four basic categories of

reports, some with official standing, some without. It is worth defining them, both to give some idea of the amount of analysis which underlies the results and to present the reports most likely to be encountered.

One kind of exercise report is for a higher level commander. It details for him those exercise objectives which were met and those which were not. It is a post-operation report to a superior. Customarily it will describe training objectives achieved (i.e., did the assigned forces complete the designated training evolutions?), the resulting increase in readiness ratings for individual units, and an overview of exercise play and events. There is little, if any, analysis of exercise events to learn of problem areas, tactical innovations, or warfighting capabilities.

Another kind of exercise report is a formal documentation of the product of the analysis process. It concentrates on the flow of battle events in the exercise instead of the "training events." These reports may or may not include word of training objectives achieved and changes in unit readiness. A report might begin with a narrative description of battle events and results for different warfare areas. Summary tables, arguments confirming or refuting hypotheses, and speculations about problems needing further investigation form the bulk of the warfare sections. Conclusions and supporting rationale in the form of evidence from exercise operations may also be present. Bear in mind that the analysis process preceding the report may have been incomplete. In this case the report will include the narrative and customarily a large collection of reconstruction data and summary tables. The report will fall short of marshaling evidence into arguments for, or against, hypotheses. These reports are really record documents of raw and processed exercise data.

It can be difficult to distinguish between these two types of report if the latter also includes items called "conclusions." Beware if there is an absence of argumentation, or if great leaps of faith are necessary for the arguments to be valid. Sometimes one gets the reconstruction plus analysis, other times just the reconstruction.

Units participating in exercises often submit their own message reports, called "Post-ex Reports" or "Commander's Evaluation." These reports seldom include any analytical results or conclusions. They do venture the unit commander's professional opinions on exercise events and operations. These opinions, tempered by years of operational experience, as well as firsthand operational experience during the exercise, are a valuable source of information. They provide the perspective of a particular player on perceived problems, suspected causes, reasons for tactical decisions, and possibly even some tentative conclusions. Statements in these reports should be tested against the data for confirmation. Sometimes the messages also contain statements entitled "Lessons Learned." Since such judgments are based upon the limited perspective of one unit, these lessons learned require additional verification too. The unit CO probably will base this report on some of the data collected by his own unit. So the CO's postexercise report is a view of the exercise based upon a partial reconstruction using one unit's data.

Finally, the Navy Tactical Development and Evaluation (TACD&E) program sanctions reports of exercise results and analyses as a formal Lessons Learned. NWP-0 defines a Lessons Learned as " . . . statements based on observation, experience, or analysis which indicates the state of present or proposed tactics." Note that a Lessons Learned is specific to a tactic or group of tactics. Evidence

Thompson: Did We Learn Anything From That Exercise? Could We?

Did We Learn Anything?   33

developed in an exercise often provides the analytical basis for such statements. NWP-0 goes on to state that " . . . the most useful Lessons Learned are brief case studies which tell what happened and why certain key outcomes resulted." Exercise operations can often provide the "cases," and exercise analysis can provide the "why" certain things happened. Again it is necessary to examine carefully the argumentation in Lessons Learned, to be sure the analysis process applied to the individual cases hasn't been curtailed after the reduction and organization step.

Variations. The analysis process for a small specialized exercise has a slightly different manifestation from that in a large free-play fleet exercise. Consider an exercise designed to test tactics for the employment of a new sonar and to train units how to execute those tactics. It might involve three or four ships outfitted with the sonar pitted against a submarine in a controlled scenario. If there is high interest in innovative ways to employ the system tactically, data collection might be better than average, since many hands can be freed from other warfare responsibilities for data collection. The operating area might be an instrumented range on which very precise ship tracks can be recorded automatically. If the planning is thorough, the design of the exercise (the particular pattern of repeated engagements with careful varying of each important factor) enables just the right data to be collected which will enable analysts to sort among the different tactics. The data which is collected would then leave fewer holes, *relative to the exact questions which are of interest.* So, one might end up with fewer errors in the data, and simultaneously, less missing data.

The quality of reconstruction will still depend on the skill of the reconstructors. With only a few ships to worry about and good data, however, not many people are required to do a good job; the job is small. If the exercise was designed carefully to shed light on specific questions, data reduction and organization work smoothly toward pre-identified goals: specific summary tables, graphs, or figures. In fact from the analytical viewpoint, the whole exercise may as well have been conducted to generate reliable numbers to go into the tables and graphs. The analysis step is more likely to proceed smoothly too, since the evidence has been designed specifically to confirm or deny the questions of interests.

The analysis process of other exercises will likely fall between these two extremes. The degree to which exercise play is controlled and constrained by the operating area's size and by various units' tactical autonomy will determine the ease with which the analysts and data collectors can finish their work. Normally, the analysis is best in the small, controlled exercises designed to answer specific questions or to train units in specific tactics. As the exercise grows in size and more free-play is allowed, it is harder to collect data to answer the host of questions which *may become* of interest.

Limitations on the Use of Exercise Analysis

The reason for analyzing exercise operations is to learn from them. One learns about tactics, readiness levels of units and groups, hardware operational capabilities, and advantages or disadvantages we may face in certain scenarios. Let us see how exercise artificialities and an imperfect analysis process limit what we can learn.

Hardware operational capabilities can be dispensed with quickly. Special exercises are designed to measure how closely systems meet design specifications. The measures are engineering quantities such as watts per megahertz, time delay in a

switching mechanism, sensitivity, and so on. As the human element enters either as the operator of the equipment or in a decision to use the system in a particular way, one moves into the realm of tactics.

Warfare Capabilities. One problem in learning about warfare capabilities from exercises lies in translating the exercise results into those one might expect in an actual battle. Setting aside the measurement errors which may crop up in the analysis process, consider the exercise artificialities. Suppose a battle group successfully engages 70 percent of the incoming air targets. This does not mean that the force would successfully engage 70 percent of an air attack in a real battle. Assuming identical scenarios and use of the same tactics, some artificialities make the exercise problem easier, others make it harder than the real-world battle problem. There is no known accurate way of adjusting for these artificialities. In fact only recently has there been general acceptance of the fact that the artificialities both help and hinder.

A second problem is the lack of a baseline expected performance level for given forces in a given scenario. A baseline level would describe how well one expected a specific force to do, against a given opposition in a given scenario *on average*. One would compare exercise results with baseline expectations to conclude that the exercise force is worse or better than expected. But no such baseline exists; that is there are no models of force warfare which can predict the outcome of an exercise battle. Thus, we don't know what the "zero" of the warfare effectiveness index is; neither do we know the forms of the adjustments necessary to translate exercise results into corresponding real-world results.

One might speculate that it would at least be possible to establish trends in warfare effectiveness from exercises. However, this too is difficult. The exercise scenarios as well as the forces involved will change over time. In any particular exercise, the missions, the geography, the forces (e.g., a CV rather than a CVN), and the threat simulation are likely to be different from those in any other exercise. Some scenarios may be particularly difficult, while others are easy. Comparing across exercises requires a way of adjusting for these differences. It requires knowing how a given force's capabilities change with each of these factors, and right now we don't know how. Of course, solving the problems of adjusting for exercise artificialities and of establishing an expected performance level for given battle problems would be a move in the right direction. But imperfections in the steps of the analysis process compound these conceptual difficulties. Recall that the data are imperfect to begin with, and errors enter during reconstruction and data reduction and organization. The numbers built from these data then have some error associated with them. These are the numbers which appear in summary tables and graphs depicting warfare effectiveness during an exercise. They are imprecise. This means that changes over time, even in exercises with roughly equivalent scenarios, must be large to be significant. Otherwise, such differences might only be statistical variations. Exactly how large they have to be, is still not clear but "big" differences bear further investigation.

What then is the usefulness of such numbers? They are useful because they result from examining the exercise from different viewpoints, and they allow judgment to be employed in a systematic manner. Without them one is completely in the dark. Clearly it is better to merge many different perspectives on how the operations went, than to rely on just one. The analysis process does this by examining objectively data

Thompson: Did We Learn Anything From That Exercise? Could We?

Did We Learn Anything? 35

collected from many different positions. It provides a framework for systematic employment of professional judgment concerning the effect of artificialities on exercise results. Recognizing each artificiality, professional judgment can be applied to assess the influence of each individually as opposed to the group as a whole. While obviously imprecise, the numbers appearing in the summary presentations, together with an understanding of the artificialities, the contextual factors, and the measurement errors, are better than a blind guess.

Evaluating an individual unit's warfighting capability (as opposed to a group's) is not easy either. The normal measures of unit readiness which come out of an exercise are at a lower mission level. An air squadron may have high sortie rates, and may be able to get on and off the carrier with ease, but the question of interest may be how effectively they contributed to the AAW defense. The link between task group AAW effectiveness and high sortie rates or pilot proficiency is not well understood. So while measurements at that level may be more precise than those at a higher level, and while the individual actions are more like actions in a real battle, it is not clear how measures of effectiveness at this level contribute to success at the group or force level. There is a need to research this crucial link between unit performance of low level mission actions and group mission effectiveness.

Tactics. As a vehicle for evaluating tactics, exercise analysis fares pretty well. Exercise artificialities and the analysis process still limit what we, conceivably, could learn and, practically, what we do learn.

The main artificiality to be careful of is threat simulation. Generally there are situations of short duration in an exercise which closely approximate those occurring in real battles, some in crucial respects. It is possible, then, to test a tactic in a specific situation which, except for the threat simulation, is realistic. The tactic may work well in the situation, but would it work against a force composed of true enemy platforms? This may be more problematic.

The limitations due to the analysis process stem more from improper execution rather than flaws in the process itself. To date, exercise analysis has failed to distinguish regularly between problems of tactical theory and those of tactical execution. If the analysis concludes that the employment of a tactic failed to achieve a desired result, it seldom explains why. There is no systematic treatment of whether the tactic was ill-conceived, or employed in the wrong situation, or executed clumsily. The idea of the tactic may be fine, it may only have been employed in the wrong situation or it may have been executed poorly. In the event that a tactic does work, that is, the overall outcome is favorable, scant explicit attention is paid to the strength of the tactic's *contribution* to the outcome. The outcome might have been favorable with almost any reasonable tactic, because, say, one force was so much stronger than the other. Remember too that the data upon which the tactical evaluation is based is the same imperfect data as before. It is true that in some evaluations, the conclusion may be so clear as to swamp any reasonable error level in the data. Even if the error is 30 percent (say in detection range, or success ratio) the conclusion still might hold.

There are certain analytical standards which are achievable for tactics evaluation in exercises. The tactic or procedure should be defined clearly. The analysis should address whether the tactic was executed correctly and whether it was employed in the appropriate situation. It should answer the question of whether the influence of

other contextual factors (aspects of the scenario for example) dominated the outcome. It should identify whether the tactic will only work when some factor is present. It should address whether the tactic integrates easily into coordinated warfare. Even if all these conditions are satisfied, the exercise may only yield one or two trials of the tactic. Definitive tests require more than one or two data points.

Scenarios. Judging how well Blue or Orange does in a scenario depends on the accuracy of the warfare capability assessments, the fidelity of the threat simulation, and the skill with which exercise results can be translated into real world expectations. It is clear from previous discussions on each of these topics that there are problems associated with each. Consequently, what we can learn about a scenario from playing it in an exercise is limited.

At best one can make gross judgments; an example might be "a CVTG cannot long operate from a Modloc in this specific area of the world without more than the usual level of ASW assets." The exercise will provide an especially fertile environment for brainstorming about the scenario, and in a systematic way. The kinds of tactical encounters which are likely to cause problems will surface. Those engagements or operations which are absolutely crucial to mission success may also become clear. Serious, thorough consideration of many courses of action may only occur in the highly competitive environment of an exercise. This can lead to the discovery of unanticipated enemy courses of action.

There are pitfalls of course in making even these gross assessments. For example, care must be taken to recognize very low readiness levels by exercise participants as a major contributor to the exercise outcome. But on the whole it should be possible to identify scenarios which are prohibitively difficult and should, therefore, be avoided. It may be possible to confirm what forces are essential for mission success and the rough force levels required.

What kinds of things might one reasonably expect to learn from exercises? First and foremost, the product of exercise analysis is well suited to correcting misperceptions about what happened during the exercise. It provides a picture of the exercise which is fashioned logically from data taken from many key vantage points instead of just one or two. As such, it is likely to be closer to the truth than a sketchy vision based on the experience of a single participant in the exercise. Second, there is a capability to make some quantitative comment on warfare effectiveness. All the caveats developed earlier in the essay still apply, of course. It is safest to assume that there is a large error in the measures of effectiveness which are used. And a single exercise usually provides but a single data point of warfare effectiveness; extrapolation from a single such point is very risky.

Exercises are a very good vehicle for identifying any procedural difficulties which attend tactical execution. The exercise and analysis also provide a fertile opportunity to rethink the rationale underlying a tactic. More definitive evidence can be developed on ill-conceived tactics if the tactic was executed correctly and employed appropriately. The exercise and analysis also present an opportunity to observe the performance of the people and the systems. Examination may uncover areas where more training is needed, where operating procedures are not well understood, or where explicit operating and coordination procedures are absent.

Sweeping conclusions and strong, definitive judgments of capabilities, tactical effectiveness, and scenario advantages should be warning flags to exercise report

Thompson: Did We Learn Anything From That Exercise? Could We?

Did We Learn Anything? 37

readers. The reader should reassure himself that the exercise scenario, the exercise goal, and the tactical context are amenable to drawing such conclusions. For example, battle group tactical proficiency cannot be easily investigated in small, controlled exercises. Nor do capabilities demonstrated in easy battle problems imply like capabilities in harder, more realistic battle problems. The message is to read exercise reports with caution, continuously testing whether it makes sense that such results and conclusions *could* be learned from the exercise.

Mr. Thompson is OEG Representative on the staff of the Commander, Sixth Fleet.