Reflections on scenario-based training in tactical command [1]

Karel van den Bosch
Johan B.J. Riemersma

TNO Human Factors
PO Box 23; 3769 ZG Soesterberg
The Netherlands

vandenbosch@tm.tno.nl
riemersma@tm.tno.nl

---

[1] Bosch, K. van den, & Riemersma, J. B. J. (2004). Reflections on scenario-based training in tactical command. In: Samuel G. Schiflett, Linda R. Elliott, Eduardo Salas, & Mike D. Coovert (Eds.), *Scaled Worlds: Development, Validation and Applications*. (pp1-21). Aldershot: Ashgate.

Contents

1 Introduction

The rapid technological developments in weapon, sensor, and communication systems have changed the character of military command substantially. The employment of advanced equipment has opened new possibilities and opportunities, but has also made task performance more complicated and often less transparent. Commanders need to select and process what is relevant in vast quantities of information in order to assess the true nature of complex situations, and to make decisions under time pressure that may have major tactical or political consequences.

All modern armies recognize the prime importance of effective command and control (Levis & Athans, 1988, Van Trees, 1989). Traditionally, the approach for training military command and control was based upon Rational-Analytic (RA) theories of human decision making (Kahneman & Tversky, 1982). Research involved asking naïve subjects to solve complex, but artificial, problems in a laboratory setting. In many studies, it is demonstrated that people’s situation assessments and decisions often deviate strongly from the mathematical optimum. A wide variety of tendencies of people to select or neglect information, to interpret information wrongly, and to make incorrect argumentation is identified in long lists of biases. The main message of RA theories is, accordingly, that people are poor decision makers (e.g. Kahneman & Tversky, 1982; Thaler, 1980). The military responded by developing analytically correct decision making procedures, designed to minimize the likelihood of biases influencing decisions. This approach is implemented in, for instance, the OODA-model (Observe, Orient, Decide, Act) of military command. Personnel is trained to use general techniques for structuring facts, generating hypotheses, to develop different relevant Courses Of Action (COAs) for the tactical problem, to define evaluation criteria, to evaluate the options against these criteria, to calculate the outcomes, and to select and execute the best option.

Although the rational-analytic approach influenced training tactical command and control, it was never completely adopted by the military (Van Creveld, 1985). The main reason is perhaps that command style is seen as the privilege of the commander, which is considered to be crucial for success (Gray, 1989-1994). But there are also practical obstacles in the application of RA-based methods. One important obstacle is that in real world problems, the analysis of a situation is always influenced by the goal for which the analysis is performed and upon the interpretation of the context, hence context-free analysis into operational units is an illusion. Furthermore, assessing a tactical situation usually requires considering a large number of variables (e.g. strength of own forces, estimated force and intentions of enemy, weather- and terrain conditions, etc.). The actual status of many of these variables is often not known or uncertain. Working out the implications of possible values of all variables relevant to the situation would produce an unworkable calculation (see Klein, 1998 for a more detailed discussion).

RA decision making research thus had limited use for organizing the training of command and control and made way for a new approach in the study of situation assessment and decision making in real-world problems: "Natural Decision Making" (NDM) (e.g. Klein, 1997; Lipshitz, 1993). Many NDM studies have shown that domain-experienced decision makers do not analyze a situation into its constituent components, nor do they generate series of different courses of action. Instead, they use their ample domain knowledge to recognize a situation as familiar to one earlier experienced and to retrieve an appropriate response (see discussion of "Recognition primed decision"; Klein, Calderwood & Clinton-Cirocco, 1986). In unfamiliar situations, domain experts treat decision making as a problem-solving process in which recognized elements are used to construct an interpretation of the situation. The plausibility of this interpretation is verified and improved where necessary (Cohen & Freeman, 1997). The approach to research and training tactical decision making advocated by NDM theories is therefore to carry out training in a contextually rich environment.

This changing view on studying tactical decision making resulted in an increased interest in field studies. This yielded significant new insights in the behavior of experts and produced many leads for research (e.g. Cannon-Bowers & Salas, 1998). However, an important limitation of conducting studies in the field is the lack of experimental power or control (Hess, MacMillan, Elliott, & Schiflett, 1999). Due to recent advances in hardware, software and network technologies, it has become more and more possible to develop Synthetic Training Environments (STEs), capturing the relevant variables of tactical command situations in realistic simulations. This technology enables the study of training tactical decision making using back-to-back experimentation (Sanders, 1991; Hess et al., 1999) where findings obtained in laboratory and/or field studies provide input for investigations using a controlled, yet operationally relevant, task environment. In addition, effects of training interventions obtained in controlled simulations can be validated in field studies.

With the growing popularity of using STE for research, and the ever-increasing capabilities to create very realistic high-fidelity simulations of tactical command situations, the discussion on requirements and specifications for reliable and operationally valid experimentation becomes more important. The investigation of training interventions in an adequate and controllable fashion requires detailed specification of the experimental training program (e.g. design and delivery of scenarios under full control of the experimenter; selection of performance measurements suitable for answering the experimental question; maintaining a standard procedure for performance feedback).

Studies and discussions on how to design, deliver, and evaluate training in synthetic environments have yielded rich information on scenario-based training in tactical command. However, it is important to acknowledge the differences in demands between STE-based training used for educational research purposes, and training used for operational purposes. For research purposes, a training program should be designed according to a prespecified structure, so that the experimenter has full control over the amount and nature of training that the trainee(s) receive. The design is generally focused on obtaining data that will provide clearly interpretable answers to the questions of interest, not to bring trainees or teams to an over-all level of task performance. Provision of training in a research context also requires strict control, so that effects of the experimental manipulations can not be confounded with extraneous factors. Finally, the effects of training are evaluated in the context of the scientific hypotheses under study.

In contrast, the objective of STE-based training for operational tasks has different demands. Instructors should be able to adapt on-line to the training needs of individuals and teams, thus requiring a more flexible design of training. Training is provided to satisfy the actual training needs effectively and efficiently, possibly requiring deviations from the prespecified training trajectory. Performance is measured to monitor the learning progress and to decide on further training, to diagnose any misconceptions or gaps in the knowledge or skills of trainees, and to use this information for providing feedback. Effects of training are evaluated on a higher level of aggregation, namely whether or not the training objectives have been achieved (internal validity) and whether effects of training generalize to the operational environment (external validity).

Studies using STEs to investigate effects of training tactical command and control skills, and the accompanying discussion about requirements for using STEs as valid research instruments, have together produced a variety of insights in scenario-based training. The present chapter gives an overview of the implications of these insights for developing scenario-based training for operational purposes, along with the identification of its major challenges.

2 Scenario-based training

Tactical commanders are, either individually or as member of a team, engaged in cognitive processes, such as: identifying and interpreting cues; collecting and selecting information, making inferences and judgments; developing, implementing, and communicating plans; and monitoring and evaluating processes. Successful task performance requires all these competencies (Oser, 1999). Once trainees have a basic understanding of concepts and principles underlying the task, they need to acquire practical skills under guided supervision in series of systematic, controlled and representative exercises. The traditional framework for instruction and training, which is based on first learning theoretical knowledge followed by practical training in field exercises, has practical and didactic drawbacks. An example of a didactic drawback is that in field exercises, the instructor has limited control over the learning environment. The selection of terrain and routes for training are limited to what is available in the exercise area. It is therefore difficult, or sometimes even impossible to structure the learning situations according to an optimal sequence. A practical drawback are the vast organizational and logistic efforts required to conduct a tactical field exercise.

In contrast, another form of training, called scenario-based training, trainees prepare, execute, and evaluate exercises that are simplified simulations of the real-world. Scenario-based training is considered to be a more appropriate approach for training competencies required in complex task environments (e.g. Fowlkes, Dwyer, Oser, & Salas, 1998, Oser, 1999). The exercises should be selected and designed in such a fashion that they systemically call upon all the knowledge and skills formulated in the training objectives (Cannon-Bowers, Burns, Salas, & Pruitt, 1998). It focuses, more than other forms of training, on the development of practice, diagnosis and feedback. Figure 1 shows a schematical overview of scenario-based training. Scenario-based training is more and more used in practical training programs. There exist, however, a number of fallacies, problems, and challenges in the practical application. These are discussed below.

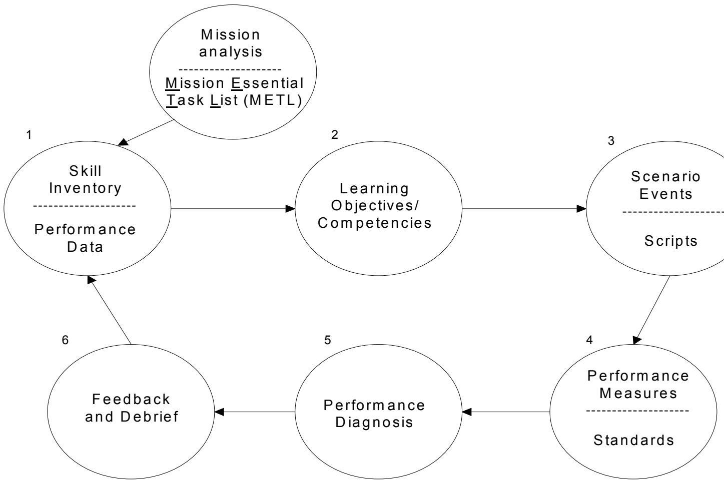

Figure 1: scenario-based training (adapted from Zachary, Bilazarian, Burns, & Cannon-Bowers, 1997)

2.1 Analysis

If we want to train novices to become experts, we have to have an adequate understanding of the content and structure of the knowledge and skills underlying effective task performance. This can be obtained by analyzing the missions and tasks, and by examining expert behavior. From learning theories and also from Naturalistic Decision Making studies we know that experts do not always have access to the processes underlying their behavior. Typical expert behavior in open cognitive tasks is more characterized by pattern recognition than by deliberate and conscious reasoning. On the road to acquiring expertise, experts have been exposed to many different situations. When confronted with a particular situation they often immediately recognize a (partial) match with one or more situations stored in memory, and immediately identify the proper course of action. This usually takes place without considering all the relevant properties of the problem. If experts are asked to explain to novices what they are doing, they tend to stop and artificially break the task down into procedural subtasks. Often they feel uncomfortable, because this segmentation does not really correspond to how they experience the task itself. Experts generally do not feel that they are performing subtasks because the integration is so strong.

The problem is then: how can we know what someone else knows? Cognitive Task Analysis (CTA) refers to a series of methods (e.g. structured interviews, interviews about the concepts experts use to think about a task, experts thinking aloud while performing the (simulated) task) for uncovering experts' relevant but tacit knowledge and knowledge structures (e.g. Hoffman, Crandall, & Shadbolt, 1998; O' Hare, Wiggins, Williams, & Wong, 1998; Schraagen, Chipman, & Shalin, 2000; Zachary, Ryder, & Hicinbothom, 1998). The result is a description of the content and structure of the knowledge and skills underlying effective task performance. CTA is generally followed by a training analysis to identify and formulate the learning objectives to be addressed by the training program (Farmer, van Rooij, Riemersma, Jorna, & Moraal, 1999).

2.1.1 Problems and challenges in analysis

Performing CTA is a complex, elaborative and time-costly activity. That explains why a thorough analysis is often omitted in the specification of (scenario-based) training programs (Militello & Hutton, 1998). The selection of method(s) for CTA is therefore not trivial. Some CTA methods are natural and non-directive (like observation of actual performance), but these tend to be time consuming and inefficient. There are more efficient techniques (like structured interviews and probed questioning), but these run the risk of being artificial and directive, and are therefore more prone to bias.

Attempts have been taken to make CTA more accessible to practitioners by streamlining the procedures (e.g. Schraagen et al., 2000). Although it is possible to reduce the load on available resources, a trade-off is generally acknowledged: the more streamlined and proceduralized CTA techniques become, the less powerful they are (Millitello & Hutton, 1998). A complicating aspect is that the approach to perform the task often differs substantially among SMEs. The content and structure of knowledge of an expert is determined by many factors: instruction, experience, and the situations experienced, may differ among experts-to-be. Consequently, not all experts therefore rely upon the same knowledge base and do not use the same strategies. To really understand all aspects of the task, it is therefore necessary to include several SMEs in the task analysis.

2.2 Event specification and scenario design

A central, and perhaps the most difficult, part of scenario-based training is specifying or selecting events and uniting them into a training scenario that enables students to achieve the learning objectives effectively and efficiently. An event can be defined as a critical moment in task execution, marked by (the absence of) a significant cue, requiring a response by the trainee as formulated in the learning goals. For example, the popping up of a target on a radar screen or the absence of an expected message can be events. The interpretation of events requires a context, including components such as environmental setting, the mission, enemy intent and capacities, logistic and organizational constraints, and many other types of information. The embedded sequence of related events within a context is called a scenario.

Scenarios must be constructed in such a fashion that trainees expand their domain knowledge, learn to recognize and assess prototypical situations, and practise the application of domain knowledge to novel situations. This requires from training developers to resolve many questions like: “how many learning objectives should I address in a training scenario?”; “how do I specify or select events that call upon the skills of interest?”; “how do I ensure that intended events actually take place in the training scenario?”; “how do I accomplish the connection between learning goal, event, performance measure, and feedback?”.

2.2.1 Problems and challenges in event specification and scenario design

There exist a number of methods supporting the process of event specification and scenario generation (e.g. Campbell & Deter, 1997; Fowlkes et al., 1998; Prince, Oser, & Salas, 1993; Oser, Gulateri, & Dwyer, 1998; Stretton & Johnston, 1997). All rightly emphasize the relationship between the learning goal, scenario event and performance measure. However, these methods unfortunately do not sufficiently support the route from the generally formulated learning goal to concrete and specific event(s). For example, a learning objective might be: “the trainee recognizes the risk that relying too much on pre-established expectations interferes with a proper situation assessment and tries to prevent such bias by actively searching for ambiguous, inconsistent, and lacking information in the problem space”. Training this objective requires the specification of events eliciting the target skills (e.g. gathering and verifying information; comparing events with previous experiences, testing assumptions), and the formulation of procedures for evaluating whether or not the trainee’s behavior is adequate. The available methods do not specify how instructional slots of scenario-based training should be filled with domain specific information in such a manner that the learning objectives are achieved.

Current methods for scenario design are all in contrast with the “train as you fight” philosophy, which is still the prevailing military approach to training. The latter pertains to the idea that to warrant transfer, the nature and content of training should correspond as much as possible to the operational environment. The consequence of that principle is that training sessions are planned and executed like operational missions. However, the objectives differ. The goal of operational missions is to minimize failure, while the goal of training is to learn and improve knowledge and skills (Farmer et al., 1999). Another drawback of training in the context of operational missions is that the training and learning process becomes intractable. Operational environments are typically complex, dynamic and unpredictable which conflicts with the fundamental requirement of training and measurement of having control over task content so that it is known which events will occur at a specified point in the scenario (Fowlkes et al., 1998).

Thus, the “train as you fight” philosophy is in conflict with the didactic principle of control. Control is necessary to make sure that the content and complexity of scenarios is in accordance with the phase of training. Scenarios at the beginning of training need to be simple and straightforward. Their function is to enable the effective acquisition of knowledge about relationships between the context, the environment, actions, and outcomes. Insight in the causal relationships between processes and outcomes is essential for effective training. This can only be achieved if the processes deliberately elicited by the events in the scenario, produce positive and clearly predictable outcomes. When trainees progress, scenarios may become more complex, allowing for multiple solutions. Thus, the nature and complexity of the scenarios change as trainees progress. Using operational scenarios from the outset of training is not compatible with the principles of controlled increase of complexity level, making it hard for the trainee to learn the relationship between processes and outcomes.

2.3 Performance measurement

The next step in the cycle is the specification of performance measures. Effective performance measures not only allow assessment of whether trainees achieve the learning objectives, but also they help to determine why performance occurred as observed and to diagnose any knowledge gaps or misconceptions. Two main types of performance measures can be distinguished: outcome and process measures. Outcome measures assess the quantity and quality of the end result (what is actually achieved?); process measures describe the strategies, steps or procedures used to accomplish a task (Sanders, 1991; Smith-Jentsch, Johnston, & Payne, 1998). Outcome measures are particularly appropriate when there is a close relationship between stimuli and response (causally, and in time), like for instance, steering a car. A steering error immediately results in a deviation of the car's optimal course. Thus, measures referring to course deviations (e.g. RMS of lateral position) are appropriate for evaluating steering competence. In tactical command and control tasks on the other hand, there is not such a clear relationship between stimuli and response. The skills of interest are usually not directly reflected in observable actions (Vreuls & Obermayer, 1985). Task performance relies for a considerable part on internal mental processes, such as reasoning, memory retrieval, knowledge integration. For example, a commander may be engaged in assessing the consequences of an anticipated change in the tactical situation, without showing any observable actions related to this process. Outcome measures are therefore not very appropriate for assessing the quality of the student's task performance. Instead, measures are needed that reveal the student's internal processes and strategies.

Another complicating characteristic of tactical command and control, is that there is often no single "right" way to accomplish a task (Hutchins, Kemple, Porter, & Sovereign, 1999). There usually exist more than one good solution for a problem, depending upon the assumptions and inferences you are willing to accept. This makes it hard to assess the quality of trainee performance. An action may be adequate from one perspective, but highly inappropriate from another. For example, a commander's mission may be to occupy a section of the enemy's territory, but confronted with severe opposition, he decides to temporarily draw back to prevent casualties. This decision is not in accordance with the assignment, but may well reflect sound judgment. To make the issue even more complex: task achievement in real life is not solely determined by human performance, but is also affected by equipment (mal)functioning, surrounding environment, and luck (Smith-Jentsch et al., 1998). Clearly, assessing tactical command performance requires that the situational context is taken into account. Finally, tactical command is often performed by teams rather than by single individuals. Team task success requires team skills like communication, coordination, providing backup, and the like. To assess the quality of the team, measures are needed that reveal the critical team processes.

2.3.1 Problems and challenges in performance measurement

Assessing trainee performance in open cognitive tasks like command and control is not a straightforward matter. The critical skills (e.g. reasoning, memory retrieval, knowledge integration) cannot be directly inferred from observable actions. Outcome data alone (i.e. good or poor situation assessment / decisions) are therefore not the most important: it is the adopted set of assumptions and the reasoning that counts. The plausibility of assumptions and argumentation should be identified in the analysis phase. For purposes of performance measurement, the interesting step is how the results of analysis are codified, and how they are made applicable for instruction and feedback. Good assessment requires more than a detailed description of expert knowledge: it requires the representation of expert knowledge in a generic model, which can be used to validly predict the behavior of an expert performer in any concrete scenario. This representation of expert knowledge can be used as norm to evaluate trainee behavior, and to use the result of the comparison as input for feedback.

Unfortunately, such complete generic expert models are seldom available for actual training. In the absence of a rational system of measures, SMEs are usually interviewed to identify task-relevant processes and to act as observers to grasp and evaluate the underlying internal processes of the trainee's task behavior. However, for useful assessment, more than subject matter expertise is needed. Just asking SMEs to observe and to evaluate trainee behavior tends to produce unsystematic, unreliable, and biased assessments (Smith-Jentsch et al., 1998). In order to achieve valid, reliable, and useful assessment of trainee performance, SMEs need instruments supporting performance evaluation and training in the use of such instruments (Van Berlo & Schraagen, 2000).
In order to systematize performance assessment, SMEs often use rating scales in which the processes of interest (e.g. *communication, monitoring, feedback, initiative, leadership, situation awareness)* are specified in the form of observable behaviors that are indicative for the processes. Unfortunately, experts tend to include too many task-processes in their assessment instrument. Another form of support refers to the timing of performance evaluation. It has been stressed in the literature that just asking SMEs to observe and to evaluate trainee behavior is not a good approach, but that ratings directly linked to behavior in specific events of the scenario produce much more reliable assessments (Smith-Jentsch et al., 1998). Furthermore, SMEs need training to systematically use the performance measurement tools (e.g. rating scales).

2.4 Performance diagnosis

Performance measures are needed for determining why performance occurred as observed and to diagnose any knowledge gaps, misconceptions, or lacking skills. The difficulty is that situational context is needed to provide meaning to the trainee's actions, or his decision to refrain from action. Elaborate mental models of tactical command and control, capable of handling situational context, are usually not available. In training programs and training research, Subject Matter Experts (SMEs) are therefore often used as Observer/Trainers (OTs), because SMEs are considered to be able to make the higher order transformations needed to evaluate the appropriateness of trainee behavior. More specifically, SMEs can:

- *infer internal processes* from evaluating the covert relationship between events in the outside world, internal process, and observable behavior.
- *take the situational context into account* that is needed to interpret trainee behavior.
- *interpret subtle cues* indicating internal processes. Determining that a commander is getting prepared for possible or likely events, or when he is considering alternative solutions, is not always easy to detect. SMEs are observant to these cues.

In fact, SME has become the "all encompassing" methodology for performance appraisal and training evaluation, largely because of the absence of a rational system of measures. They can provide the subjective but knowledgeable interpretation of the trainee's action in the particular context, thus compensating for the absence of model-based solutions.

It should be emphasized that SMEs should not only be able to establish what task-relevant processes are performed by the trainee (performance assessment), but also to infer from this the mental model that the trainee must have of the situation (performance diagnosis). In case of errors, they should be able to diagnose the performance by reconstructing it in relation to preceding, or anticipated, events. For example, if an exercise demands the trainee to make a decision whether or not to send supporting forces, he must take many factors into consideration, like magnitude of threat to endangered forces, passableness of route, safety of passage for support troops to endangered forces, logistic possibilities, etc. The trainee's decision does, in itself, not reveal the underlying arguments of the decision process. It is the task of an SME to uncover the cause of an incorrect decision. Sometimes SMEs can determine the cause fairly easy, in other cases additional questions may be needed. Making a good diagnosis about an error (e.g. knowledge gap, implausible reasoning, ignoring conflicting information) is necessary for providing effective feedback.

2.4.1 Problems and challenges in performance diagnosis

Their domain knowledge makes SMEs not by definition good OTs (Observer/Tainers, or Observer/Controllers (OCs). The observation that experts often have difficulty making their assessment principles explicit, is an obstacle for training, because trainees do need explicit explanations (in the form of feedback) to understand why a particular course of action is right or wrong. To be able to do this in actual training, SMEs need training regarding how to attach clues to the assessment for providing feedback. Thus, they should not only evaluate trainee behavior, but, in case of errors, they also should *diagnose* the performance by reconstructing it in terms of preceding, or anticipated, events that resulted in that behavior. The assessment must be suitable for delivering feedback to trainees so that they themselves can make that reconstruction. However, in actual training programs, SMEs tend to focus on performance assessment, and pay too little attention to performance diagnosis.

2.5 Feedback and debrief

For scenario-based training a distinction has to be made between the phases of briefing, exercise and debriefing. In the briefing phase the instructional strategy of articulation (refreshing relevant knowledge, setting goals, explanation) and modeling (providing demonstrations, examples) are adequate. The learning goals should specify the intended level of performance, which should be demonstrated in examples. In the exercise phase forms of coaching (instruction, feedback) and scaffolding (providing reminders, hints) are suited. In the debrief phase reflection (asking questions, evaluation, feedback) is an appropriate strategy to enhance learning. All these methods rely on performance measures in some way or the other, because feedback can only be based on some kind of evaluation, preferably by measuring performance.

In the development of a novice to expert level there is ideally a shift from external control and evaluation to self-control and evaluation, reduced scaffolding and collaborative assessment techniques. The early phases of training focus the acquisition of knowledge about relationships between the context, the environment, actions, and outcomes. The trainee needs explicit feedback (usually provided by the instructor) to acquire this knowledge. In the more advanced phases of training, the emphasis is on the quality of task processes (e.g. information processing, assumption testing, reasoning). At this stage, trainees should no longer be entirely dependent upon their instructor for performance evaluation and diagnosis, they should be provided with tools for self-assessment and correction, and should be encouraged to discuss solutions with other trainees.

But even in these advanced stages of training, SMEs continue to play an important role as trainees often ignore important aspects of task behavior, simply because they lack the knowledge to appreciate their relevance. In his feedback, the SME makes his expert mental model accessible to the trainee by explaining why a particular course of action is right or wrong. This should enable trainees to improve and refine their situation understanding until this has reached expert level quality.

2.5.1 Problems and challenges in feedback and debrief

For any scenario-based training exercise, the debrief phase is the one in which most learning activities can take place. This phase is the one in which the individual as well as the unit can reflect on what happened, why it happened, what alternative courses of action were open, what the underlying strengths and weaknesses were and how strengths can be sustained and weaknesses overcome. Unfortunately, many training sessions are debriefed in a way that focuses on overall success or failure and not on the underlying strengths and deficiencies in knowledge, skills or attitude of participants. These strengths and deficiencies may not easily be diagnosed due to the complexity of the training processes and outcomes or the opaqueness of the underlying processes. The problem with too much reliance on outcome however, is that the overall success or failure of the trained mission or teamtask may be quite unreliable as an indicator for quality of performance, deficiencies or training progress. As an example, consider a bomber mission with its complex sequence of phases and overall and phase specific tasks. When such a mission is evaluated solely on the missile impact accuracy, this may well provide an irrelevant criterion for most phases and tasks. Furthermore, such an outcome may be highly dependent on chance factors as changing environmental influences or the inherent variability in aiming accuracy outcome.

Too much reliance on a probabilistic overall outcome can institute even a level of unwarranted uncertainty about followed procedures and behavior which actually may have been up to standards. Chance failure then can lead to erratic trial-and-error behavior, which gives negative training outcomes.

In the early 80's After Action Review (AAR) was introduced in the US Army as a review and not as a critique. The best AARs are those which lead to self-correcting of individuals as well as units. For this to happen, a clear and unambiguous, valid overview of the events, what actually happened during the training exercise, has to be made explicit. It serves as a basis for finding out what happened as intended, and what deviations of the intended course of action were identified. From these, key issues can be selected for further discussion and elaboration, focusing on underlying causes or weaknesses, always relating these to the training objectives and doctrine. An AAR is not complete when an elaboration of solutions for maintaining strengths and reducing weaknesses has been omitted. Good AAR thus facilitates exploration of issues and the development of plans for self-correction. Blankmeyer & Blakely (1998) provide the following scheme:

Table 1: Structure for After Action Review (from: Blankmeyer & Blakely, 1998)

AAR sequence
Introduction and rules (briefly)
Review of Objectives and intents
• Training objectives; what was the objective of the training session?
• Commander's Mission/Intent (what was supposed to happen); what was the intended course of action?
• Opposing forces (OPFOR) Commander's Mission/Intent; how well was taken into account the intent of the opposing force?
• Relevant Doctrine/Tactics, Techniques and Procedures
Summary of recent events (what happened)
Discussion of Key Issues
• Chronological Order of Events
• Battlefield Operating Systems (BOS)
• Key Events/Themes/Issues

Discussion of Optional Issues

- Soldier/Leader Skills; what other courses of action could have been followed?
- Tasks to Sustain/Improve
- Statistics; number of hits; percentages,
- Others

Discussion of Force Protection (Safety)

- Risks involved in the course of action
- Measures to reduce risks.

Closing Comments (Summary)

3 War gaming

The increasing capabilities of simulation and networking technology makes is possible to provide complete virtual operational environments with scenarios covering the range of military operations to support distributed single service, joint, and combined force task based training. It is believed that these platforms, like e.g. WARSIM 2000 (Lacey & O'Brien, 1997), will positively affect the effectiveness of commander and staff training because of its increased realism and scope. The idea is that because the task of commanding forces is multidimensional, the training should also be multidimensional by practicing skill in high fidelity simulations including many entities (war games). In such simulations, the course of events is never predetermined, but evolves as a direct result of interactions among the simulated and real participants to the simulation (Lane & Alluisi, 1992). Fowlkes et al. (1998) argue that the factors making wargames exciting (complex, dynamic and unpredictable) make it troublesome from a training standpoint. Control of task content is a fundamental requirement of training so that what is measured is being known. This requirement is clearly at odds with the intentionally dynamic nature of war fighting that is simulated in war games.

War game's feature of unpredictability is not only impeding control on training, it is also unwanted from an instructional point of view. Often a large amount of effort and money is spent on modeling the hit probability and the precise effects of weapon hits to calculate the effects of a tactical decision on the course of events as realistically as possible, and to feed the results of the calculation back into the simulation. The same principle is used in many management games. In our view a certain instructional paradox is connected to these efforts. Let us be more explicit. For reasons of clarity we give an example on the individual skill level in gunnery tasks, but the same reasoning applies to staff decisions in a wargame environment. In a gunnery simulator, the result of a gunner firing his weapon is calculated using several factors: aim point, the distance to the target, characteristics of used ammunition and target, and -empirically sound- probabilistic variation. The result can be a hit or no hit. If it is a hit, the result can be mobility kill or total kill. The simulation supplies the result as feedback to the gunner. The instructional paradox is that the outcome feedback (although very much conform reality) cannot be interpreted by the gunner in a way that improves his skills. Training can even have detrimental effects if the simulator assigns a non-hit, due to the influence of probabilistic variation, to accurate performance, and the gunner decides to change his task behavior. So the paradox is that although the simulation provides highly realistic outcomes to task performance, by including a probabilistic variation component in the calculation, the learner requires many, many engagements to build up knowledge for himself about the relationship between task performance and result. So, instead of actually applying probability distributions, a more sound instructive way would be to use a deterministic decision rule for determining the probability of success: all shots with a hit-rate probability above 50 % are real hits, all those below are misses. Feedback then addresses the underlying model and makes it more explicit. Only when trainees incorporate this knowledge and adequate skills, probabilistic misses can be introduced as a cue or trigger for replanning actions, e.g. a second engagement.

A positive feature of wargaming is that it may enhance motivation by inducing challenge, curiosity and fantasy. This may, in effect, foster the retention of learned knowledge and skills. However, training in games may also bring about unwanted effects. The risk inherent in any gaming situation is a shift in emphasis from learning to winning. Whether trainees will perceive winning, rather than learning, as the primary objective of the games depends upon the way it is designed and whether or not pre-instructional information is able to induce the desired attitude towards the game. Lundy (1985) describes several learner orientations with respect to management games. She classified them as:

- Opportunistic, competitive; trainees see the game only as a game, an opportunity to prove themselves. Main goal is to win and no means to this end are being eschewed, including gambling
- Knowledge based; trainees try to develop skills in applying the earlier learned theoretical knowledge. They use a genuine strategy, which is sound reasoning based on acquired knowledge. Even this group with the intended orientation may fall victim to the restricted time horizon of most games and to gambling when they do not succeed
- Specific skills; trainees try to use the gaming situation for their own objectives, for instance acquiring skill in negotiating
- Exclusion; trainees (15-20 %) disconnect physically or mentally because they ‘don’t like games’, have trouble in communicating properly, or fail to collaborate
- Confusion; trainees cannot grasp the rules, do almost nothing and report little learning and much confusion

The closed design of many games and the usually complex and even opaque way of determining the outcome of certain actions often lead to the development of game-specific skills, which have no transfer to the reference (simulated) real environment. A game-specific skill may be, for instance, the ability to grasp the way the outcome is determined by the implemented algorithm and to capitalize on that knowledge, or to make use of inherent weaknesses in the game design. Most devilish is the use of game construction knowledge in making the system crash in case one has no longer the possibility to win, thereby preventing opponents of reaching deserved victory.

4 Team level training

Nowadays, complex tasks can no longer be performed by subdividing them in part-tasks, allocating these to several specialists and integrating the part-task results afterwards into an overall result. Due to technological and organizational changes, much more emphasis has to be placed on the ways teams of experts operate in concert towards a common goal and to the crucial team processes of communication and co-ordination.

4.1 Fallacies

Training a team like a military unit, is an endeavor surrounded with much vagueness and even mythical connotations. One myth is the idea of a strong leader as a sufficient condition for optimal team performance. This is one of the most persevering ideas about how to train military units, since in these types of training the commander is usually the focal point of evaluation, not only with respect to his operational role but also as the one responsible for training outcomes.

Another fallacy is the idea that team training is best performed in environments and with scenarios closely resembling the work- or operational environment. For the military this is the 'train as you fight' doctrine. The roots of this fallacy are clearly legitimate: contextual similarity does enhance transfer to the work situation, and as long as there is not a more developed theory of transfer of training, it seems a well-founded idea, certainly as a reaction to earlier forms of military exercises. A drawback, however, is that no distinction is made with respect to the already acquired ' level of expertise' and further, that in such situations the transparency of the learning situation is too low for the trainee to gain useful learning and the ongoing behaviors are too intractable for the training staff for giving effective feedback. A further drawback is the lack of focus on well specified training goals, which address identified weaknesses. Effectively, this may lead to a form of trial-and-error learning, known to be one of the most inefficient types of learning. The also widely held belief that just more practice automatically leads to better skills is not valid, since with just more practise, no real learning focus or intent has been given to the group of trainees. Essential is thus a good idea about what a team training program has to accomplish beyond the enhancement of knowledge about a certain domain or how to operate in certain circumstances.

Thus, for team training, training goals have to be explicated on the level of the team.

4.2 Strategies

As a prerequisite condition for team training one should demand proficiency at individual levels of competencies. The real aim of team training is to learn to work as a collective of experts with a shared goal, but with different relevant expertise. Central to such a collective effort are effective communication and co-ordination of behavior. This is only possible when the team members not only share goals, but also have a harmonized idea about the situation, the means the team can bring in, the way tasks are interrelated in the overall team task, the weaknesses and strengths of other team members and a calibrated way of (also non-verbal) communication. How to reach these conditions? There is no standard procedure invented yet. One has to think about carefully constructed scenarios, in which the need for developing complementary but shared mental models becomes obvious and after which the AAR are focused on further developing the shared mental models on which good team performance depends.

Two key design parameters in scenario construction are, first, the linking of scenario elements to well defined training goals, and, second, the manipulation of interdependencies between team members, such that a build-up of shared mental models is facilitated. Not operational fidelity but instructional efficiency should be the leading principle in scenario construction. For example, when the learning goal is the development of skills in monitoring the behavior of other team members and assisting them in their task performance, events have to be chosen such that one team member is overloaded, while another one has enough spare capacity to intervene. Only then a lack of skill can be identified and corrected. In the same vein, when a learning goal is to foster a (shared) mental model of the team task, it is less efficient to start with scenarios in which all team members are dependent on each other. It will be more transparent for the trainees to start with scenarios with only pair-wise dependencies.

Also for building up skills in situation assessment, one has to start with situations with a large amount of constancy and only variations in one or two (un-correlated) dynamic variables. Otherwise it is almost impossible for trainees to develop an adequate model of the effective variables to consider in situation assessment.

4.3 Team performance measures and feedback

Depending on the instructional strategy chosen for a training session, different forms of feedback are necessary. However, feedback can only be given on the basis of adequate and reliable performance measurements or assessments in terms of outcomes as well as processes. Furthermore, feedback will be more effective when a proper diagnosis of lacking (team)knowledge or (team)skills or misconceptions can be made.

For scenario-based training a distinction has to be made between the phases of briefing, exercise and debriefing. In the briefing phase the instructional strategy of articulation (refreshing relevant knowledge, setting goals, explanation) and modeling (providing demonstrations, examples) are adequate. In the exercise phase forms of coaching (instruction, feedback) and scaffolding (providing reminders, hints, team work facilitator) are adequate. In the debriefing phase, one should use reflection (e.g., asking questions, evaluation, feedback, guided team self-correction) as the strategy to enhance learning. All these methods rely on performance measures in some way or the other. Setting goals demands a certain intended level of competence in terms of content, speed or accuracy. In the demonstrations and examples a level of competence has to be demonstrated, which is based on adequate performance levels. Also feedback can only be based on some kind of evaluation, preferably by measuring and diagnosing performance. In the development of a team to an expert level, there should ideally be a shift from external control and evaluation to team self-control and self-evaluation, fading scaffolding and improved goalsetting by the team itself. Even then the need for performance measures is paramount since during actual team work much can go unnoticed or cannot be precisely assessed. The form of representing feedback should reflect the already acquired level of competence and be adapted to the use by the team itself. Thus, performance measures are needed to monitor training progress and diagnose weaknesses in task as well as teamwork skills of an actual training team.

4.4 Use of simulated team members in team training

Team training often faces a number of organizational problems, which limit the frequency of training sessions with the complete team. These are:

- budget limitations; the cost of providing a training environment (training staff, role-players, observers) for training a team are tremendous.
- agenda problems; the necessity of having all team members at the same place at the same time often severely limits the number of opportunities for training the whole team.
- availability of the team environment for training; training disrupts the normal use of the resources needed.

There are also inherent drawbacks of always having to train the complete team or using other trainees for role-playing positions in the team. Although in the latter case a form of collaborative learning is stimulated, it may not be effective in the long run. Time spent in role playing is not spent in training the own task. Furthermore, trainees may not be proficient enough to really perform well as role players in positions, in which they have little or no experience (Schaafstal *et al.,* 2000). This may result in biased team work models.

Once the choice is made to use a networked simulation based training environment, the incorporation of intelligent agents into such an environment to overcome the difficulties of assembling all team members, suggests itself. They can play the part of missing members of the team but they can furthermore also play the part of external resources, such as higher and lower control and even part of the training staff. Thus they may enable training when no training would be possible without them.

Simulated teammates are thus a promising alternative to human teammates, because they are always available, may be modeled after experienced training personnel, and may be more cost effective. Furthermore, the use of simulated team members can actually enhance the quality of training.
For event-based training there has to be a consistent link between events in the training scenario to training goals and assessment and diagnosis of behavior. In such an approach training needs, critical tasks, learning objectives, scenario design and performance measurement, diagnosis and feedback have to be systematically and methodically linked in a tight manner. Since at the team level, practice in teamwork skills calls for 'events' originating in the behavior of other team members, the control thereof is easier when simulated team members are used: they can be programmed to represent a certain level of competence, to make errors, thereby eliciting (or not) corrective behavior of the trainee, to ask questions, e.g. about priorities et cetera. In a very consistent, and learning goal related way a simulated team member can provide all the clues needed for the specific practice in the teamwork skills the session is focused on.

Since the systematic and controlled approach to train well specified teamwork skills also enables much more focused performance measurement and diagnosis, it is suggested (Schaafstal et al, 2000) that a simulated team member can be enhanced and can serve as an aid in diagnosing the team's or the team member's performance. It takes only one step further to also provide intelligent feedback or to prepare focused After Action Reviews on the basis of this on-line diagnostic function.

5 Conclusion

Competency in tactical command is not easily acquired nor assessed. It comprises: adequate *assessment* of tactical situations and enemy intent, the *planning* of initiatives and responses to fulfill the mission with available resources, the *command and control* during the execution of the intended actions, and the *re-planning* given contingencies. Although tactical command is generally considered a core competency of armed forces, the military seldom allocates the attention, or the resources, needed to train its personnel to a sufficient level of skill. In this paper we have reflected upon some issues involved in developing and implementing effective training in this area.

The "train-as-you-fight" philosophy is very popular in military training institutions. One implication is the use of complex operational scenarios for training purposes. We have argued that, from the viewpoint of the learner, operational scenarios are much too complex on the input side, which results in a lack of focus on the part of trainees. Furthermore, it yields an almost impossible task for observer/trainers to select and fruitfully feedback their observations.

Instead of operational scenarios, more controlled and systematic series of training scenarios which are specifically designed to train one or more components of tactical command are needed for effective training. Instructors should be able to tailor the content and structure of the series to the training needs of the individual trainee or unit.

A comprehensive representation of expert knowledge, including all phases of the command cycle, is needed for designing scenarios providing the events eliciting the target behaviors, and the standards for evaluating trainee performance. Effectively using such scenarios for training purposes demands knowledge about optimal training strategies. Unfortunately, our current understanding of these two elements is far from complete. More analytic research is needed, using adequate and controlled research designs, focusing on developing theories about what comprises competency in tactical command and about effective ways to develop such competencies. Based on such empirically validated theories, more effective training can be designed and implemented, using the paradigm of scenario-based training in Synthetic Training Environments.

We have argued that training scenarios have to be built by specifying focused events. These should be coupled tightly to learning goals, performance measures and performance diagnosis, and feedback. Due to the complexity of tactical command, the relations may easily remain opaque to the trainee. To enhance transparency, training scenarios should initially use events addressing a single certain specified skill, and by augmenting outcome measures and feedback to compensate for the inherent noise. As learning progresses, the operational complexity of scenarios may be increased.

Tactical command is often performed in teams. Team tasks require the utilization of cues that do not originate in the external environment (e.g. terrain, enemy or neutral forces), but in the internal environment of the own organization and team. For learning the relevance of these cues, and learning to respond appropriately, teams consisting of trainees are not very effective environments. The use of simulated team members may provide a more controlled and transparent learning situation. Simulated team members (agents) may even evolve to diagnostic agents in their own right.

References

Berlo, M. P. W. van, & Schraagen, J. M. C. (2000). A generic assessment tool for evaluating C2 exercises. In: *Proceedings of the 22th Interservice/Industry Training Systems and Education Conference (I/ITSEC)*. Held at: Orlando, Florida.

Blankmeyer, W., & Blakely, T. (1998). Leaders conducting after-action reviews (AAR) often deliver substandard feedback. *ARMOR, November-December*, 15-18.

Campbell, C. H., & Deter, D. E. (1997). *Guide to development of structured simulation-based training.* HumRRO.

Cannon-Bowers, J. A., Burns, J. J., Salas, E., & Pruitt, J. S. (1998). Advanced technology in scenario-based training. In: J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: implications for individual and team training* (pp. 365-374). Washington, DC: APA.

Cannon-Bowers, J. A., Salas, E., & Johnston, J. H. (1998). Lessons learned from conducting the TADMUS pogram: balancing science, practice, and more. In: J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: implications for individual and team training* (pp. 409-413). Washington, DC: APA.

Cohen, M. S., & Freeman, J. T. (1997). Improving citical thinking. In: R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 161-169). Brookfield, Vermont: Ashgate.

Creveld, M. van. (1985). *Command in war*. Cambridge, MA: Harvard University Press.

Farmer, E., Rooij, J. C. G. M. van, Riemersma, J. B. J., Jorna, P., & Moraal, J. (1999). *Handbook of simulator-based training*. Hamspshire, UK: Ashgate.

Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *The International Journal of Aviation Psychology, 8*(3), 209-222.

Gray, A. M. United States Marine Corps. (1989-1994). *Warfighting: tactics for managing confrontation*. NewYork, NY: Doubleday.

Hess, S. M., MacMillan, J., Elliott, L. R., & Schiflett, S. (1999). Team-in-the-loop synthetic simulation: bridging the gap between laboratory and field research. *43nd Annual meeting of the Human Factors and Ergonomics society*. Held at: Houston, TX.

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of critical decision method to elicit expert knowledge: a case study in the methodology of cognitive task analysis. *Human Factors, 40*(2), 254-276.

Hutchins, Susan G., Kemple, William G., Porter, Gary R., & Sovereign, Michael G. (1999). Evaluating Human Performance in Command and Control Environments. In: *Command and Control Research & Technology symposium*. Held at: Newport, RI.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11*, 123-141.

Klein, G. (1997). The current status of the naturalistic decision making framework. In: R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 11-28). Brookfield, Vermont: Ashgate.

Klein, G. (1998). *The source of power: how people make decisions*. Cambridge, MA: MIT Press.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. In: *Proceedings of the 30th Annual Human Factors Society Conference*(pp. 576-580). Dayton, OH: Human Factors Society.

Lacey, L. W., & O'Brien, L. (1997). Conceptual models for WARSIM 2000. In: *Proceedings of the 19th Interservice/Industry Training, Simulation, and Education Conference*(pp. 495-502). Held at: Orlando, FL, on: December 1-4, 1997.

Lane, N. E., & Alluisi, E. A. (1992). *FDidelity and validity in distributed interactive simulation: questions and answers*. (Report No. IDA Document D-1066). Alexandria, VA: Institute for Defence Analysis.

Levis, H., & Athans, M. (1989). The quest for a C3 theory: dreams and reality. In: S. Johnson, & A. Levis (Eds.), *Science of command and control: part II. Coping with complexity* (pp. 4-9). Washington, DC: AFCEA International Press.

Lipshitz, R. (1997). On-line coping with uncertainty: beyoynd the reduce, quantify and plug heuristics. In: R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: emerging themes and applications* (pp. 149-160). Brookfield, Vermont: Ashgate.

Lundy, J. (1985). The effects of competition in business games. In: M. van Ments, & K. Hearnden (Eds.), *Effective use of Games in simulation*. Sagset.

Militello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

O' Hare, D., Wiggins, M., Williams, A., & Wong, W. (1998). Cognitive task analyses for decision centred design and training. *Ergonomics, 41*(11), 1698-1718.

Oser, R. L. (1999). A structured approach for scenario-based training. *43nd Annual meeting of the Human Factors and Ergonomics society*. Held at: Houston, TX.

Oser, R. L., Gulateri, J. W., & Dwyer, D. J. (1998). Guidelines for assessing exercise design: implications for scenario-based training. In: *Proceedings of the 20th Interservice/Industry Training, Simulation, and Education Conference*(pp. 509-518). Held at: Orlando, FL.

Prince, C., Oser, R., & Salas, E. (1993). Increasing hits and reducing misses in CRM/LOS scenarios: guidelines for simulator scenario development. *International Journal of Aviation Psychology, 3*(1), 69-82.

Sanders, A. F. (1991). Simulation as a tool in the measurement of human performance. *Ergonomics, 34*(8), 995-1025.

Schaafstal, A. M., Lyons, D. M., & Stroomer, S. S. (2000). Training teams with simulated teammates. In: *Proceedings of the 22th Interservice/Industry Training Systems and Education Conference (I/ITSEC)*. Held at: Orlando, Florida, on: 28-30 November 2000.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). *CognitiveTask Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Smith-Jentsch, K. A., Johnston, J. H., & Payne, S. C. (1998). Measuring team-related expertise in complex environments. In: J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: implications for individual and team training* (pp. 61-87). Washington, DC: APA.

Stretton, M. L., & Johnston, J. H. (1997). Scenario-based training: an architecture for intelligent event selection. In: *Proceedings of the 19th Interservice/Industry Training, Simulation, and Education Conference*(pp. 108-117). Held at: Orlando, FL.

Thaler, R. H. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1*, 39-60.

Trees, H. L. van. (1989). C3 systems reserach: a decade of progress. In: S. Johnson, & A. Levis (Eds.), *Science of command and control: part II. Coping with complexity* (pp. 24-44). Washington, DC: AFCEA International Press.

Vreuls, D., & Obermayer, R. W. (1985). Human-system performance measurement in training simulators. *Human Factors, 27*(3), 241-250.

Zachary, W., Bilazarian, P., Burns, J., & Cannon-Bowers, J. A. (1997). Advanced embedded training concepts for shipboard systems. In: *Proceedings of the 19th Interservice/Industry Training, Simulation, and Education Conference*(pp. 670-679). Held at: Orlando, FL.

Zachary, W. W., Ryder, J. M., & Hicinbothom, J. H. (1998). Cognitive task analysis and modeling of decision making in complex environments. In: J. A. Cannon-Bowers, & E. Salas (Eds.), *Making decisions under stress: implications for individual and team training* (pp. 315-344). Washington, DC: APA.