

# Training Experienced Teams for New Experiences

Wayne Shebilske (Wright State University)  
Kevin Gildea, Jared Freeman, & Georgiy Levchuk (Aptima)

We report the development and testing of a laboratory paradigm to study advanced training for a simulated Air Force Intelligence, Surveillance, and Reconnaissance (ISR) and Dynamic Targeting Cell (DTC) team. In Phase I (50 hr), the simulated DTC team learned background. In Phase II, they performed about 49 hours of simulated missions, which had many differences, such as different time sensitive targets, and many similarities, such as particular areas having a higher probability of specific enemy threats. In Phase III, the enemy increased their weapons and changed tactics so that difficulty increased and the pattern of differences and similarities changed. The differences, similarities, and the changes in their patterns represented those that occur in operational settings. We report the team's experience and proficiency in Phase II, their drop in proficiency at the beginning of Phase III, and their training to regain proficiency during their new experiences in Phase III.

## INTRODUCTION

Training plays an important role in experienced teams preparing for new experiences. For example, professional football teams train extensively to prepare for their next game. They watch films of their next opponent; they try to anticipate their opponent's game plan; they develop a game plan; and they practice the tactics needed for their game plan. This advanced training becomes even harder in post-season games when the best-of-the-best play one another. Similarly, experienced military teams train for new experiences related to a) their current enemy escalating or changing tactics, b) new enemies, and c) confronting their current and/or new enemies in different contexts.

After an ISR nominates an object of interest as a potential time sensitive target (TST), an unexpected threat or opportunity that demands immediate action, the main responsibilities of a DTC team are to differentiate time sensitive targets from other unexpected objects, to designate an object of interest as a TST, and to submit a plan to prosecute all designated TSTs. The plan includes the order in which the TSTs should be attacked and strike packages for each attack.

Until recently, research on such complex domains was limited to explorations of operational settings with little rigorous hypothesis testing. Now

Aptima's DDD 4.0, enables us to investigate complex DTC simulations leveraging the power of several disciplines to analyze emergent information processing systems that are invisible from any single perspective. We simulate and explore complex domains, precisely measure variables that will guide modeling of human information processing during complex activities, and rigorously test well founded hypotheses about complex processes.

Similarly, training of experienced DTC teams for new experiences usually happens on the job. Training with realistic simulations could have at least three advantages over training on the job. First, training systems would be highly accessible to promote frequent practice. Second, practice scenarios in simulations could be systematically structured and focused on training objectives, to promote deliberate practice. Third, performance assessments could be relevant to training objectives, and systematically delivered to train and maintain competencies.

A large body of literature from cognitive and instructional psychology supports this approach and provides valuable training principles (e.g. Piorolli and Anderson, 1985; Schmidt and Bjork, 1982). Much of this literature, however, concerns basic training on tasks that are well-defined (e.g. mathematics) and that are executed mainly by

individuals. In contrast, DTC functions and other command and control operations involve human performance by teams in domains that are ill-defined. Furthermore, the DTC training under consideration is not initial training of new operators, but advanced training of experienced operators learning new experiences. Can such advanced training be represented in a realistic laboratory simulation? As a step toward answering this question, we trained college students for about 100 hours and measured their proficiency while performing as a simulated DTC team. Then we simulated challenging new experiences and measured their drop in performance. Finally, we measured their learning of the new experiences for about 8 more hours.

The new experience had two parts. The first, and most extensive part, related to the enemy changing strategies to exploit vulnerabilities in current defense strategies. For example, during original training, the DTC developed the strategy of using refueling tankers to support their proposed attack packages. This strategy was helpful because the proposed attack packages utilized some assets that were low on fuel due to their execution of operations related to the ATO. A challenging change in enemy strategy might be to have the enemy destroy the refueling tankers, which would force attack assets to return to base to refuel. This requirement would force the DTC to choose different attack package patterns than those to which they had become accustomed. The second part of the new experience was to have the enemy increase their weapons. This part was introduced after the strategy change was introduced and after its effect was measured. The additional weapons added to the challenge of learning the new strategic experience.

## **METHODS**

### **Participants**

The participants were 7 undergraduate college students (2 women and 5 men, mean age = 22.6 years), who were paid \$7.25 per hour for 108 hours each. Their participation was part of their

responsibilities as research assistants. Although the trainees were research assistants, they did not know the purpose of the experiment.

### **Apparatus and Task**

Office dividers separated seven work stations with four stations in one row and three in the other. Each station had an IBM compatible PC with a 17 in. monitor, a mouse and keyboard for inputs, and a headset linked with an Aardvark sound system audio net that enabled open and recorded communication within teams and isolation of sounds outside the team.

We employed Aptima's Dynamic Distributed Decision making (DDD) platform to simulate the Dynamic Targeting Cell (DTC) task. We chose this task because of its relevancy to modern Air Force operations and to general online command and control operations. Our advanced trainees performed the role of the ISR and the four roles in a DTC cell, the DTC Chief, Ground Track Coordinator (GTC), Attack Coordinator (AC), and Target Duty Officer (TDO). Two confederates played the roles of Chief of Combat Operations (CCO) and Senior Offensive Duty Officer (SODO) who have critical roles in the larger offensive team with which the DTC interacts.

During operational missions and during the DDD simulated missions, these offensive team members usually communicate by text messaging and occasionally communicate by voice. The background training for the present trainees included standard Air Force brevity procedures for both communication modes. For example, the trainees were taught that the standard text or verbal acknowledgment to a message is "copy."

The DTC must operate in the face of uncertain information and within the constraints of an existing mission plan, the Air Tasking Order (ATO). In operational settings, the ATO changes every 8 to 12 hours. In the present experiment, it changed for each mission. The ATO includes guidance on what kinds of TSTs might be encountered, and how they should be prioritized relative to the ATO and relative to each other.

The synthetic task environment was based on a task analysis that Aptima conducted on

operational DTC teams. The Air Force structures DTC operations as a kill chain, which includes multiple stages and multiple task objectives. Our research focuses on three task objectives (TO): TO1, detect and differentiate TSTs, TO2, prioritize the TSTs, and TO3, coordinate attack assets, with emphasis of TO3, which entails the others. These task objectives relate to complex teamwork occurring in three of the kill chain stages: Find, Fix, and Target. Operations occurring during the Find stage are limited to the ISR who initiates operations related to TO1. DTC operations in the Fix phase include complex communications among the four DTC team members, GTC, Chief, AC, and TDO. This teamwork relates to TO1 and TO2. DTC operations in the Target phase also include complex communications among the four DTC team members. This teamwork relates to TO3. In operational settings and in the DDD simulation, the proposed order of attacking TSTs and the proposed strike package for each attack is approved or disapproved first by the SODO and then by the CCO. These two officers also control the execution of accepted proposals. In the present experiment, expert confederates played the roles of CCO and SODO proficiently so that variability related to approvals and to executions was minimal.

## Design

The within-participant independent variables were pretests and posttests in each of Phases II and III. The planned comparisons were the pretest versus the posttest in Phase I, which reflected the change in performance during that phase; the posttest in Phase I versus the pretest in Phase II, which reflected the drop in performance during the initial exposure to the new experience; and the pretest versus the posttest in Phase III, which reflected change in performance during Phase III.

The dependant variable was the quality of the proposed strike package (TO3) for each TST, which was determined by expert ratings of the strike package. Specifically, two experts evaluated each strike package. They independently rated the strike package as correct or incorrect with high or low confidence. The ratings were converted as

follows: incorrect with high confidence = 1, incorrect with low confidence = 2, correct with low confidence = 3, correct with high confidence = 4. In addition, submitting no plan before the TST disappeared = 0, and submitting a plan for prosecuting a target that was on the no strike list = -1. The expert ratings were the same for 96% of the ratings, and the few disagreements were resolved by conference so that one number was entered into the analysis for each TST. This measure is comprehensive in that the DTC had to accomplish TO1, TO2, and TO3 in order to get a high score.

## Procedure

Training missions included planning and debrief sessions with checklists analogous to those used by the Air Force (Elliott, Cardenas, & Schiflett, 1999). During planning sessions, the teams commit to plans including the ATO and possible TSTs. During mission execution, the offensive team judiciously adapts plans when unexpected events occur, and the DTC plays the key role of proposing specific adaptations. During debriefs, the team extracts lessons from previous plans and adaptations. The team then incorporates these lessons during subsequent planning, missions, and debriefs. Each session in the present experiment included these stages: planning (10 min), mission execution (40 min), and debriefing (10 min).

During Phase I (50 hours), research assistants taught background information using materials from Air Force publications. The background included the characteristics of enemy weapons and friendly assets, communication brevity procedures, and communication dynamics among the team members. The communication dynamics came from the Aptima task analysis. It included why and when teammates communicate with one another and the protocol for who communicates what to whom. Phase I ended with sessions during which the trainees or the instructors could stop the action to address issues. The trainees might stop action to ask questions. The instructors would answer the questions and then resume the action. The instructors might stop action to point out errors.

Training before these assisted missions taught what to do. The assisted missions taught how to do it.

During Phase II (49 hours), the trainees performed 49 unassisted sessions. The debriefing included feedback on how well the team was accomplishing its task objectives. Each mission was unique, but they had consistent patterns of differences and similarities, which were created by simulating the same enemy forces attacking with a consistent strategy. The mission difficulty was affected by the number of TSTs per mission and the number of threats that protected the TSTs. Phase II started with warm up trials, which had 3-7 TSTs and 10 threats to the strike package that attacked the TSTs. Two pre-test missions had 8 and 9 TSTs and 10 threats. These threats were 4 fighters, 1 long range SAM, and 5 short range SAMS. Two post-test missions in Phase II had 8 and 9 TSTs and 37 threats to the strike package. These threats were 24 fighters, three long range SAMS and 10 short range SAMS. These enemy threats required the strike package to include enough assets to take out not only the TST but also the threats to the strike package. Difficulty increased gradually over the 49 missions.

During Phase III (9 hr), the trainees performed 9 more unassisted missions, during which they trained for new experiences. The main change at the beginning of Phase III was a strategic variation similar to that described earlier. The enemy used its new strategy consistently throughout Phase III, so that the trainees had an opportunity to practice defending against the new strategy. The number of TSTs (8 and 9) and Threats (37) for the posttest in Phase II was similar to the number of TSTs (9 and 10) and Threats (33 and 35) for two pretest missions in Phase III. In contrast, each of two posttest missions for Phase III had 13 TSTs and 45 Threats.

## RESULTS

Before reporting quantitative results, a qualitative result is noteworthy. Scientists who had performed the task analysis of DTC dynamics in operational settings, observed the present simulated DTC near the end of Phase II. The scientists observed that the simulated DTC team performed DTC dynamics at about the level of an experienced operational DTC with a medium-high skill level.

This skill level was higher than we had anticipated causing a ceiling effect during Phase II. Specifically, the mean accuracy for TO3 increased from 2.9 to 3.5, as predicted, but the increase was not significant ( $t(34) = .09, p > .05$ ). In contrast, the ratings of TO3 accuracy were sensitive measures between Phases II and III, and within Phase III.

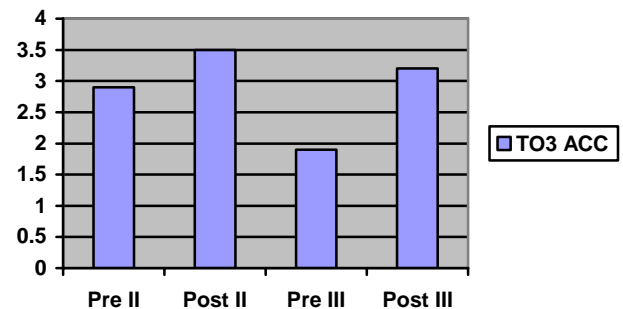


Figure 1. Mean accuracy ratings for training objective 3 (TO3) for the pretest and posttest in Phase II and for the pretest and posttest in Phase III.

The ceiling effect in Phase II was due to high performance on the pretest. In addition to showing this high initial performance, Figure 1 shows that TO3 accuracy fell between Phase II and III and then increased within Phase III. Between the posttest in Phase II and the pretest in Phase III, the TO3 accuracy decreased significantly from 3.5 to 1.9 ( $t(32) = 3.59, p < .001$ ). In Phase III, TO3 accuracy rose significantly from 1.9 on the pretest to 3.2 on the posttest ( $t(40) = 3.09, p < .01$ ).

## DISCUSSION

The initial high performance in Phase II, suggests that we gave too much familiarization training before we took our initial measures. We erred in this direction, because our main concern before the experiment was whether or not undergraduate research assistants could learn a task as complex as our DTC simulation.

The unmistakable improvement during Phase III is even more impressive when one considers that the posttest was more difficult than the pretest, which had fewer TSTs and threats.

The results suggest that a realistic laboratory simulation can represent the training of experienced operators to learn new experiences. Scientists and

practitioners can use this evidence to improve skills training in operational settings and to address important scientific questions. For example, the present results set the stage for scientists and practitioners to investigate whether simulating new experiences will help actual DTC teams adapt more quickly to real changes. At least two sets of hypotheses should be investigated. One set relates to the potential immediate benefits of simulating a specific new experience, such as confronting a new enemy in a new context. The second set relates to potential learning-to-learn benefits of simulating multiple new enemy forces, new contexts, and new enemy tactics. Such learning-to-learn effects for actual DTC teams simulating many new experiences might make the DTC teams generally more adaptive to new experiences in operational settings (cf. Ness, Tepe, and Ritzer, 2004.)

During experiments similar to the present laboratory experiment and to the proposed field experiments, scientists and trainers should address important scientific questions. For example, we are conducting time series analyses to investigate the task work and teamwork events underlying the DTC's proposed strike packages during original learning and during learning new experiences.

In addition, we are extending the present experiment to test two different training protocols for experienced teams learning new experiences. In a control condition, trainees will follow a predetermined path from the pretest to the posttest. In an experimental condition, a Partially Observable Markov Decision Process (POMDP) model will adaptively guide the selection of training scenarios between the pretest and posttest. We will test the hypothesis that experienced teams training for new experiences will perform better in the adaptive POMDP condition. Experiments with less experienced trainees and less complex tasks provide an empirical, theoretical, and applied foundations for this hypothesis ( e.g. Anderson, J. R., Douglass, S. & Qin, Y.,2004; Anderson, J. R. et al., 1995; Conati, C. et al., 1997; Schulze, et al., 2000). These foundations, which are based on fine-grained analyses and fine-grained interventions, consistently support the hypothesis that adaptive training improves learning. Testing this hypothesis in the present coarse-grained context will continue

the process of extending these foundations to experienced teams learning new complex experiences.

## REFERENCES

- Anderson, J. R., Douglass, S. & Qin, Y. (2004). How should a theory of learning and cognition inform instruction?. In A. Healy (Ed.) *Experimental cognitive psychology and it's applications*. American Psychological Association; Washinton, D. C.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences*, 4(2), 167-207.
- Conati, C., Gertner, A., VanLehn, K., and Druzdzel M. (1997). On-line student modeling for coached problem solving using Bayesian networks. In A. Jameson, C. Paris and C. Tasso (Eds.), *User Modeling: Proceedings of the Sixth International Conference, UM97*, 231-242. Vienna: Springer.
- Ness, J.W., Tepe, V., and Ritzer, D.R., 2004, *The Science and Simulation of Human Performance*. (Boston: Elsviver).
- Piorolli, Peter L. and Anderson, J.R., 1985, The role of practice in fact retrieval. *Journal of experimental psychology: Learning, memory, and cognition*, 11, pp. 136-53.
- Schulze, K. G., Shelby, R. N., Treacy, D. J., and Wintersgill, M. C. (2000a). Andes: A coached learning environment for classical Newtonian physics. In J. A. Chambers (Ed.) *Selected Papers from the 11<sup>th</sup> International Conference on College Teaching and Learning*, 151-162, FCCJ, Jacksonville, FL.
- Schmidt, R.A., and Bjork, R.B., 1992, New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, 4, pp. 207-217.