

EFFECTS OF FATIGUE ON SIMULATION-BASED TEAM DECISION MAKING PERFORMANCE

CHRISTOPHER BARNES
Air Force Research Laboratory

LINDA R. ELLIOTT
Army Research Laboratory

MICHAEL D. COOVERT
University of South Florida

DONALD HARVILLE
Air Force Research Laboratory

ABSTRACT

This paper describes a study examining the effects of fatigue on team decision-making performance in a command and control context. Ten three-person teams participated in an investigation of sleep deprivation on physiological state, cognitive function, and simulation-based performance. Teams participated in the study from 6:30 pm through 10:30 am the next morning. In this report, we describe preliminary analyses, focused on effects of sleep loss. Despite the small number of teams, significant results were found with regard to time, scenario, oral temperature, and math total points.

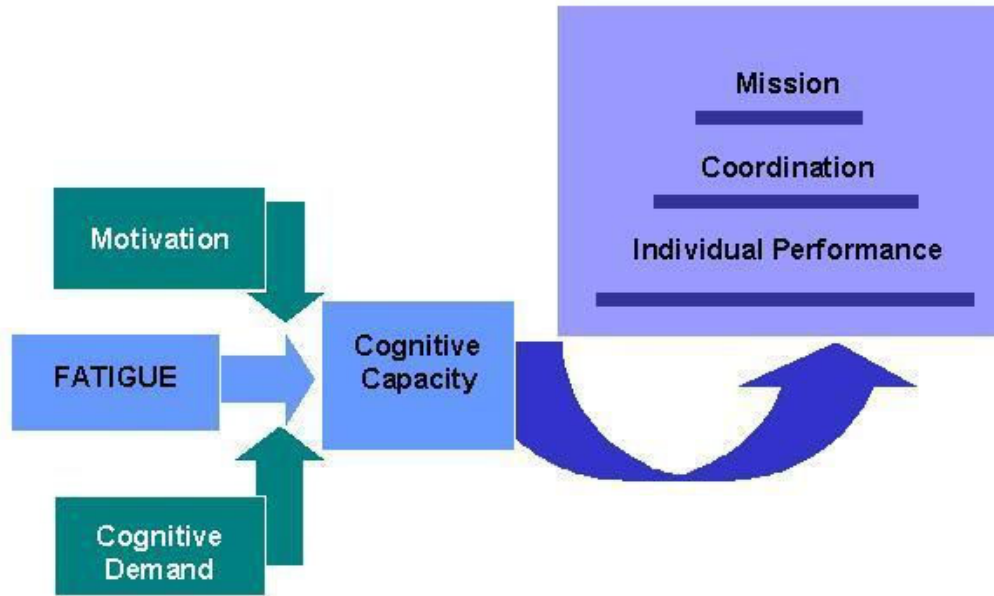
INTRODUCTION

United States Air Force (USAF) command and control (C2) warfighters face increasingly complex environments that represent the essence of decision making—multiple demands for enhanced vigilance, rapid situation assessment, and coordinated adaptive response. There are many perspectives on decision making, however, all would agree that decision making contexts are typified by expert, complex, interdependent and dynamic decision making, often under conditions of time pressure and/or uncertainty (Beach & Lipshitz, 1993; Cohen, 1993; Klein, 1993; Mitchell & Beach, 1990; Orasanu & Salas, 1991; Orasanu & Connolly, 1993; Rasmussen, 1993).

Sustained operations are integral to command and control—combat missions require vigilance over time and adaptive performance under stress. Situations requiring close coordination and adaptive replanning are increasingly prevalent and challenging. Requirements for multi-service coordination are increasing, in maneuvers that are mobile, rapid, dynamic, and constantly evolving. Current examples

include tactics such as battlefield interdiction and close air support in situations requiring rapid movement of troops and armament (Elliott et al., 2002).

While extensive data are available on effects of sleep loss on physiological, attitudinal, and cognitive function (Kryger, Roth, & Dement, 2000), very few studies reported data regarding sleep loss effects on particular aspects of information processing in complex decision making tasks (Mahan, 1992; 1994). Even fewer have reported on effects on team performance (Elliott, Coovert, Barnes, & Miller, 2003; Harville, Elliott, Barnes, & Miller, 2003); however, a few preliminary studies, based on team simulation-based performance, provide some introductory results (Mahan, Elliott, Dunwoody, & Marino, 1998; Elliott, Coovert, & Miller, 2003). To continue this stream of research, the Chronobiology and Sleep Laboratory at Brooks City-Base, San Antonio, TX has initiated a program of research on effects of sleep loss on information processing, communication, coordination, and decision making in complex simulation-based tasks. Figure 1 provides a representation of our overall approach to

Figure 1. Modeling Fatigue Effects on Performance

constructs, measures, and relationships, across a sequence of studies.

The model predicts that fatigue interacts with cognitive demand to influence decision making and mission performance. More specifically, cognitive demands are expected to utilize cognitive resources from individual cognitive capacity (knowledge and ability), consistent with resource allocation models such as the Kanfer-Ackerman model of learning and motivation (Kanfer & Ackerman, 1989; Kanfer, 1990). An underlying and general assumption is that fatigue is expected to reduce individual cognitive capacity. As this capacity is reduced, performance will be affected negatively. Motivation moderates the relationship between capacity and performance.

In the overall model, fatigue diminishes total cognitive capacity, with increasing decrement over time. This systems view is consistent with quantitative research on effects of fatigue and chronobiology which supports the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model, which outlines effects of fatigue and chronobiology in more specific detail (Eddy & Hursh, 2001; Hursh, 1998)

METHOD

Participants

Research participants were drawn from a pool of USAF officers awaiting Air Battle Management

Training at Tyndall Air Force Base, FL. A total of ten 3-person teams participated in this study. All of the participants had already attended the Aerospace Basics Course, which however provided them with little training or knowledge useful for the current study.

Each participant participated in a 40-hour training session occurring during a one-week period. The week included one hour of administrative processing, nine hours training on the Automated Neuropsychological Assessment Metric (ANAM) cognitive test battery (Reeves, Winter, Kane, Elsmore, & Bleiburg, 2001) to reach specified performance levels, and 30 hours training on Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) assets, capabilities, and tactics, along with Agent Enabled Decision Group Environment (AEDGE™) interface functions. The subjects were trained in three distinct C2 functional roles: ISR, Sweep, and Strike. The ISR role owns assets related to ISR functions, such as unmanned aerial vehicles (UAV). The Strike role owns assets such as air-to-ground bombers and airborne jammers, while the Sweep role owns assets such as air-to-air fighter aircraft. During this time, the various ergonomic features of the chairs and workstations were explained and demonstrated to the participants. Participants learned principles of these ergonomics and how they can be adjusted to aid with fatigue effects.

The experimental session began at 6pm on the last day of training (always a Friday) and ended at 11am the following morning. With one subject in the role of



Strike, one as Sweep, and one as ISR, they participated as three-person teams, every other hour, in eight 40-minute team-based C4ISR decision making scenarios, with 20 additional minutes during each session for debriefing, data collection, and mission planning for the next session. Their roles as Strike, Sweep, or ISR did not change during the experimental sessions. Every other hour, between each scenario session, they performed on the ANAM cognitive test battery that assesses reaction time, working memory, simple mathematical processing, and multitasking (Reeves et al., 2001). After each cognitive battery session, they provided physiological data (e.g., temperature, actigraphy), and self-reports on mood-state, and sleepiness. All email and audio communications were digitally captured for transcription. This resulted in extensive cognitive performance and simulation-based process and performance data.

Preliminary criterion measures of simulation-based performance were generated from a PC-based synthetic team task environment developed for investigations of C4ISR team performance. The AEDGE™ was developed based on cognitive and functional analysis of C3 mission, tactics, team member roles, and role interdependencies (Chaiken, Elliott, Dalrymple, & Schiflett, 2001, Barnes, Petrov, Elliott, & Stoyen, 2002). Tactical scenarios were developed to capture core team coordination, decision-making and problem-solving task demands. Platforms such as the AEDGE™ provide an advanced PC-based platform for research and/or training. The advantages of these capabilities are increased experimental control, manipulation, and operational relevance (Bowers, Salas, Prince, & Brannick, 1992; Cannon-Bowers, Burns, Salas, & Pruitt, 1998; Covert, Craiger, & Cannon-Bowers, 1995; Schiflett & Elliott, 2000). Functional and cognitive fidelity was based on cognitive task analyses (Chaiken et al., 2001).

Mission scenarios were typified by a strong demand for communication, shared awareness, coordinated action, and adaptive response to time-critical situations. Scenarios requiring dynamic replanning were carefully constructed to ensure equivalence in task demand and difficulty. This is particularly critical and challenging within this repeated-measures context. Two critical issues must be addressed: that of fidelity and equivalence of scenarios and event-based measures.

Equivalence of Measures

Sustained operations research has particular demands with regard to repeated-measures.

Measures have to be repeated over time in order to ascertain effects of fatigue. However, measures often cannot be replicated because of the need to minimize practice or learning effects. Even relatively simple cognitive tests that assess reaction time, working memory, or attention-switching require preliminary training to asymptote performance prior to the experimental session. Measures of more complex performance, such as logic or problem solving, are more difficult to assess over time, as most available tests do not have many equivalent forms. For many types of problems, repetition will elicit recognition-based performance: participants are more likely to increase performance because they remember the problem.

Performance in the C4ISR scenarios will also increase, if the same scenario is used repeatedly. This prevents the assessment of fatigue effects. Once participants realize the same scenario is repeated, they will anticipate events and create strategies to increase performance while minimizing effort.

In the current study, each team of participants experienced only one overnight session. During the session they performed in eight different scenarios. The challenge inherent in this experimental design was the requirement of equivalence in scenario difficulty. It was important to avoid confounding results with scenarios varying in workload complexity or demand, and it can be quite difficult to craft scenarios that result in similar mean outcome scores.

Equivalent scenarios were constructed by assuring that all scenarios had (a) similar roles, (b) equivalent friendly assets, (c) equivalent hostile assets, (d) equivalent timing and tempo of events, (e) equivalent timing and tempo of additional hostile and friendly assets, and (f) equivalent geographic distances between hostile and friendly assets. Geographic distances affect the timing of hostile-friendly encounters and thus affects the tempo of workload demand. Each scenario had an ISR (Intelligence, Surveillance, and Reconnaissance), Strike, and Sweep role played by participants. Each role had similar assets and tactical goals. Assets were allocated across hostile and friendly roles in the same manner. For example, the ISR role had the same number and type of UAV assets at the beginning of each scenario, and had additional assets appear at the same time through each scenario. He/she would face similar threat events, with regard to the number, type, and timing of hostile events. The same kinds of coordinating actions



among the friendly roles were required in each scenario.

Recognition of the underlying "deep" structure of each scenario is minimized by changing the "surface" structure of each. One way this was achieved was by changing the geographic context and placement of assets. One scenario may be located in the geographic region of Taiwan, while another would be situation in Sri Lanka. The number and placement of assets would be equivalent, but not readily recognized. Another way this was achieved was by changing the type of hostile threat. In one version of the scenario, hostile threats were comprised of enemy surface-to-air missile sites. This situation is equivalent to a military tactic described as SEAD (suppression of enemy air defense). In another version, the hostile targets were theatre ballistic missile launchers. Identification and targeting of these targets are often referred to as "scud-hunting." The third version used in this study had hostile ships as enemy targets.

Scenario events were also timed to be equivalent. Assets appeared at particular times in each scenario. For example, in each scenario, hostile fighter aircraft appeared at specified times. Other scenarios have the same type and timing of events, where only the names of the assets change. Thus, in each scenario, the same cognitive and functional demands are presented to each role.

Measurement of Performance

A variety of measures were collected, including individual scenario score, team scenario score, oral temperature, and math score on a cognitive test battery. The math score consisted of number of correct addition problems in a set time period.

Mission Outcomes

Raw measures of mission outcome and team process were captured and time-stamped by the simulation. This includes descriptions and counts of events and actions, which then form the basis for various assessments of performance. For example, mission outcome scores were represented by the type, number, and relative value of assets that were lost, by "friendly" and "hostile" roles. Friendly assets included air bases, cities, surface-to-air missile launchers, uninhabited aerial vehicles, tanker aircraft, high-value reconnaissance aircraft, fighter aircraft, and bomber aircraft. Each asset was given a relative score value, generated by our weapons director expert, and validated by other experienced weapons directors. The loss of any friendly asset detracts from the score of the friendly team, and adds to the score of the enemy. In turn, hostile

assets are similar. The loss of hostile assets adds to the score of the friendly team, and detracts from the score of the hostile. For these research participants, the overall mission outcome score was based on the point value obtained after subtracting all friendly "losses" from the total hostile "losses."

Audio Capture of Communications

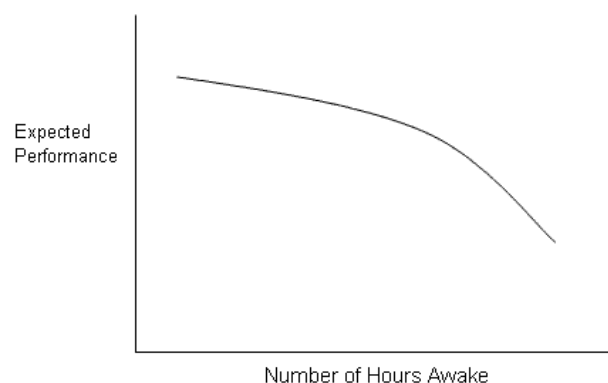
Communications were recorded in digital format to ease coding and analyses of data. Communications were initially coded for indications of teamwork, such as sharing of information or assets, sequencing actions, acknowledgements, requests for repeats, task-related encouragement, expressions of fatigue, and social comments (positive and negative). All comments were coded as to whether they requested or provided information.

Additional measures of individual characteristics include the Stanford Sleepiness Scale, the Profile of Mood States, the NEO-PI (all subscales), and performance on the ANAM cognitive test battery. The ANAM includes measures of reaction time, working memory, and multi-tasking ability. In addition, all subjects provided estimates prior to each scenario, regarding the likelihood of attaining differing categories of performance outcomes, and afterward, their satisfaction with their outcomes.

Multilevel Modeling

Multilevel modeling was particularly suited to fatigue research due to the necessity of repeated measures testing. Hierarchically structured data also occurs when the same individuals or units are measured on more than one occasion. A common example occurs in studies of animal and human growth. Here the occasions are clustered within individuals that represent the level 2 units with measurement occasions the level 1 units.

Figure 2



RESULTS

The data for this study are arranged hierarchically. The outcome variable of interest is the team performance score and there are 240 observations total, however, three scenario scores for three teams were deleted due to administrative problems. This gave an effective sample size of 231 cases. As described earlier, teams completed several scenarios across the night. Figure 2, depicts what we would expect the longer one is kept awake -- performance declines. This leads to negatively accelerated growth curve for performance across the weekend. On the other hand, we would expect the team's performance on the task to increase as they become more proficient on the task and develop better teamwork skills. This results in a positive growth curve.

A series of multilevel models from least to most complex is tested to examine team performance. The data are hierarchical in that *occasion* (which repeated measure administration, 1-8) is nested within the *individual*, which is nested within *team*. So *occasion* is a level-1 variable indicating the testing session. *Individual* is a level-2 variable and indicates which research participant, and *team* is a level-3 variable. To ensure scores across the eight scenarios are comparable, team scores were centered on each scenario (for a discussion on the importance of this see Kreft & de Leeuw, 1998 pp. 109-115 or any multilevel textbook). All analyses were conducted with the MLWIN software package.

The first model is a null model and is computed for comparison purposes. The model states that:

$$\text{teamscore}_{occasion, individual, team} = b_0 \text{occasion, week, team} \text{ cons (1)}$$

where: *teamscore* is the score obtained by the team, the subscripts *occasion*, *individual*, and *team* are as defined above, b_0 is a regression weight, and *cons* refers to a constant. (Due to space constraints we do not present the variance component estimates $W_{team}W_{individual, team} W_{occasion, individual, team}$ that correspond with equations 1-4.) Overall model deviance (lack of fit) is 4262.14. Two level-1 predictors of interest are the amount of time into the experimental session the participants are (how long they have been awake) and which scenario is being run. Adding in the level-1 predictor *time* and *scenario* into the model results in equation (2) and the solution reduces overall model deviance to 4250.46. Differences in model deviance are distributed as a chi-square so this reduction is significant, $c^2_2=11.68, p < .01$.

$$\text{teamscore}_{occasion, individual, team} = b_0 \text{occasion, individual, team} \text{ cons} + b_1 \text{time}_{occasion, individual, team} + b_2 \text{scenario}_{occasion, individual, team} \text{ (2)}$$

It is useful to examine the beta weights for the substantive variables and determine if they are significant. Significance is determined by dividing the beta by its standard error of measurement (sem). If greater than |2| the beta is significantly different than zero. The beta for time is .803 with a sem of .254, so the estimate is significant. What this means is that for each unit increase in the amount of time kept awake, team performance declines by .8 points. Since *scenario* is a dummy coded variable, it is not interpreted for the present purposes.

Equation three represents the addition of the individual's oral temperature into the model. Oral temperature is thought to mirror the stage of the individual's circadian rhythm.

$$\text{teamscore}_{occasion, individual, team} = b_0 \text{occasion, individual, team} \text{ cons} + b_1 \text{time}_{occasion, individual, team} + b_2 \text{scenario}_{occasion, individual, team} + b_3 \text{oral-temp}_{occasion, individual, team} \text{ (3)}$$

Overall model deviance is reduced to 4231.1 which is a highly significant reduction, $c^2_1 = 19.36, p < .001$.

Another series of models was run to look at the effect of staying awake on the cognitive performance battery tests and if any were predictive of team score. None of the variables further decreased overall model deviance except for the *math total points* score.

$$\text{teamscore}_{occasion, individual, team} = b_0 \text{occasion, individual, team} \text{ cons} + b_1 \text{time}_{occasion, individual, team} + b_2 \text{scenario}_{occasion, individual, team} + b_3 \text{oral-temp}_{occasion, individual, team} + b_4 \text{math-totalpoints}_{occasion, individual, team} \text{ (4)}$$

Results from the addition of using the math-total points as a predictor is a significant reduction in overall deviance to 4225.82, $c^2_1 = 6.28, p < .02$.

A final series of models was run to see if these slopes and intercepts might be modeled better as random coefficients. Overall deviance decreased, but not significantly. Parsimony argues for keeping the results as non-random coefficients.

Table 1 provides a summary of the results for overall model fit and incremental improvements.



Table 1
 Summary of the results for overall model fit and incremental improvements

<u>Model</u>	<u>Overall Deviance</u>	<u>Improvement</u>
Null	4262.14	
added Time, Scenario	4250.46	11.68**
Time, Week, Oral-Temperature	4231.1	19.36***
Time, Week, Oral-Temperature, Math-Total Points	4225.82	6.28*

*p < .02. **p < .01. ***p < .001

DISCUSSION/CONCLUSION

Despite the small number of teams, significant results were found with regard to time, oral temperature, and math total points. Each contributed to reducing the overall deviance of the team score. These results were as expected, indicating an effect of fatigue on team performance. Results suggest a decrease in cognitive capacity under fatigued conditions, which shows effects at both the individual and team levels, consistent with circadian rhythm models. It was also expected that more of the cognitive battery tests would be associated with the team scores, but only the math total points scores were significant.

Further stages of this study are currently in the planning process. The next stage will increase the sample size, providing more statistical power. It is encouraging that significant results have already been found at this early stage, and it is expected that future stages will further clarify the effects of fatigue on team performance. It is already clear at this point that fatigue has an effect on team performance. Future steps will include better quantifying these effects and eventually creating strategies to minimize and counter such effects. Other analyses utilizing data collected as a part of these efforts are currently being conducted, including communications analysis and command and control scenario process and outcomes measures.

REFERENCES

- Barnes, C. M., Petrov, P. V., Elliott, L. R., & Stoyen, A. (2002). Agent based simulation and support of C3 decisionmaking: Issues and opportunities. *Proceedings of the Conference on Computer Generated Forces and Behavior Representation*.
- Beach, L. R. & Lipshitz, R. (1993). Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. Norwood, NJ: Ablex Publishing Corporation.
- Bowers, C., Salas, E., Prince, C., & Brannick, M. (1992). Games teams play: A method for investigating team coordination and performance. *Behavior Research Methods, Instruments & Computers*, 24, 503-506.
- 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: American Psychological Association.
- Chaiken, S., Elliott, L. R., Dalrymple, M., & Schiflett, S. (2001). Weapons director intelligent agent-assist task: Procedure and findings for a validation study. *Proceedings of the 6th International Command and Control Research and Technology Symposium*.



- Cohen, M. S. (1993). The naturalistic basis of decision biases. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsombok (Eds.) *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.
- Covert, M. D., Craiger, J. P., & Cannon-Bowers, J. A. (1995). Innovations in modeling and simulating team performance: Implications for decision making. In R. Guzzo & E. Salas (Eds.), *Team Effectiveness and Decisionmaking in Organizations* (pp. 149-203). San Francisco, CA: Jossey Bass.
- Elliott, L. R., Barnes, C., Brown, L., Fischer, J., Miller, J. C., Dalrymple, M., Whitmore, J., & Cardenas, R. (2002). Investigation of complex C3 decisionmaking under sustained operations: Issues and analyses. *Proceedings of the 7th International Command and Control Research and Technology Symposium*.
- Elliott, L. R., Covert, M., Barnes, C., & Miller, J. C. (2003). *Modeling performance in C4ISR sustained operations: A multi-level approach. Proceedings of the 8th International Command and Control Research and Technology Symposium*.
- Elliott, L. R., Covert, M., & Miller, J. C. (2003, April). Ascertaining effects of sleep loss and experience on simulation-based performance. Poster session presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Harville, D. L., Elliott, L. R., Barnes, C., & Miller, J. C. (2003). Communication and decisionmaking in C4ISR sustained operations: An experimental approach. *Proceedings of the 8th International Command and Control Research and Technology Symposium*.
- Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 75-170). Palo Alto, CA: Consulting Psychologists Press.
- Kanfer, R. & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative / aptitude approach to skill acquisition. *Journal of Applied Psychology*, 74, 657-690.
- Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decisionmaking. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsombok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.
- Kreft, I., & De Leeuw, J. (1998). *Introducing Multilevel Modeling*. Thousand Oaks, Sage.
- Kryger, M., Roth, T., & Dement, W. (2000). *Principles and practices of sleep medicine* (3rd ed.). Philadelphia, PA: W. B. Saunders Company.
- Mahan, R. P. (1992). Effects of task uncertainty and continuous performance on knowledge execution in complex decision making. *International Journal of Computer Integrated Manufacturing*, 5(2), 58-67.
- Mahan, R. P. (1994). Stress-Induces strategy shifts toward intuitive cognition: A cognitive continuum framework approach. *Human Performance*, 7(2), 85-118.
- Mahan, R. P., Elliott, L. R., Dunwoody, P., & Marino, C. (1998, April). Team decision making under stress: The effects of sleep loss, continuous performance, and absence of feedback on hierarchical team decisionmaking. Paper presented at the Aerospace Medical Panel Symposium on Collaborative Crew Performance in Complex Operational Systems, Edinburgh, Scotland.
- Mitchell, T. R., & Beach, L. R. (1990). ...Do I love thee? Let me count ... Toward an understanding of automatic decision making. *Organizational Behavior and Human Decision Processes*, 417, 1-20.
- Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsombok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.
- Orasanu, J., & Salas, E. (1991). Team decision making in complex environments. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsombok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.
- Rasmussen, J. (1993). Deciding and doing: Decision making in natural contexts. . In G. Klein, J. Orasanu, R. Calderwood, & C. Zsombok (Eds.) *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.
- Reeves, D., Winter, K., Kane, R., Elsmore, T., & Bleiberg, J. (2001). *ANAM 2001 User's Manual*. (Special Report NCRF-SR-2001-1). San Diego, CA: National Cognitive Recovery Foundation.



Schiflett, S. G., & Elliott, L. R. (2000). Synthetic team training environments for command and control. In Dee Andrews and Mike McNeese (Eds.), *Aircrew Training Methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

Editor's Note: Pagination in this article may be different from the citation source due to formatting and layout conversions

