Risk Assessment War-game (RAW)

By

Kent Miller, U.S. Army

Don Brown, University of Virginia

During Military Decision Making Process (MDMP) Course of Action Analysis (COAA), the commander and staff rely on interactive-active role-playing, Simulated Interactions (SI), to analyze each prospective Friendly Course of Action (FCOA). In an effort to predict whether an FCOA will meet the goals and objectives at a critical event, the "blue" team and the "red" team act out the exchanges between the "friendly" and threat forces, respectively. Although the stated purpose of the doctrinal war-game is to determine which "COA accomplishes the mission with minimum casualties while best positioning the force to retain the initiative for future operations" (U. S. Army 1997), the Center for Army Lessons Learned (CALL) was unable to find records of any formal study assessing the effectiveness of the doctrinal war-game. (CALL 2003)

In the doctrinal COAA war-game, the friendly actions are fixed and all the uncertainty is associated with the threat reactions. However, the war-game does not capture the uncertainty in these threat reactions. A recent Army Research Laboratory (ARL) study identified this failure to account for the uncertainty in the threat's reactions as a significant weakness of the MDMP. (Kaste, Bodt et al. 2000)

In this paper, we present our methodology for using SI to quantitatively assess risk during COAA. We will show that the Risk Assessment War-game (RAW) provides the theoretical framework for a thorough, quantitative risk assessment whereas the doctrinal war-game does not. We will demonstrate how the commander and staff can use the quantitative risk assessment for risk management within each alternative and for selecting a Pareto optimal COA based upon the commander's degree of risk aversion.

We realize that regardless of the theory, we must demonstrate to military planners that RAW is both practical and of greater value than the doctrinal war-game. We conducted three experiments in the military domain comparing RAW to the doctrinal war-game. In this paper, we present our findings from these three experiments.

A risk assessment is an integral part of a risk analysis. Risk is the measure of the probability and severity of adverse effects. (Lowrance 1976) A thorough risk assessment is critical to effectively managing risk. (Haimes 1998) A risk assessment should answer three questions: what can go wrong, what is the likelihood it will go wrong, and what are the consequences. (Kaplan and Garrick 1981) As applied to the COAA, a thorough risk assessment would identify the threat reactions which could lead to mission failure and assess the likelihood of these reactions.

Nobel laureate Harry Markowitz proved one could effectively manage financial risk through asset diversification. Applying his portfolio theory, one could distinguish between efficient and inefficient portfolios based upon an analysis of reward and risk. As a result of his pioneering work, an investor can select a Pareto optimal portfolio based upon their individual degree of risk aversion. (Markowitz 1987)

Decision-makers in conflict environments should also weigh reward and risk in selecting a strategy. If commanders could somehow quantify reward and risk, they could select a Pareto COA based upon their degree of risk aversion. However, unlike the financial world, estimating means and variances based upon historical data is highly suspect. As evidenced by the ongoing insurgency in Iraq as well as our continued struggle against Al Qaeda, adversaries are not restricted to selecting a response based upon a set of rules developed largely from observance of past analogous situations.

In this paper, we use the Bayesian interpretation of probability. A probability represents a degree of belief in a hypothesis. (Bernardo and Smith 1994) During the interactive role-play, the subjective probability associated with a threat reaction represents the red team's degree of belief as to which reaction the enemy will execute.

The use of role-playing in conflict forecasting is widespread in business (Armstrong 1987), the legal profession (Cooper 1977), as well as the military (Golhamer and Speier 1959). Role-playing is a judgmental forecasting methodology which involves people assuming positions to understand or predict behavior. (Armstrong 2001) Inter-active role-playing (i.e., SI) involves the acting out of exchanges between individuals as well as groups. (Green 2003) Proponents of SI belief it facilitates a realistic examination of the interaction among parties and is superior to other judgmental forecasting methods for predicting decisions and outcomes in conflict situations. (Armstrong 2002)

Prior to the late 1970's, few scientific studies attesting to the effectiveness of role-playing in forecasting existed. Studies such as the one involving hypnotic behavior (Orne 1968) and a prison simulation (Zimbardo 1972) added significant credibility to the claims of realism. Limited experimentation with mock trials (Kerr 1979) and in psychology (Mixon 1972) (Armstrong 1987) along with anecdotal evidence from the Vietnam War (Halberstam 2001) suggested accuracy. However, surveys of the use of role-playing in political gaming (Mandel 1977), the legal profession (Gerbasi 1977), and psychology literature (Kerr 1979) found no strong evidence of its' effectiveness.

Beginning in the late seventies, Armstrong pioneered experiments comparing the accuracy of role-playing in predicting a decision to other conflict forecasting methods in actual situations. In each of the four retrospective and one prospective study he conducted, role-players more accurately predicted the actual decision than the subjects rendering an individual opinion. (Armstrong 2001) In more recent studies, Green extended Armstrong's research to provide evidence that role-playing is more accurate than game theory and unaided judgment in predicting a decision. (Green 2002)

Although these experiments suggest role-playing is the most accurate conflict forecasting method, the accuracy rate is low. On average, role-players correctly forecasted the decision only 56% of the time. (Armstrong 2001) If this result is indicative of the staff's ability to accurately forecast the enemy's decisions (i.e., threat reactions), we submit the staff must consider more than the "most likely" enemy course of action.

Therefore, we submit that an interactive role-play which provides the framework for forecasting the feasible set of threat reactions is of greater value to the commander and staff than a methodology which facilitates only the forecast of a single reaction. Forecasting a single threat reaction is analogous to point estimation. Forecasting the set of feasible set of threat reactions and associating likelihoods results in a probability distribution function. Given the set of possibilities and associated probabilities, the commander and staff can better assess and manage risk. In our search of the conflict forecasting literature, we did not find any study investigating the effectiveness of SI in assessing risk.

The doctrinal war-game is intended to facilitate the analysis of a single FCOA considered against a single Enemy COA (ECOA). Prior to the start of the game, the blue team decomposes the FCOA into critical events. For each critical event, the doctrinal war-game follows an action-reaction-counteraction cycle.(U. S. Army 1997) The blue team states the blue force action or counter-action. As appropriate, the red team responds with the "probable" or most dangerous threat reaction to each action or counteraction. The sequence is continued until the critical event is completed. (U. S. Army 1997)

In a thorough risk analysis, the commander and staff should war-game each FCOAs against all feasible enemy courses of action.(U. S. Army 1997) If the red team had identified n feasible ECOAs during the initial IPB, the commander and staff should conduct n independent war-games, which would result in the forecast of at most n sequences for each critical event. However, in a time-constrained planning environment, the commander and staff often cannot war-game each FCOA against all feasible enemy courses of action. To save time, they may consider only one FCOA against the most likely and/or most dangerous ECOA. Although developing only one FCOA and/or reducing the number of ECOAs war-gamed saves time, limiting the number of COAs increases risk to the command.(U. S. Army 1997)

RAW is designed to facilitate the analysis of a single FCOA simultaneously considered against all feasible ECOAs in a single war-game in a time constrained environment. As with the doctrinal war-game, the blue team decomposes the FCOA into critical events. RAW follows the same action-reaction-counteraction cycle. Two significant differences are that in RAW, the red team identifies the feasible set of threat reactions to each action or counteraction and associates a subjective probability with each red force reaction in the threat reaction set.

We will now present the RAW methodology. For each critical event, the commander and staff complete steps 1-7. Sequences resulting in failure are terminal. If the critical events are dependent and in series, each successive critical event continues from each successful sequence carried forward from the previous critical event. Within each critical event, the commander and staff repeat steps 2-4, as necessary.

Step 1. State the blue force action

Step 2. State the set of mutually exclusive and exhaustive, feasible red force responses to the blue force action/counteraction

Step 3. Elicit the red team's degree of belief in each red force reaction

Step 4. For each red force reaction, if appropriate, state the blue force counteraction

Step 5. Calculate the likelihood of each sequence

Step 6. Determine the outcome for each sequence

Step 7. Identify the sequences that lead to mission failure

Step 8. Calculate the likelihood of critical event failure

Step 9. After each critical event is war-gamed, compute the overall likelihood of mission failure.

A thorough risk assessment should identify all action-reaction-counteraction sequences that could lead to mission failure. In order to identify these sequences, the commander and staff must consider all possible threat reactions. In the doctrinal war-game, the red team states only one of the possible red force reactions to each blue force action. This exchange yields only one possible sequence. Using RAW, the red team can state the set of possible red force reactions, to each blue force action or counteraction, resulting in, m , possible sequences. If $m > 1$, RAW provides the framework for identifying all possible failure sequences (i.e., what can go wrong); whereas, the doctrinal war-game does not.

As emphasized earlier, a thorough, quantitative risk assessment must identify what can go wrong, what is the likelihood, and what are the consequences. The doctrinal war-game does not identify all the sequences that could lead to mission failure or capture the uncertainty in the threat reactions; therefore, the doctrinal war-game does not provide the framework for a thorough risk assessment. RAW facilitates the identification of all the sequences which could lead to mission failure and captures the uncertainty in the threat reactions; therefore, RAW provides the framework for a thorough, quantitative risk assessment.

With a quantitative risk assessment, commanders and staff can better manage risk within a COA. With knowledge of the sequences that could lead to mission failure, military planners could make modifications to the COA to mitigate the risk. Based upon likelihoods associated with the problematic sequences, the commander can systematically decide upon acceptable risk.

With a quantitative risk assessment, commanders can select a Pareto optimal COA based upon their degree of risk aversion. According to current doctrine, the staff recommends the FCOA with the highest score based upon the ranking and weighting during the COA comparison phase. One can think of this score as the reward associated with each FCOA. Selecting an FCOA based only upon this reward is like choosing an investment based only upon the expected return, given favorable market conditions. Commanders should select an FCOA based upon a trade-off analysis between reward and risk. RAW provides the quantitative measure of risk necessary to conduct this analysis.

Given the added complexity of RAW, we developed a software proto-type for use in our experiments. The software captures the action-reaction-counteraction cycle in a relational database, graphically displays the sequences in an event tree, facilitates the outcome assessment of each sequence, and performs all probability calculations. By default, the software employs the Analytical Hierarchy Process (AHP) in eliciting the subjective probabilities and ensuring the consistency of the pair-wise comparisons. (Monti and Carenini 1999) Given that the S2 is already required to arrange the ECOAs in probable order of execution, AHP appears an appropriate choice of methods. (U. S. Army 1994)

We conducted three experiments designed to investigate the utility and practicability of RAW from the perspective of potential users (i.e., military planners.) The test subjects ranged from novice to experienced military planner. We conducted the first experiment using University of Virginia Army ROTC cadets; the second experiment with the officers of the 30th Engineer Battalion at Fort Bragg; and the final experiment with students enrolled in the USMC School of Advanced War-fighting (SAW) at Quantico.

We used an anonymous web based survey to elicit information from each study participant. Using a five point Likert Scale, respondents were requested to state their level of agreement with a series of statements. We tested the reliability and validity of the survey using established principles and guidelines. (Walonick 2003)

The first experiment took place on 18 March 2004, during a scheduled 2-hour Military Science (MS) lab. Twenty, MS1 and MS2, ROTC cadets participated. The cadet battalion commander divided the cadets into nine, 2-3 person groups, and then randomly assigned the groups to either section A or B. Other than classroom instruction, none of the cadets had experience as military planners.

The experiment consisted of two orders drills. For each drill, the cadet battalion commander issued the participants a platoon level Operations Order (OPORD) in which he identified three possible ECOAs, the most likely, the most dangerous, and another feasible alternative. For each platoon OPORD issued, each group generated a squad level OPORD. The first issued OPORD was a platoon movement-to-contact mission. The second issued OPORD was a platoon zone reconnaissance mission.

Each group produced both a movement-to-contact and reconnaissance OPORD. For the movement-to-contact mission, groups assigned to section A used the doctrinal war-game while groups assigned to section B used RAW. For the zone reconnaissance mission, groups assigned to section B used the doctrinal war-game while groups assigned to section A used RAW.

Because of a scheduling conflict, three of the twenty cadets had to depart the lab early and were unable to complete the user survey. Of the actual survey respondents, five were in their first year of MS studies and twelve were in their second year.

The second experiment was conducted at Fort Bragg, North Carolina on 7 April 2004 with the leadership of the U.S. Army's 30th Engineer Battalion as part of a 3-hour Officer Professional Development (OPD). The twenty participants consisted of: 2 majors, 5 captains, 11 lieutenants, and 2 warrant officers. Eighteen of the officers were engineers. Of the remaining two, one was a military intelligence officer and the other a chemical officer. Two of the twenty had extensive planning experience, 4 had significant experience, 9 had some and the remainder had no experience.

As was the case in the ROTC experiment, the experiment consisted of two orders drills. The battalion S3 issued the participants two task force level OPORDs. For each issued task force OPORD, each group prepared a company level OPORD. The first mission involved securing key terrain surrounding an airfield. For this mission, the battalion S2 identified three possible enemy courses of action: the most likely, the most dangerous, and another plausible alternative. The second OPORD was a follow-on mission from the first.

The twenty officers were divided into ten, 2 person groups. The ten groups were randomly assigned to either section A or B. For the initial mission, groups assigned to section A used the doctrinal war-game while groups assigned to section B used RAW. For the follow-on mission, groups assigned to section B used the doctrinal war-game while groups assigned to section A used RAW.

We conducted the third and final experiment at Quantico, Virginia with students from the USMC's School of Advanced Warfighting (SAW). The School of Advanced Warfighting provides graduate-level military education tailored to complement and expand on the attendees previous year's study at the Command and Staff College or other Intermediate Level School. The School's focus is on decision-making and complex problem-solving experience at the operational level. Attendees are preparing for high-impact, MEF-level and higher service, joint and multinational billets. (School of Advanced Warfighting 2004)

Twenty-one officers, three lieutenant colonels and eighteen majors participated. Fifteen were Marine Corps officers, the remainder were from sister services and allied countries. Several military specialties were represented: eight infantrymen, five artillerymen, three aviators, three logisticians, one military intelligence officer, and one Special Forces officer.

All were experienced military planners. Three said they had extensive planning experience. Another eighteen claimed significant military planning experience.

The students were divided into four Operational Planning Teams (OPTs). For this experiment, two of the OPTs served as a control and used the doctrinal approach to war-gaming. The other two groups employed RAW. Only surveys from the OPTs that used RAW are included in the results.

The four OPTs played the role of III MEF staff officers for the planning associated with the movement of the 9th MEB from Okinawa to the III MEF area of operations and the associated MPF operation necessary for the conduct of military operations in concert with adjacent joint forces. The Head, School of Advanced Warfighting (SAW) played the role of the Commanding General, III MEF. Other SAW staff members simulated the roles of United States Transportation Command (USTRANSCOM) personnel. Each OPT produced two briefs—one Deployment Mission Analysis Brief and one detailed Concept of Deployment Brief.

In all three experiments, the majority of military planners said that RAW was practical in a deliberate planning process. (See Table 1) Ninety-four percent of the ROTC cadets and 70% of the 30th Engineer Battalion officers believed that RAW was practical in a tactical planning environment. Eighty percent of the SAW students believed that RAW was practical in an operational planning environment. Only 10% of the Engineer population and 20% of the SAW population stated that RAW was impractical.

	UVA ROTC	30th EN BN	USMC SAW	TOTAL
Strongly Agree	3	4		7
Agree	13	10	8	31
Neutral	1	4		5
Disagree		2	2	4
Strongly Disagree				

Table 1: RAW is practical

In all three experiments, the majority of military planners said that RAW adequately captured the uncertainty in the threat's reactions. (See Table 2) Sixty-five percent of the ROTC cadets as well as 60% of both the 30th Engineer Battalion officers and SAW students agreed or strongly agreed. Only 6% of the ROTC population and 10% of the SAW population disagreed.

	UVA ROTC	30th EN BN	USMC SAW	TOTAL
Strongly Agree	6	3	1	10
Agree	5	9	5	19
Neutral	5	8	3	16
Disagree	1		1	2
Strongly Disagree				

Table 2: Raw adequately captures the threat uncertainty

Many of the military planners were uncertain with respect to whether or not the elicited subjective probabilities were meaningful. (See Table 3) The majority of both the ROTC cadets and 30th Engineer Battalion officers felt they were meaningful, 71% and 60%, respectively. However, only 20% and SAW students agreed while the other 80% were uncertain. Only 6% of the ROTC population and 10% of the Engineer population were convinced that the probabilities were not meaningful.

	UVA ROTC	30th EN BN	USMC SAW	TOTAL
Strongly Agree	3	5		8
Agree	8	7	2	17
Neutral	5	6	8	19
Disagree	1	2		3
Strongly Disagree				

Table 3: RAW probabilities are meaningful

In all three experiments, the majority of military planners stated that RAW was more likely that the doctrinal war-game to result in a robust COA. Seventy-six percent of the ROTC cadets, 55% of the 30th Engineer Battalion officers, and 70% of SAW students agreed or strongly agreed. Only 6% of the ROTC population, 10% of the Engineer population, and 10% of the SAW population disagreed.

	UVA ROTC	30th EN BN	USMC SAW	TOTAL
Strongly Agree	4	3	1	8
Agree	9	8	6	23
Neutral	3	7	2	12
Disagree	1	2	1	4
Strongly Disagree				

Table 4: RAW is more likely to result in a robust COA

In summary, a thorough risk assessment is essential for effective risk management. During MDMP COAA, the blue force actions are “fixed” while the red force actions are “uncertain.” Because the doctrinal war-game is deterministic, it cannot capture this uncertainty. If it cannot capture this uncertainty, the MDMP COAA cannot result in a thorough risk assessment. If the commander and staff do not thoroughly assess risk, the commander cannot effectively manage risk.

RAW provides the framework for a thorough, quantitative risk assessment whereas the doctrinal war-game does not. With a thorough, quantitative risk assessment, the commander and staff can effectively manage risk within a COA. With a thorough, quantitative risk assessment, the commander and staff can perform risk versus reward trade-off analysis between COAs.

The results from three experiments clearly demonstrate that the military planners surveyed perceived that RAW is both practical and of greater utility than the doctrinal war-game. Both inexperienced and experienced military planners believed that RAW: was practical in a time constrained, deliberate planning environment; adequately captured the uncertainty in the threat's reaction, and is more likely to result in a robust COA. Though few disagreed, a number of respondents, especially the experienced military planners, did express that they were unsure whether or not the probabilities were meaningful. However, we believe that this perception will change as military planners become more comfortable with the probabilities, given adequate training and education.

The commander is responsible for effectively managing risk in any military operation. RAW facilitates effective risk management whereas the doctrinal war-game does not. We believe we have provided sufficient evidence through designed experiments with novice and experienced military planners to encourage the U.S. Military to explore the feasibility of upgrading the doctrinal war-game to provide a systemic framework for a thorough, quantitative risk assessment. RAW provides such a framework.

Bibliography

Armstrong, J. S. (1987). Forecasting Methods for Conflict Situations, John Wiley & Sons, Ltd.

Armstrong, J. S. (2001). Principles of forecasting : a handbook for researchers and practitioners. Boston ; London, Kluwer Academic.

Armstrong, J. S. (2001). Role playing: a method to forecast decisions. Principles of forecasting : a handbook for researchers and practitioners. J. S. Armstrong. Boston ; London, Kluwer Academic**:** 15-30.

Armstrong, J. S. (2002). "Assessing game theory, role playing, and unaided judgment." International Journal of Forecasting **18**(3): 345-352.

Bernardo, J. M. and A. F. M. Smith (1994). Bayesian theory. Chichester ; New York, Wiley.

CALL (2003). RFI 03-0714090636a. K. Miller, Center for Army Lessons Learned.

Cooper, R. (1977). Shadow Jury used by IBM in Big Anti-Trust Case. Wall Street Journal. New York.

Gerbasi, K. C., M. Zuckerman, et al. (1977). "Justice Needs a New Blindfold: A Review of Mock Jury Research." Psychological Bulletin(84): 323-345.

Golhamer, H. and H. Speier (1959). "Some Observations on Political Gaming." World Politics **12**: 71-83.

Green, K. (2003). Conflict Forecasting. **2003**.

Green, K. C. (2002). "Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgment." International Journal of Forecasting **18**(3): 321-344.

Haimes, Y. Y. (1998). Risk modeling, assessment, and management. New York, Wiley.

Halberstam, D. (2001). The best and the brightest. New York, Modern Library.

Kaplan, S. and B. J. Garrick (1981). "On the quantitative definition of risk." Risk Analysis **1**(1): 11-27.

Kaste, R., B. Bodt, et al. (2000). Research in Modeling and Simulation for Command and Control. Aberdeen Proving Ground, MD, Army Research Laboratory.

Kerr, N. L., D. R. Nerenz, et al. (1979). "Role-Playing and the Study of Jury Behavior." Sociological Methods and Research(7): 337-355.

Lowrance, W. W. (1976). Of acceptable risk : science and the determination of safety. Los Altos, Calif., W. Kaufmann.

Mandel, R. (1977). "Political Gaming and Foreign Policy Making During Crises." World Politics(29): 610-625.

Markowitz, H. (1987). Mean-variance analysis in portfolio choice and capital markets. New York, B. Blackwell.

Mixon, D. (1972). "Instead of Deception." Journal of the Theory of Social Behavior(2): 145-177.

Monti, S. and G. Carenini (1999). "Dealing with the Expert Inconsistency in Probability Elicitation." IEEE Transactions on Knowledge and Data Engineering **12**(4): 499-508.

Orne, M. T., P. W. Sheehan, et al. (1968). "Occurrence of Post-Hypnotic Behavior Outside the Experimental Setting." Journal of Personality and Social Psychology(9): 189-196.

Reimer, D. J. (1995). Risk Mangement.

School of Advanced Warfighting (2004). School of Advanced Warfighting.

U. S. Army (1994). FM 34-130: Intelligence Preparation of the Battlefield. Washington, DC, Headquarters Department of the Army.

U. S. Army (1997). FM 101-5: Staff Organization and Operation. Washington, DC, Headquarters Department of the Army.

Walonick, D. S. (2003). Survival Statistics. Minneapolis, MN, StatPac, Inc.

Zimbardo, P. (1972). "The Pathology of Imprisonment." Society(9): 4-8.