# Temporal Updates with Decision Theory for Military Decision Making with Risk Considerations

*Jeremy D. Jordan, PhD*

*Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, Ohio, 45433, USA, +1 937-255-3636, jeremy.jordan@afit.edu*

*Sharif H. Melouk, PhD*

*Department of Information Systems, Statistics, and Management Science, Box 870226, The University of Alabama, Tuscaloosa, AL 35487, USA, smelouk@cba.ua.edu*

*Marcus B. Perry, PhD*

*Department of Information Systems, Statistics, and Management Science, Box 870226, The University of Alabama, Tuscaloosa, AL 35487, USA, mperry@cba.ua.edu*

**Abstract**

We introduce a new decision making criterion using decision theory, where optimal decision policies update as new information becomes available. The approach captures change in a perceived optimal strategy of a decision-maker based on the information known at the time of the decision. Furthermore, we incorporate a utility function to model the different risk behaviors of a decision-maker which allows for the examination of optimal decision-making while accounting for different risk strategies. The new decision criterion are then applied to a military decision making process providing further contribution through visualization of the effects of risk behavior. The techniques presented can be utilized in a precursory analysis to forecast different decisions a soldier or decision maker may encounter, during an engagement, or in an *a posteriori* analysis to determine the effectiveness of the actions taken. The procedures are easily transitioned to assist military leaders in making better decisions quickly through quantitative modeling.

# 1 Introduction

In this manuscript, a method to update optimal decisions, as new information becomes available and measure the difference between the optimal and perceived optimal decisions during a decision-making process is proposed. A natural application of the proposed model is implementation into military combat simulation models or for use in decision-making during combat operations. Consider a representative combat scenario where a tank is engaged in combat, is receiving inputs from its sensors as to the conditions of the overall surrounding operating environment (i.e., nature), observes an unknown object in the distance and must decide on an action to take. Initially, the tank sensors indicate that the object could be an enemy tank, enemy armored personnel carrier (APC), or a friendly tank. Thus, the tank commander must choose whether to shoot at the object or investigate the situation further. Depending on the mission, the tank personnel may choose to do one or the other. After an initial decision is made, the tank sensors give updated information that the object is an enemy tank or an enemy APC. The optimal decision of the tank may then be to shoot at

the object. However, if the true identity of the object is of a friendly nature, the perceived optimal decision to shoot differs from the true optimal decision which may be to advance. The difference between these two decisions can be thought of as the regret of the decision-maker. This type of combat scenario is quite common, and analysis of the decision-making in this scenario is of great importance. This research suggests an improved, quantitative modeling approach (as compared to learning or case-based approaches) that can be employed during a live action, decision-making situation or implemented in a simulation training model.

# 2 Game Theory

This paper expands the area of applied game theoretic work with a unique contribution in the context of military operations. There are many facets of game theory, for general terminology and a thorough overview of game theory, please refer to [1], [3], and [4]. In this paper, we assume a 2-player game with simultaneous play, perfect information, and a zero-sum structure.

Initially, all of the actions possibly performed by each player during a game are identified. The set of actions is denoted $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ for Player 1 and $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_n\}$ for Player 2. All possible combinations of the action set result in the reward matrix ($\mathbf{R}$).

Table 1: Normal Form of a Game

| | $\beta_1$ | $\beta_2$ | $\ldots$ | $\beta_{n-1}$ | $\beta_n$ |
|---|---|---|---|---|---|
| $\alpha_1$ | $r_{11}$ | $r_{12}$ | $\ldots$ | $r_{1,n-1}$ | $r_{1,n}$ |
| $\alpha_2$ | $r_{21}$ | $r_{22}$ | | $r_{2,n-1}$ | $r_{2,n}$ |
| $\vdots$ | $\vdots$ | | $\ddots$ | | $\vdots$ |
| $\alpha_{m-1}$ | $r_{m-1,1}$ | $r_{m-1,2}$ | | $r_{m-1,n-1}$ | $r_{m-1,n}$ |
| $\alpha_m$ | $r_{m1}$ | $r_{m2}$ | $\ldots$ | $r_{m,n-1}$ | $r_{m,n}$ |

Using the reward matrix and action sets, the normal form of the game is developed as shown in Table 1. Initially the normal form is examined to determine whether a pure strategy emerges for each player. If a pure strategy exists, this is referred to as a saddle point, or Nash Equilibrium. The condition that must hold (in the normal form) for a saddle point to exist is the maximum of the row minimums must equal the minimum of the column maximums, or:

$$\max_j \min_i \{r_{ij}\} = \min_i \max_j \{r_{ij}\}$$

If a saddle point is not present, the game is formulated as a linear program and solved to determine the optimal mixed strategy for each player of the game. In particular, the primal problem is used to determine the optimal strategy of Player 1, whereas, the dual problem is used to determine the optimal strategy of Player 2. This technique is employed due to the assumption of the minimax/maximin principle. Formulation 1 shows the primal optimization

problem.

$$\begin{aligned}
\max z = \nu + 0w_1 + 0w_2 \ldots + 0w_m \\
s.t.\ \nu \leq r_{11}w_1 + r_{21}w_2 + \cdots + r_{m1}w_m \\
\nu \leq r_{12}w_1 + r_{22}w_2 + \cdots + r_{m2}w_m \\
\vdots \\
\nu \leq r_{1n}w_1 + r_{2n}w_2 + \cdots + r_{mn}w_m \\
\sum_i w_i = 1; \forall i = 1, \ldots, m \\
w_i \geq 0; \forall i = 1, \ldots, m
\end{aligned} \tag{1}$$

In this formulation, the objective is to maximize $\nu$, the value of the game with respect to player one, by determining its optimal mixed strategy. Player 1s mixed strategy, $\gamma = \{w_1, w_2, \ldots, w_m\}$ denotes a probability distribution assigned to the set of actions $\boldsymbol{\alpha}$. The constraints represent player 1s expected reward given that the player chooses one of its n actions. The optimal solution to the linear program corresponding to the primal formulation is easily found via any of a number of commercially available solvers. Similarly, to compute the mixed strategy, $\delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$ of the action set $\boldsymbol{\beta}$ of Player 2, the dual problem is formulated. Formulation 2 shows the dual optimization problem.

$$\begin{aligned}
\min z = \omega + 0\delta_1 + 0\delta_2 \ldots + 0\delta_n \\
s.t.\ \omega \geq r_{11}\delta_1 + r_{12}\delta_1 + \cdots + r_{1n}\delta_n \\
\omega \geq r_{21}\delta_2 + r_{22}\delta_2 + \cdots + r_{2n}\delta_n \\
\vdots \\
\omega \geq r_{m1}\delta_1 + r_{m2}\delta_2 + \cdots + r_{mn}\delta_n \\
\sum_j \delta_j = 1; \forall j = 1, \ldots, n \\
\delta_j \geq 0; \forall j = 1, \ldots, n
\end{aligned} \tag{2}$$

Formally stated, a pure strategy occurs when the strategy for the action sets $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ have the following properties: $w_i = 1$ for 1 of $m$ actions and $w_i$=0 for all remaining $i$'s or $\delta_j = 1$ for 1 of $\delta$ actions and $\delta_j = 0\ \forall$ remaining $j$'s, respectively. A mixed strategy occurs when action sets $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are assigned a probability distribution over the actions such that the assigned probability does not represent a pure strategy; that is, the probability of any particular action is strictly less than one.

## 2.1 Value of the Game

The players' mixed strategies result in a floor value for Player 1 for this original formulation, denoted as $\pi$. In other words, Player 1 is guaranteed to receive no less than $\pi$ if mixed

strategy $\gamma$ is played. The variable $\pi$ is also the ceiling value for Player 2, guaranteeing this player from losing no more than $\pi$ by playing mixed strategy $\delta$. When the value of the game to each of the players is equal, a Nash equilibrium occurs. Consequently, any mixed strategy that results in equal values of $\pi$ meets this criteria and is considered an optimal strategy. It is important to note that $\pi$ is the value to each of the players over time. At each play of a game, there will be some variation from $\pi$. However, as time approaches infinity, each player can expect its reward to approach $\pi$.

The common value of the game, $\pi$, for each player is actually the solution to their respective linear programs, $\pi = \nu = \omega$. However, since the reward matrix $\mathbf{R}$ will be manipulated in upcoming sections to account for player risk preferences, the value of the game is reported as

$$\pi = \gamma \, \mathbf{R} \, \delta' \tag{3}$$

Note that, the $\delta$ used for nature, in the one-player versus nature case, is a uniform distribution across $\beta$, or nature's action set. This is a reasonable assumption since information is unavailable regarding the likelihood of nature's strategy. Thus, it is reasonable to assume that each of the actions of nature are equally likely to occur.

# 3 Proposed Model

The optimal strategy will change, or update, depending on the data available about the action sets or strategy of the game players. An updated strategy is actually a perceived strategy since it is dependent on a players perception of what actions are available to the other player. We denote a perceived strategy as $\hat{\gamma}$. Thus, the strategy $\gamma$ for the action set $\boldsymbol{\alpha}$ at time step $s$ is dependent on a player's perception of what actions are available to the other player. That is,

$$\hat{\gamma}^{(0)} = \gamma | \hat{\beta}^{(0)}$$

for $s = 0$ at the start of the game. This shows that the strategy of Player 1, $\gamma$, is based on his perception of the actions available to Player 2, which is all of the possible actions of Player 2 initially. In general,

$$\hat{\gamma}^{(s)} = \gamma | \hat{\beta}^{(s)} \tag{4}$$

Note that Player 1 can only perceive as many actions of which Player 2 is capable. This continues up to time step $q$, the number of time steps in the game. As $\gamma$ updates, it is dependent on data obtained from some information source, or combination of sources, that is perceiving the situation, or information about $\boldsymbol{\beta}$. Thus $\hat{\boldsymbol{\beta}}$ is dependent on

$$\boldsymbol{\zeta} = \{\zeta_1 \cap \zeta_2 \cap \ldots \cap \zeta_k\},$$

where $\zeta_i$ is source $i$ of $k$ number of sources. Thus,

$$\hat{\beta}^{(s)} = \beta | \boldsymbol{\zeta}^{(s)} \tag{5}$$

where $\boldsymbol{\zeta}^{(s)}$ is the set of information sources available at time $s$. Thus,

$$\hat{\gamma}^{(s)} = [\gamma | \{\hat{\beta}^{(s)} | \boldsymbol{\zeta}^{(s)}\}] \tag{6}$$

shows that the strategy of Player 1 is dependent on the information received from the source, or sensor, and its perception of the action set of Player 2.

The reward matrix $\mathbf{R}$ and strategy of Player 2, $\delta$, will update at each time step as well, depending on the player's perceptions of the action sets available to the other players, denoted $\hat{R}^{(s)}$ and $\hat{\delta}^{(s)}$, respectively.

## 3.1 Optimal vs. Perceived Optimal Decisions

During a game, a player may not have true information about the set of available actions to the opponent. The mixed strategy may not be the proper mixed strategy to use since it may be based on false information, making it a perceived optimal mixed strategy (i.e., $\hat{\gamma}^{(s)}$). The true optimal mixed strategy based on perfect information may then be used textita posteriori to determine how poor this perceived mixed strategy is.

The true optimal strategy is thus denoted $\gamma^{(s)}$. Similarly, the true reward matrix is $\mathbf{R^{(s)}}$ and the true strategy of Player 2 is $\delta^{(s)}$. If the information received via sensors or some other source, $\boldsymbol{\zeta}$, is less than perfect, a difference will occur between the optimal strategy and the perceived optimal strategy. That is, the quality of $\hat{\gamma}^{(s)}$ is less than that of $\gamma^{(s)}$. The magnitude of the difference between the perceived optimal and true optimal strategies can be measured by using the value of the game, $\pi$, as a comparison measure between strategies. The value corresponding to the perceived optimal strategy, $\hat{\pi}^{(s)}$, is calculated using the perceived optimal strategy, the true reward matrix, and the true strategy of Player 2, as shown in (7)

$$\hat{\pi}^{(s)} = \hat{\gamma}^{(s)}\, \mathbf{R^{(s)}}\, \delta^{(\mathbf{s})'}. \tag{7}$$

The value of the true optimal strategy $\pi^{(s)}$ is similarly calculated as shown in 8

$$\pi^{(s)} = \gamma^{(s)}\, \mathbf{R^{(s)}}\, \delta^{(\mathbf{s})'}. \tag{8}$$

The difference between the values of the two strategies is

$$\vec{\pi}^{(s)} = \pi^{(s)} - \hat{\pi}^{(s)}. \tag{9}$$

The value of $\vec{\pi}^{(s)}$ will change as the game progresses. Initially, the difference between the value of the game of the perceived and the true optimal decision is $\vec{\pi}^{(0)} = 0$. The variable, $\vec{\pi}^{(s)}$, can be used to determine the value of obtaining perfect information and also to measure the value of the information obtained from sources $\boldsymbol{\zeta} = \zeta_1, \zeta_2, \ldots, \zeta_k$. Note, this value may or may not be representative of the actual value because of the nature of the measure. For example, in a ranking system, $\vec{\pi}^{(s)}$ will provide a frame of reference for which two different decisions can be compared and/or the value of the source can be observed over time.

## 3.2 Determining User Preferences via a Utility Function

Traditional zero-sum game theory assumes that each player will approach the game in an identical fashion, however, this is not always the case. While it is true that each player will attempt to maximize its minimum gain and minimize the maximum gain of the other players,

the values of the reward matrix **R** will not always reflect the true value of the reward to each player. This difference in value is accounted for by determining the players risk preference and changing the values in the reward matrix to reflect this preference. We use utility theory to represent the reward value a situation has to each player. In other words, the true values of the situations to the players in some cases, when all other influencing factors are considered, are different than the general reward matrix. The original reward matrix produces strategies that can be thought of as the expected case, or the strategy a rational decision-maker would employ. So, the use of a utility function allows for a plethora of decision-makers to be represented based on their individual risk taking preference. That is, a decision-maker may be risk averse, risk neutral, or risk prone. The risk averse individual avoids risk more so than the risk neutral individual in the expected case. The risk prone individual approaches situations with great risk in comparison with the risk neutral individual. A risky player attempts to maximize payoff regardless of the chance for loss. The risk neutral individual approaches the situation as the average rational individual would by trying to maximize minimum payoff for the given reward matrix.

A utility matrix is conjectured using a utility function to alter the original payoff values of the reward matrix. We use an exponential utility function to transform the original reward matrix using a parameter, $\rho$, which accounts for the players' risk behavior. From [2], for the monotonically increasing measure,

$$u(r_{ij}) = \frac{1 - exp[-(r_{ij} - Low)/\rho]}{1 - exp[-(High - Low)/\rho]}\,, \tag{10}$$

where $u(r_{ij})$ denotes the utility of the $(ij)^{th}$ element of the reward matrix, *Low* is the lowest level of the measure $m$, *High* is the highest level of $m$, and $\rho$ is the exponential constant for the value function (i.e., the risk tolerance of a player). Recall, $m$ is the measure used to compile the original reward matrix and is the basis for computing the expected case strategy. The function in (10) alters this original measure to produce 'new' values in the reward matrix based on the risk preference of the decision-maker.

As the game updates, the risk behavior of the players will evolve. A player may initially approach the game with a risk averse attitude, then transition to a risk prone attitude as the game progresses. In a one player versus nature scenario, only the risk attitude of Player 1 needs to be considered, and the results are only dependent on this risk attitude. In (11), we calculate the value of the game for different risk tolerances of the players.

$$\hat{\pi}^{(s)}_{\rho_1(\rho_2)} = \hat{\gamma}^{(s)}_{\rho_1}\, R^{(s)}\, \delta^{(s)'}_{\rho_2}. \tag{11}$$

In general, $\rho \to 0^+$ indicates a more risk averse behavior, whereas $\rho \to 0^-$ indicates a more risk prone behavior. Risk neutral behavior is represented by $\rho = \infty$ or $\rho = -\infty$. The exponential utility function is undefined at $\rho = 0$

In the one-player versus nature game, the effects of the risk behavior of Player 1 is based on its risk preference and the likely outcomes of nature. If information is unknown regarding the probable outcomes of the situation being faced, we assume a uniform distribution across the possible actions of nature. If we have prior knowledge of the likely probabilities of the outcomes of nature, we can adjust our risk preferences accordingly. Understanding the

effects of risk behavior is thus accomplished through adjusting the probabilities of natures outcome and then enumerating the solutions across levels of $\rho$ for player 1. This analysis can be done before engagement, in battle, or as a post-battle analysis to determine better risk preferences in similar future situations. In general, as one becomes more risk prone, the chance of greater gain increases as the does the chance of greater loss. A risk neutral approach generally produces a certain value with less variation. The question of interest then becomes, what is the optimal risk preference to assume given knowledge of the situation? To address this in future research, we plan to adopt robust parameter design principles to minimize the variability in the value of the game across levels of $\rho$, subject to some desired mean that must be achieved. Alternatively, the mean can be maximized subject to a desired limit on the amount of variation one is willing to accept. A decision maker can then examine the effects of risk behavior *a posteriori* to inform future decision making.

# 4 Conclusion

In this paper, we presented a game theoretic framework with which to approach situations where updating information would be beneficial to decision-makers. Although the methodology presented in this paper is applicable in various fields of study, many assumptions need to be addressed in order to ensure a wider application use. Military situations, any naturally arising two player games (e.g., sports games), strategic games, and proper allocation of resources are a few simple examples that can be modeled using the proposed methodology. Updating the optimal decision based on the reception of new information occurs in a broad range of areas, and this research opens a number of possible threads.

# References


[1] Prajit K. Dutta. *Strategies and games: theory and practice*. MIT press, 1999.

[2] Craig W Kirkwood. *Strategic decision making: multiobjective decision analysis with spreadsheets*, volume 59. Duxbury Press Belmont, CA, 1997.

[3] Martin J Osborne and Ariel Rubinstein. *A course in game theory*. MIT press, 1994.

[4] Wayne L Winston and Jeffrey B Goldberg. *Operations research: applications and algorithms*, volume 3. Thomson Brooks/Cole Belmont, 2004.