Behavioral Modeling of Attackers Based on Prospect Theory and Corresponding Defenders Strategy

Ziyi Chen, Chunqi Wan, Bingfeng Ge, Yajie Dou(✉), and Yuejin Tan

College of Systems Engineering, National University of Defense Technology, Changsha, Hunan 410073, China
chenziyi13@nudt.edu.cn, yajiedou_nudt@163.com

Abstract. Many methods focused on describing the attackers' behavior while ignoring defenders' actions. Classical game-theoretic models assume that attackers maximize their utility, but experimental studies show that often this is not the case. In addition to expected utility maximization, decision-makers also consider loss of aversion or likelihood insensitivity. Improved game-theoretic models can consider the attackers' adaptation to defenders' decisions, but few useful advice or enlightenments have been given to defenders. In this article, in order to analyze from defenders' perspective, current decision-making methods are augmented with prospect theory results so that the attackers' decisions can be described under different values of loss aversion and likelihood insensitivity. The effects of the modified method and the consideration of upgrading the defense system are studied via simulation. Based on the simulation results, we arrive at a conclusion that the defenders' optimal decision is sensitive to the attackers' levels of loss aversion and likelihood insensitivity.

Keywords: Prospect theory · Decision analysis · Game theory

1 Introduction

Classic game-theoretic models often assume that attackers and defenders are rational in the sense of the von Neumann-Morgenstern expected utility theory [1]. Methods proposed by Parnell et al. [2] and Rios Insua et al. [3] combine decision analysis and game theory to allow some deviations from the rationality. However, Simon [4] and Kahneman and Tversky [5], have already show that decision-makers are not strictly rational when they engage in unaided decisions. Similarly, players in games deviate from the rationality principles in such ways such as being averse to losses, having diminishing sensitivity, being dependent on reference points, and distorting probability when they face uncertain outcomes. Therefore, descriptive models have been proposed predict decision-makers' actual behavior. Merrick and Parnell [6] give a review of methods used in attacker/defender models to represent the adaptation of the attacker to the defender's decisions. In addition to traditional decision analysis, game-theoretic models and hybrid methods, prospect theory has been widely applied [7, 8]; for example, Mazicioglu and Merrick [9] represented adversaries with multiple objectives

Y. Chen et al. (Eds.): GDN 2018, LNBIP 315, pp. 179–189, 2018.
https://doi.org/10.1007/978-3-319-92874-6_14

in counter-terrorism using multi-attribute prospect theory. Edouard [10] modeled terrorists' behavior with modified decision weights and proposed the strategic logit risk analysis (SLRA) method to solve allocation of scarce defense resources. Merrick [11] modified existing methods used in counter-terrorism decisions, in particular, the attacker's decision problem, with a descriptive model that accounts for the attacker's loss aversion.

An effective confrontation decision framework incorporates the defender's decisions and the attacker's adaptation to them. The literatures mentioned above consider using improved model to describe the behavior of the attacker more realistically, but the people's research is to provide advice to the defender, so we should consider how the defender can arrive the optimal result by changing his behavior. Ideally, this article introduces a "Wait and see" region through a random upgrade approach, which means that defender can actively change their behavior to achieve optimal results based on the type of the attacker's behavior.

The organization of this paper is organized as follows. In the next section, a brief introduction of prospect theory is presented. A comprehensive description of the model is given in Sect. 3. In order to provide advice to the defender the impact of changes in the behavior of the defender, confidential information, and random promotion is considered in Sect. 4. Finally, conclusions and suggestions are given in Sect. 5.

2 Brief Introduction of Prospect Theory

It is a well-known fact that human beings are not purely rational (and not entirely irrational or arbitrary) without the use of decision analysis methods. Therefore, we extend the method of Parnell et al. [2] and Rios Insua et al. [3] to the descriptive model of attackers' decision making so that attackers follow rationality in decision making. First, we will use prospect theory to model attackers' decisions behavior. Prospect theory has been shown to represent loss aversion, framing effects, deterministic effects, and likelihood insensitivity. A prospect is a gamble that has a probability p to get x and a probability $1 - p$ to get 0. According to original prospect theory [4], we set this prospect as $\pi(p)v(x)$, where v is a S-shaped value function, with its inflection point(the reference point) at 0. Thus, it represents diminishing sensitivity and reference dependence. It is also defined that $v(x) - v(0) < v(0) - v(-x)$ represents loss aversion. The probability weighting function π also includes reference dependence for $p = 0$ and $p = 1$, diminishing sensitivity away from these references. Function $v(x)$ is concave for low probability values and convex for high values.

For the more complex situation with more than two outcomes, the original prospect theory cannot obey stochastic dominance. Tversky and Kahneman [8] introduced the now standard form of prospect theory to overcome this shortcoming. They placed the weights on the cumulative probabilities, or ranked probabilities. To apply this form of prospect theory to a prospect X with n outcomes, we first order the outcomes $x_1 \geq \ldots \geq x_k \geq 0 \geq x_{k+1} \geq \ldots \geq x_n$ along with their respective probabilities $p_1, \ldots, p^n$. Note that there are k outcomes which are gains and $n - k$ outcomes which are losses. If $k = 0$, then all outcomes are losses; if $k = n$, then all outcomes are gains. In the improved form of prospect theory, weights are applied to cumulative probabilities:

$$w^{+}(x_j) = \pi(P(X \geq x_j)) - \pi(P(X > x_j)) \quad (1)$$

for gains $(j \leq k)$, and

$$w^{-}(x_j) = \pi(P(X \leq x_j)) - \pi(P(X < x_j)) \quad (2)$$

for losses $(j \leq k)$. Thus, the value of a prospect is:

$$\sum_{j=1}^{k} w^{+}(x_j)v(x_j) + \sum_{j=k+1}^{n} w^{-}(x_j)v(x_j). \quad (3)$$

In the following example, we will use this prospect theory.

3 Behavioral Modeling of Attackers

A-type missile is a new type of weapon with a nuclear warhead, which can effectively break through the defenses of the current conventional defense systems. C-type defense system is designed specifically for A-type missile, which can effectively intercept its attacks. We use d1 represent the defender decided to upgrade to the C-type defense system, and d2 for not upgrade. The attacker decision represents the choice of attacks using the A-type missile, a1, and attacks using a conventional weapon (not missile), a2. The loss caused by the A-type missile attack is assumed to be 40 billion dollars; parameter r is used to represent the loss. The loss caused by a conventional weapon attack is assumed to be a quarter of this amount.

The cost of upgrading the defense system is assumed to be \$100 million. We assume that the attackers want to inflict the largest damage on the defender, but they also suffer losses if the defenders' defense systems intercept their attacks. For illustrative purposes, this loss, f, is assumed to be one-tenth of the value of the successful A-type missile attack. The probability p of successful interception of a Type A missile by the C-type defense system is assumed at 0.8. Furthermore, if the A-type missile successfully breaks through the intercept, that the probability of successful attack is assumed to be $q = 0.5$. If the defenders are not equipped with a C-type defense system, the original defensive forces have a probability S of a successful intercept, which is assumed to be 0.3. As for the conventional weapon, the probability of successful attack is $q = 0.5$.

The attacker and defender decisions are represented as decision trees in Fig. 1. The optimal decision for the defender is simple to solve as if we know the attacker's decision.

We assume that the attacker can observe the defender's choice. If the defender does not upgrade the C-type defensive system, the attacker selects the A-type missile, and if the defender upgrades the C-type defensive system then the attacker switches to conventional weapons, then we consider that the attacker was deterred by the defender's upgrade strategy, upgrade C-type defense system is the optimal strategy. We also think that the upgrade decision is dictated by the preference of the attacker. In the

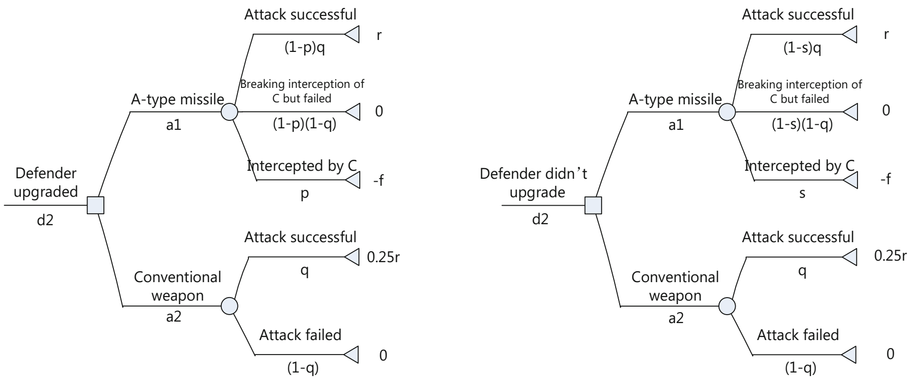

Fig. 1. The attacker decision tree if the defenders upgraded (left) and if the defender did not upgrade (right) their defenses.

following, we study the decisions of the attacker and the effect of loss aversion and likelihood insensitivity to it.

We must choose an appropriate form for the value function v and the probability weighting function π. We will choose parameterized forms to demonstrate the effects of the magnitude of loss aversion and likelihood insensitivity. We choose the simplest parametric forms of each for illustrative purposes. Tversky and Kahneman [8] introduced a value function that represents loss aversion, written as:

$$v(x)=\begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\beta} & x < 0 \end{cases} \tag{4}$$

where $\alpha, \beta, \lambda \geq 1$. This form represents loss aversion by the increased steepness for negative values of x.

If $\alpha = \beta = 1$ then the attacker is risk neutral for prospects involving only gains and for prospects involving only losses; otherwise the value function is S shaped. This form represents no loss aversion when λ = 1, with the level of loss aversion increasing in λ. Chateauneuf et al. [12] discuss additive probability weighting function, defined by:

$$\pi(p) = \begin{cases} 1 & p = 1 \\ \kappa p + \frac{1}{2}(1-\kappa) & 0 < p < 1 \\ 0 & p = 0 \end{cases} \tag{5}$$

where $0 < \kappa \leq 1$. This form can represent no probability weighting when $\kappa = 1$.

Tversky and Kahneman [8] experimentally estimated λ = 2.25. Novemsky and Kahneman [13] found individual values of λ to vary between 1 and 3, with most people's behavior corresponding to values between 2 and 3. Baillon et al. [14] finds likelihood insensitivity for the new additive weighting function equivalent to values of κ between 0.6 and 0.8. Thus, in the following the effects of loss aversion and likelihood insensitivity are tested for λ between 1 and 3 and κ between 0 and 1. The effect of adding risk attitude is tested for values of 1 (piece-wise linear value function) and 0.5

(S-shaped value function). Note that these values are estimated from experiments with a general population and are not specific to terrorist choice under uncertainty.

When the defender upgraded the C-type defensive system, a prospect-theoretic the attacker will choose the A-type missile if:

$$\begin{aligned}&(\pi((1-p)q)-\pi(0))v(r)+(\pi((1-p)(1-q)+(1-p)q)-\pi((1-p)q))v(0)\\&+(\pi(p)-\pi(0))v(-f)\geq(\pi(q)-\pi(0))v(\frac{1}{4}r)+(\pi((1-q)+q)-\pi(q))v(0)\end{aligned} \tag{6}$$

Using the $\pi(0)=0$ and $v(0)=0$, we simplify it to:

$$\pi((1-p)q)v(r)+\pi(p)v(-f)\geq\pi(q)v(\frac{1}{4}r) \tag{7}$$

Substituting the parameters of the value function, the probability weighting function and the value of p,q,r,f mentioned above ($p=0.8,\ q=0.5,\ r=40,\ f=4$), then simplifying, we obtain the following inequality:

$$\begin{aligned}\lambda&<\frac{(\kappa(q-pq)+\frac{1}{2}(1-\kappa))r^{\alpha}-(\kappa q+\frac{1}{2}(1-\kappa))(\frac{1}{4}r)^{\alpha}}{(\kappa p+\frac{1}{2}(1-\kappa))f^{\beta}}\\&=\frac{((1-(\frac{1}{4})^{\alpha})(\kappa q+\frac{1}{2}(1-\kappa))-\kappa pq)r^{\alpha}}{(\kappa p+\frac{1}{2}(1-\kappa))f^{\beta}}\\&=\frac{(\frac{1}{2}(1-(\frac{1}{4})^{\alpha})-\frac{2}{5}\kappa)40^{\alpha}}{(\frac{4}{5}\kappa+\frac{1}{2}(1-\kappa))4^{\beta}}\end{aligned} \tag{8}$$

Similarly, if the defender didn't upgrade (right-hand-side of Fig. 1), setting the parameters of the value function, the probability weighting function and the value of $p,\ q,\ r,\ f,\ s$ mentioned above ($p=0.8,\ q=0.5,\ r=40,\ f=4,\ s=0.3$), a prospect-theoretic attacker will choose the A-type missile if:

$$\begin{aligned}\lambda&<\frac{(\kappa(q-sq)+\frac{1}{2}(1-\kappa))r^{\alpha}-(\kappa q+\frac{1}{2}(1-\kappa))(\frac{1}{4}r)^{\alpha}}{(\kappa s+\frac{1}{2}(1-\kappa))f^{\beta}}\\&=\frac{((1-(\frac{1}{4})^{\alpha})(\kappa q+\frac{1}{2}(1-\kappa))-\kappa sq)r^{\alpha}}{(\kappa s+\frac{1}{2}(1-\kappa))f^{\beta}}\\&=\frac{(\frac{1}{2}(1-(\frac{1}{4})^{\alpha})-\frac{3}{20}\kappa)40^{\alpha}}{(\frac{3}{10}\kappa+\frac{1}{2}(1-\kappa))4^{\beta}}\end{aligned} \tag{9}$$

We set $\alpha=\beta=1$ and $\alpha=\beta=0.5$ to represent two different kinds of risk attitude. If $\alpha=\beta=1$, the inequalities (8) and (9) can be reduced to, respectively:

$$\lambda<\frac{37\frac{1}{2}-40\kappa}{5+3\kappa} \tag{10}$$

and

$$\lambda < \frac{75 - 30\kappa}{20 - 4\kappa} = 7.5 \tag{11}$$

When $\alpha = \beta = 0.5$, inequalities (8) and (9) can be reduced to, respectively:

$$\lambda < \frac{\sqrt{10}(\frac{5}{2} - 4\kappa)}{5 + 3\kappa} \tag{12}$$

and

$$\lambda < \frac{\sqrt{10}(5 - 3\kappa)}{10 - 4\kappa} \tag{13}$$

Figure 2 shows the different the attacker's decisions regions as a strategy plot, setting α and β, and varying λ and κ. Since $p > s$ and $\kappa > 0$, the right side of Inequality (8) is less than the right side of Inequality (9), the attacker prefers A-type missile attack to have higher loss aversion thresholds if the C-type defensive system is upgraded. This means that for the values of λ and κ which Inequality (9) does not hold, Inequality (8) also does not hold. For such values, as shown in the white regions to the right of Fig. 2, the attacker will select the conventional weapon irrespective of the defender's decision. The level of loss aversion in this area means that the possibility of successful interception will reduce the value of A-type missile attack, regardless of whether or not the defender upgraded. However, there is a tradeoff between loss aversion and likelihood insecurity at the borders of the region.

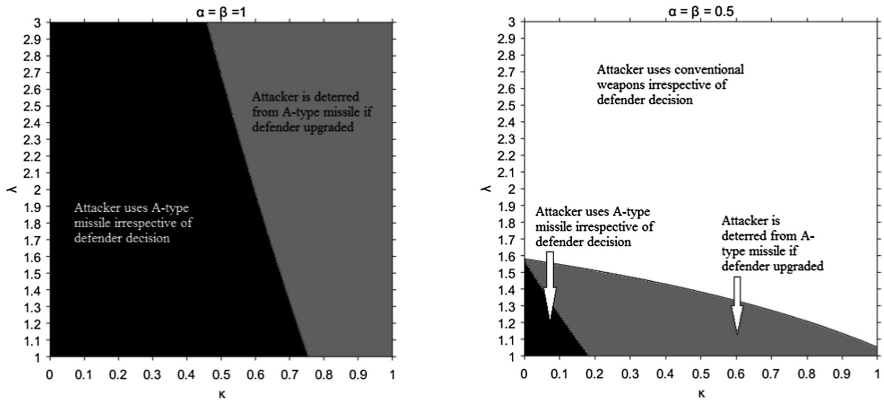

Fig. 2. Strategy plots showing the effect of loss aversion and likelihood insensitivity on the attacker's reaction to the defender's decision.

The left side of Fig. 2 shows the case of $\alpha = \beta = 1$ and the right side shows the case of $\alpha = \beta = 0.5$. For a neutral risk ($\alpha = \beta = 1$), there is no white area, so the attacker either always chooses A-type missile or is deterred by promotion and choose A-type missile only if the defender doesn't upgrade. A risk attitude involving $\alpha = \beta = 0.5$ reduces the area of the black area where the attacker always uses A-type missile and the gray area that the attacker uses A-type missile only if the defender doesn't upgrade; the white area where the attacker always chooses conventional weapons grows accordingly. In the observed regions of the experimental values of λ and κ ($2 \leq \lambda \leq 3$ and $0.6 \leq \kappa \leq 0.8$), the attacker mostly deters by promotion with the neutral risk and chose conventional weapons regardless of promotion when $\alpha = \beta = 0.5$.

Likelihood insensitivity increases the relative value of A-type missile attacks if defender did not upgrade, and the choice of conventional weapons requires more loss aversion. For values of λ and κ where inequality (8) holds, we have that (9) also holds. For such a value (represented as a black area in Fig. 2), the attacker chooses an A-type missile attack, regardless of the defender's decision. For the remaining values of λ and κ, if the defender upgraded, the attacker will select the conventional weapon, and if the defender did not upgrade, the attacker will select the A-type missile attack, so the promotion deters it from the A-type missile attack (shown as the gray area in Fig. 2).

4 Secrecy and Upgrade Randomly

The defense system is not upgraded in all locations. In this section, we consider the security measures, and only select a certain percentage of sites to upgrade. We consider the case when the scale and location of the upgrade are chosen at random and the attackers do not know whether the target is upgraded or not. Figure 3 shows the decision tree of the attacker's decision as the defender randomly upgrades the defense system with probability d.

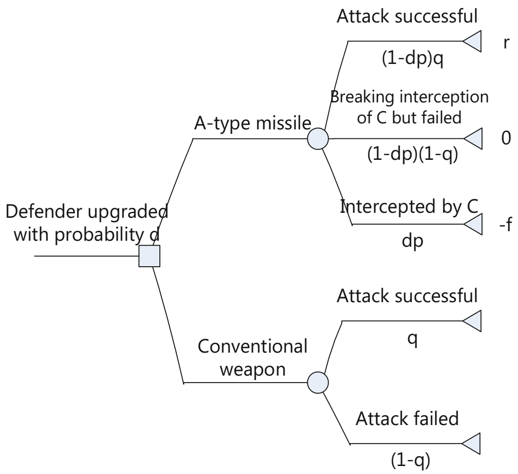

Fig. 3. The attacker decision trees if the defender upgraded with probability d

When the defender upgraded the C-type defensive system with probability d, a prospect-theoretic attacker will choose the A-type missile if:

$$(\pi((1-dp)q)-\pi(0))v(r)+(\pi((1-dp)(1-q)+(1-dp)q)-\pi((1-dp)q))v(0)$$
$$+(\pi(dp)-\pi(0))v(-f)\geq(\pi(q)-\pi(0))v(\frac{1}{4}r)+(\pi((1-q)+q)-\pi(q))v(0) \tag{14}$$

Substituting the parameters of the value function, the probability weighting function and setting $p=0.8, q=0.5, r=40,\ f=4$, we obtain:

$$\begin{aligned}\lambda &< \frac{((1-(\frac{1}{4})^{\alpha})(\kappa q+\frac{1}{2}(1-\kappa))-\kappa dpq)r^{\alpha}}{(\kappa dp+\frac{1}{2}(1-\kappa))f^{\beta}} \\ &= \frac{(\frac{1}{2}(1-(\frac{1}{4})^{\alpha})-\frac{2}{5}d\kappa)40^{\alpha}}{(\frac{4}{5}d\kappa+\frac{1}{2}(1-\kappa))4^{\beta}}\end{aligned} \tag{15}$$

We still consider the two risk attitudes of $\alpha=\beta=1$ and $\alpha=\beta=0.5$, and indicate the region between $d=0$ and $d=1$, which is called the "Waiting and see area" of the attacker.

The "Wait and see" of the attacker is shown as the gray region in Fig. 4, which is a middle area contained by two different values 0 and 1 of the parameter d in Inequality (15), which means when the attacker in this region, his decision-making is infected by the defender's value, if the point of λ and κ corresponding to the attacker is above Inequality (15), then the attacker uses conventional weapons irrespective of defender's decision, on the contrary, the attacker uses A-type missile regardless of the promotion. It is useful that the defender could choose the value of d depend on the attacker's value of λ and κ.

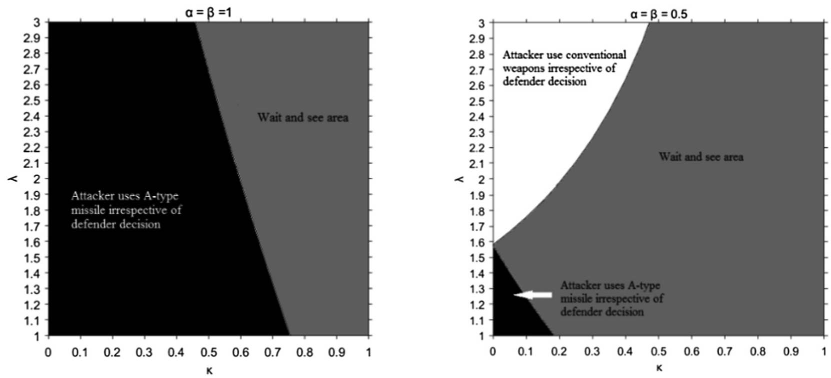

Fig. 4. Strategy plots showing the effect of loss aversion and likelihood insensitivity on the attacker's reaction to the defender decision (Defender upgrade randomly).

The case of an attacker with $\lambda = 1.4$ and $\kappa = 0.3$ is shown in Fig. 5. In this case, the defender can choose the Wait and see area under the point, and force the attacker not to use A-type missile. Thus, understanding the level of loss aversion and likelihood insensitivity of a possible attacker is important to determine the optimal level of d.

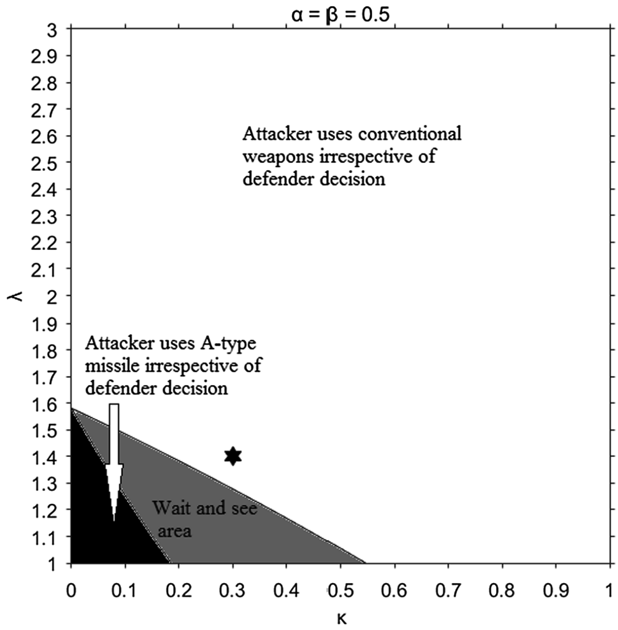

Fig. 5. Strategy plots when $\lambda = 1.4$, $\kappa = 0.3$ and $d = 0.5$

5 Conclusions

In this article, we have reviewed the current methods in confrontation. We chose the prospect theory to describe the loss aversion and likelihood insensitivity in the attacker's decision behavior and provide some suggestions.

We first studied the strategy of whether or not a defender upgraded into a C-type defense system, used prospect theory to build descriptive modeling of the attacker's decisions. When the attacker is at the most common level of loss aversion, likelihood insensitivity and risk attitude, he uses conventional weapons regardless of the promotion. This means that in the face of such opponents, the defender need not to upgrade the defense system. When loss aversion is low, likelihood insensitivity and the risk attitude are neutral. The Attacker uses A-type missile when the defender did not upgrade and conventional weapons if the defender upgraded. So promotion prevents such attackers from attacking to the lower consequences. Therefore, it is important to understand the level of loss aversion, likelihood insensitivity and risk attitude of the potential attackers.

The attackers do not know whether their target was upgraded and upgrading only some of the sites. This approach resulted in a large "Wait and see area", meaning that the defender could take the initiative based on the type of attacker's behavior to upgrade. This shows that understanding the level of loss aversion, likelihood insensitivity and risk attitude is important to defender's decision-making.

There are several open areas for research revealed in this work. Only one evaluation standard for the attacker and defender is considered here. Multiple objectives which represent the attacker can be included. This, however, requires the development of a descriptive model for attacker. This method requires that attributes be ordered according to their importance and considered one at a time. When considering a particular attribute, then alternatives below the expected level are removed. And finally, we get a set from which we can choose a decision that fits our preference.

Note that the results are obtained in one step. What if the attacker considers multiple sequential events in his decision, not only a sum probability? How to apply prospect theory to the decision tree with multiple sequential uncertainty nodes?

Last, prospect theory is not the only descriptive decision model. In fact, there are several alternative models of choice behavior, including rank dependent utility [15, 16] and regret theory [17, 18]. Further, there are behavioral game theory models that represent human behavior in strategic interactions, such as k-level thinking and cognitive hierarchy theory [19]. Thus, our contribution represents a first step in applying descriptive attacker models in confrontation research.

Acknowledgments. We are thankful to the Editors and the reviewers for their valuable comments and detailed suggestions to improve the presentation of the paper. Further, we also acknowledge the support in part by the National Natural Science Foundation of China under Grants No.71690233, NO.71501182.

References

1. Von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1945)
2. Parnell, G.S., Smith, C.M., Moxley, F.I.: Prospect theory intelligent adversary risk analysis: a bioterrorism risk management model. Risk Anal. **30**(1), 32–48 (2010)
3. Rios Insua, D., Rios, J., Banks, D.: Adversarial risk analysis. J. Am. Stat. Assoc. **104**(486), 841–854 (2009)
4. Simon, H.A.: A behavioral model of rational choice. Quart. J. Econ. **69**(1), 99–118 (1955)
5. Kahneman, D., Tversky, A.: Judgment under uncertainty: heuristics and biases. Science **185** (4157), 1124–1131 (1974)
6. Merrick, J., Parnell, G.S.: A comparative analysis of PRA and intelligent adversary methods for counter-terrorism risk management. Risk Anal. **31**(9), 1488–1510 (2011)
7. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**, 263–292 (1979)
8. Tversky, A., Kahneman, D.: Advances in prospect theory: cumulative representation of uncertainty. J. Risk Uncertain. **5**(4), 297–323 (1992)
9. Mazicioglu, D., Jrw, M.: Behavioral modeling of adversaries with multiple objectives in counter-terrorism. Risk Anal. (2) (2017)

10. Kujawski, E.: Accounting for terrorist behavior in allocating defensive counter-terrorism resources. Syst. Eng. **18**(4), 365–376 (2015)
11. Merrick, J.R., Leclerc, P.: Modeling adversaries in counter-terrorism decisions using prospect theory. Risk Anal. **36**(4), 681–693 (2016)
12. Chateauneuf, A., Eichberger, J., Grant, S.: Choice under uncertainty with the best and worst in mind: neo-additive capacities. J. Econ. Theory **137**, 538–567 (2007)
13. Novemsky, N., Kahneman, D.: The boundaries of loss aversion. J. Mark. Res. **42**, 119–128 (2005)
14. Baillon, A., Bleichrodt, H., Keskina, U., L'Haridonb, O., Lia, C.: Learning under Ambiguity: An Experiment Using Initial Public Offerings on a Stock Market. CREM (2014). http://EconPapers.repec.org/RePEc:tut:cremwp:201331
15. Quiggin, J.: A theory of anticipated utility. J. Econ. Behav. Organ. **3**(4), 323–343 (1982)
16. Quiggin, J.: Generalized Expected Utility Theory: The Rank-Dependent Model. Kluwer Academic Publishers, Boston (1993)
17. Bell, D.E.: Regret in decision making under uncertainty. Oper. Res. **30**(5), 960–981 (1982)
18. Loomes, G., Sugden, R.: Regret theory: an alternative theory of rational choice under uncertainty. Econ. J. **92**(4), 805–824 (1982)
19. Camerer, C.: Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton (2003)