2017

Adversarial Decision Making in Counterterrorism Applications

Dogucan Mazicioglu
Virginia Commonwealth University

Adversarial Decision Making in Counterterrorism Applications

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in Systems Modeling and Analysis at Virginia Commonwealth University.

by

John Mazicioglu
Doctor of Philosophy in Systems Modeling and Analysis

Director: Jason Merrick,
Professor, Department of Supply Chain Management and Analytics

Virginia Commonwealth University
Richmond, Virginia
August, 2017

Acknowledgment

First and foremost, I would like to thank my advisor, Dr. Merrick, whom I am forever indebted to for his continued support. Dr. Merrick has been a friend, a mentor, and most importantly a role model who has supported me through both personal and academic hardships over the years. I would also like to thank my committee members for volunteering their time to provide me feedback and guidance on my research. I would like to thank my mother, father, and sister for the sacrifices they have made by providing me this opportunity, a chance to realize the American dream. Miles upon miles walked by my mother every day to get to her job stemming from our inability to afford a car, my father holding four jobs and working 100 hours a week, and my sister taking a full time job as young as 16 years old in addition to her school responsibilities mark our humble beginnings in this country. The sacrifices you have made resonate with me, and they mark my foundation as they have formed my work ethic. These sacrifices remind me daily how grateful I am to have my family. I would also like to thank my girlfriend and best friend, Reina Abboud for bringing positivity to my life and a level of happiness that I before thought wasn’t possible. I am blessed to have her and her family whom I have come to call my family. Last but not least, I would like to thank my brother, Chris Dommert, for all of his help during my undergraduate years at the Virginia Military Institute. You put me in a position to succeed by supporting me at times when I was truly struggling in life. Thank you.

Contents

Abstract

Adversarial Decision Making in Counterterrorism Applications

By Dogucan “John” Mazicioglu, Doctor of Philosophy in Systems Modeling and Analysis

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in Systems Modeling and Analysis at Virginia Commonwealth University.

Virginia Commonwealth University, 2017.

Director: Jason Merrick, Professor, Department of Supply Chain Management and Analytics

Our main objective is to improve decision making in counterterrorism applications by implementing expected utility for prescriptive decision making and prospect theory for descriptive modeling. The areas that we aim to improve are behavioral modeling of adversaries with multi objectives in counterterrorism applications and incorporating risk attitudes of decision makers to risk matrices in assessing risk within an adversarial counterterrorism framework. Traditionally, counterterrorism applications have been approached on a single attribute basis. We utilize a multi-attribute prospect theory approach to more realistically model the attacker’s behavior, while using expected utility theory to prescribe the appropriate actions to the defender. We evaluate our approach by considering an attacker with multiple objectives who wishes to smuggle radioactive material into the United States and a defender who has the option to implement a screening process to hinder the attacker. Next, we consider the use of risk matrices (a method widely used for assessing risk given a consequence and a probability pairing of a potential threat) in an adversarial framework – modeling an attacker and defender risk matrix using utility theory and linking the matrices with the Luce model. A shortcoming with modeling the attacker and the defender risk matrix using utility theory is utility theory’s failure to account for the decision makers’ deviation

from rational behavior as seen in experimental literature. We consider an adversarial risk matrix framework that models the attacker risk matrix using prospect theory to overcome this shortcoming, while using expected utility theory to prescribe actions to the defender.

Chapter 1: Introduction

Merrick and Parnell (2011) review attacker/defender models, concentrating on different methods to represent the adaptation of the attacker to the defender's decisions. Traditional decision analysis approaches represent such adaptation by varying the probability of the attacker choosing between their attack options (von Winterfeldt and Sullivan 2006, Bakir 2008, Merrick and McLay 2010). Game theoretic approaches study the equilibrium of the attacker's and defender's optimal decisions (Kunreuther and Heal, 2003; Bier, 2005; Heal and Kunreuther, 2005; Banks and Anderson, 2006; Zhuang and Bier, 2007; Zhuang et al., 2007; Bier et al., 2007a, 2007b). Paté-Cornell and Guikema (2002), Parnell, Smith, and Moxley (2010), and Rios Insua, Rios, and Banks (2009) developed hybrid approaches that explicitly model the defender's decisions, the attacker's decisions, and uncertainties using game theoretic concepts. The three types of research do share some commonalities, however. They assume that each decision is made by maximizing expected value or expected utility and they use a single attribute to measure the outcomes.

This dissertation is comprised of three chapters on a common theme of attacker/defender models. In the following, we introduce the context of each chapter before explaining our contribution.

1.1 Overview of Chapter 2

Decision making in counterterrorism applications has been studied primarily on a single attribute basis, with the financial damage inflicted on a defender being the primary focus of analysis. Many of the studies in counterterrorism applications have considered an attacker and a defender who are rational. Studies suggest that it is more realistic to consider an attacker with multiple objectives. Further, empirical research suggests that decision makers are not purely rational, as they exhibit loss aversion and likelihood insensitivity, phenomena that are modeled using Prospect theory (Kahneman and Tversky 1979).

In Chapter 2, we consider an attacker with multiple objectives who wishes to maximize the damage inflicted on the defender by smuggling radioactive material into the United States and a defender who wishes to minimize the damage caused by the attacker. The three objectives that the attacker considers are economic damage in terms of financial cost, psychological damage in terms of fear, and organizational growth in terms of recruitment rates. The defender has the option of implementing a screening process to hinder the attacker. We use prospect theory to describe the attacker while using expected utility theory to prescribe actions to the defender. We evaluate the impact of varying reference point, loss aversion and likelihood insensitivity values for the attacker in regards to the defender's screening choices to conduct a complete analysis of prioritization of attacks and trade-offs.

1.2 Overview of Chapter 3

A risk matrix is a risk evaluation method that is widely used in engineering applications. Risk matrices enable the decision maker by providing a ranking framework that helps evaluate the probability and consequence pairing of each risk scenario. The ease of use and ability to make quick decisions render risk matrices an attractive tool for industry use. Cox (2008a) states several shortcomings with the risk matrix approach, while concluding its discontinuation unlikely due to its widespread adoption. Ruan et al. (2015) improves risk matrices by integrating the risk attitudes of decision makers via utility theory.

In chapter 3, we bridge the work of Ruan et al. (2015) and Cox (2008b) to provide an adversarial risk matrix framework. An adversarial risk matrix enables hierarchical decision making where the attacker and the defender is modeled using expected utility theory. We use the Luce model to provide a link between the attacker and defender matrices. We incorporate loss aversion for both the attacker and defender using a utility function with a risk aversion parameter. The impacts of varying attacker/defender loss aversion values on risk prioritization is analyzed to provide a complete understanding of how adversarial risk matrices can be adopted in practice.

1.3 Overview of Chapter 4

In Chapter 4, we extend Chapter 3 but following the approach used in Chapter 2. We consider adversarial risk matrices where the attacker is described using prospect theory and the defender is prescribed actions using expected utility theory. Prospect theoretic modeling of the attacker provides a descriptive framework that is more consistent with empirical research. Similar to Chapter 3, we evaluate the impact of varying attacker/defender loss aversion values on risk prioritization to provide a complete understanding of how adversarial risk matrices with a prospect theoretic attacker and expected utility defender can be adopted in practice.

Chapter 2: Behavioral Modeling of Adversaries with Multiple Objectives in Counterterrorism

2.1 Introduction

An effective counter-terrorism decision framework incorporates the defender's decisions and the attacker's adaptation to them (Brown and Cox 2011). Keeney (2007) proposes that we should consider multiple objectives in the analysis of counterterrorism decisions. This has led to research on terrorists' objectives, using intelligence experts as proxies for the terrorist decision makers (Rosoff and John 2009), with studies considering general terrorists (John and Rosoff 2011, 2014), Al Qaeda (Keeney and von Winterfeldt 2010), and the Islamic State of Iraq and the Levant (ISIL) (Seibert et al. 2016). However, there is a lack of research on incorporating multiple objectives into attacker/defender models.

The aim of attacker/defender models is to provide prescriptive recommendations to the defender, making the common use of expected utility theory appropriate. Recent research has focused on descriptive modeling of the attacker's decision (Nguyen et al. 2013, Yang et al. 2013, Kar et al. 2015, Merrick and Leclerc 2016) using descriptive decision theories like prospect theory (Kahneman and Tversky 1979) and Quantal Response Equilibria (McKelvey and Palfrey 1995). Such approaches are called asymmetrically prescriptive/descriptive (Raiffa 1982) and, thus far, have focused solely on the economic impact of terrorist events on a single attribute basis.

In this paper, we extend attacker/defender models to incorporate multiple objectives. However, rather than assuming full rationality of the attacker in making such complex trade-offs, we follow the approach in asymmetrically prescriptive/descriptive research in modeling the multiple objectives of the attacker. Yang et al. (2013), Merrick and Leclerc (2016), and Kar et al. (2015) chose prospect theory as the descriptive paradigm to model the attacker's decisions in a single-attribute asymmetrically prescriptive/descriptive analysis. We extend this work to multiple attributes using a multi-attribute version of prospect theory. The application of such descriptive decision theories in counter-terrorism is not without difficulty, as we cannot perform decision making experiments on terrorists and it is not clear that the observed behavior of other populations is directly transferrable to modeling terrorists. However, the results of our analysis make it clear that ignoring decision behaviors that are commonly observed in experiments on many different types of decision makers can lead to incorrect predictions of the attacker's actions and incorrect prescriptive recommendations for the defender.

In the next section, we review attacker/defender models and the role of descriptive decision theories in modeling attacker decisions. We review specific descriptive theories for multiple objective decisions in Section 2.3. In Section 2.4, we describe our example decision, extending the work in Merrick and McLay (2010) on screening for radioactive dispersal devices to three attributes. We demonstrate the impact of multi-attribute prospect theory and various parameter choices on the example decision in Section 2.5. We discuss the results in Section 2.6 and their implications for both practitioners and researchers, before drawing conclusions and discussing potential future research directions in Section 2.7.

2.2 Bringing Descriptive Decision Theories to Attacker/Defender Models

Merrick and Parnell (2011) review attacker/defender models, concentrating on different methods to represent the adaptation of the attacker to the defender's decisions. Traditional decision analysis approaches represent such adaptation by varying the probability of the attacker choosing between their attack options (von Winterfeldt and O'Sullivan 2006, Bakir 2008, Merrick and McLay 2010). Game theoretic approaches study the equilibrium of the attacker's and defender's optimal decisions (Kunreuther and Heal, 2003; Bier, 2005; Heal and Kunreuther, 2005; Banks and

Anderson, 2006; Zhuang and Bier, 2007; Zhuang et al., 2007; Bier et al., 2007a, 2007b). Paté-Cornell and Guikema (2002), Parnell, Smith, and Moxley (2010), and Rios Insua, Rios, and Banks (2009) developed hybrid approaches that explicitly model the defender's decisions, the attacker's decisions, and uncertainties using game theoretic concepts. The three types of research do share some commonalities, however. They assume that each decision is made by maximizing expected value or expected utility and they use a single attribute to measure the outcomes.

Raiffa (1982) describes three types of game theory research. The first is symmetrically descriptive research where one aims to describe the behavior of all decision makers in a competitive situation, now known as behavioral game theory (Camerer 2003). The second is symmetrically prescriptive research where one aims to prescribe the normatively best actions of each decision maker. The third is asymmetrically prescriptive/descriptive research where one aims to prescribe the action one decision maker should take, while describing the actions that competing decision makers will take. Expected-utility theory (EUT) (Ramsey 1931, Savage 1954) is the accepted model for prescriptive research. von Neumann and Morgenstern (1944) use EUT to model human preferences over risk in their theory of games, an assumption carried forward in the concept of a Nash equilibrium (Nash 1950). This makes classical game theory symmetrically prescriptive.

However, the descriptive validity of EUT has been questioned, and many experimental anomalies have been identified that EUT cannot satisfy. In an effort to explain these anomalies, a number of alternative models of preferences under risk have been developed. The descriptive research of prospect theory (PT) by Kahneman and Tversky (1979) and Tversky and Kahneman (1992) is a generalization of EUT aimed at accommodating three commonly observed systematic deviations of human behavior from EUT: reference dependence, probabilistic insensitivity, and rank dependence. PT replaces EUT's assumption of linearity in probabilities with a probability weighting function and replaces the utility function with a reference-dependent value function. Here the reference point is the outcome below which the decision maker considers the consequence a loss and prospect theory incorporates the phenomenon of loss aversion by weighting losses more heavily than gains. For monetary outcomes, the reference point is usually assumed to be the decision maker's current state of wealth or zero change. However, it can vary for monetary and other outcomes depending on the decision context. Kadane and Larkey (1982a, 1982b) discuss asymmetrically prescriptive/descriptive analysis in game theory and propose the use of an "empirically supported psychological theory making at least probabilistic predictions about the strategies people are likely to use" for descriptive analyses. There is a small but growing literature that seeks to address how descriptive preferences should impact game theory (Ritzberger 1996, Berejikian 2002a, 2002b, Colman 2003, Butler 2007, Metzger and Rieger 2009).

A growing literature applies asymmetrically prescriptive/descriptive methods in attacker/defender models (Nguyen et al. 2013, Shan and Zhuang 2013, Yang et al. 2013, Merrick and Leclerc 2016, Kar et al. 2015), but this work has not considered an attacker with multiple objectives.

2.3 Descriptive Models of Multiple Attribute Decision Making

There are examples of multi-attribute utility (MAU) models in counter-terrorism. Leung et al. (2004) use a multi-attribute model to help the defender identify critical bridges. Feng and Keller (2006) use a multi-attribute to help the defender determine the best way to distribute Potassium Iodide after a nuclear incident. Merrick and McLay (2010) consider a two-objective value function in examining whether the defender should screen cargo containers for nuclear material. They find that depending on the multi-attribute value function used, the optimal container screening decision can change significantly from that obtained with a single cost objective. MAU functions are appropriate for the defender decision model as we are working in a prescriptive mode and these models are well developed. Keeney and Raiffa (1976) outlined the main theory and approaches for building MAU functions in a prescriptive setting (see Keeney 1992, Kirkwood 1997, Abbas and Matheson 2005, Abbas 2009, Abbas 2011, Abbas and Bell 2011, 2012, and Bond et al. 2008, 2010 for more recent extensions).

MAU models are referred to as compensatory, since an increase in one attribute can compensate for a reduction in another and vice versa. Are such representations appropriate for the attacker decision model that is behavioral in nature? Simon (1955) proposed that people look for alternatives that meet their aspiration level for their most important objective and then their aspiration level for their second most important objective, etc. until they reach a single alternative. Tversky (1972) proposed that people look at aspects of an alternative, essentially features that may include meeting an aspiration level on an alternative. They then pick an aspect and eliminate all alternatives without this aspect, before moving on to the next aspect. However, the order of consideration is random, so the probability that an alternative is chosen is equal to the proportion of times the alternative remained at the end of this process across all possible random orderings of the aspects. Gigerenzer et al. (1999) propose a take-the-best (TTB) heuristic, where the attributes are ordered from most to least valid. The formulations in Simon (1955), Tversky (1972), and Gigerenzer et al. (1999) are non-compensatory as they incorporate different types of lexicographic ordering. Einhorn (1970, 1971) and Tversky et al. (1988) find that decision makers often exhibit non-compensatory effects in their choice patterns. However, this research has not incorporated decision making under uncertainty, a critical emphasis for counter-terrorism decisions and is thus not yet applicable in our application.

Fishburn (1984) and Miyamoto and Wakker (1996) studied MAU under non-expected utility formulations. Zank (2001) and Bleichrodt and Miyamoto (2003) introduced multi-attribute prospect theory, but their formulation of loss aversion considered whether the overall multi-attribute outcome was above or below a reference point, i.e. an overall reference point is elicited in value units. Bleichrodt et al. (2009) proposed a different form of multi-attribute prospect theory where each attribute has its own reference point, i.e. each attribute's reference point is specified in the attribute's units. Each of these behavioral proposals deviates from prescriptive models like MAU. The formulations in Fishburn (1984), Miyamoto and Wakker (1996), Zank (2001), Bleichrodt and Miyamoto (2003), and Bleichrodt et al. (2009) are compensatory, but include behavioral effects such as likelihood insensitivity, reference dependence, and loss aversion that are commonly observed in the literature on decision making under uncertainty. As these compensatory models explicitly represent the uncertainty inherent in counter-terrorism decisions, we will apply these formulations in the ensuing development. Specifically, we will use the formulation of Bleichrodt et al. (2009) as it can represent different reference points for each attribute.

Let us represent an attacker with three objectives as X_1, X_2, and X_3. The outcomes under each objective are represented by $x_{i,j}$, for the i-th attribute and the j-th outcome ($j = 1, \dots, n_i$), and the outcomes are ordered by preference (so $x_{i,j} \geq x_{i,j+1}$). Outcome $x_{i,j}$ occurs with probability $p_{i,j}$ and attribute X_i has m_i outcomes that are gains. We can then write the prospect theoretic value of a multiple attribute prospect under Bleichrodt et al. (2009) as

$$\begin{aligned} PT(X_1, X_2, X_3) = {} & k_1 \left(\sum_{j=1}^{m_1} \pi_1^+(p_{1,j})\, U_1(x_{1,j}) + \sum_{j=m_1+1}^{\mathrm{n}_1} \pi_1^-(p_{1,j}) U_1(x_{1,j}) \right) \\ & + k_2 \left(\sum_{j=1}^{m_2} \pi_2^+(p_{2,j})\, U_2(x_{2,j}) + \sum_{j=m_2+1}^{\mathrm{n}_2} \pi_2^-(p_{2,j}) U_2(x_{2,j}) \right) \end{aligned}$$

$$+k_3\left(\sum_{j=1}^{m_3}\pi_3^+(p_{3,j})\,U_3(x_{3,j})\sum_{j=m_3+1}^{\mathrm{n}_3}\pi_3^-(p_{3,j})U_3(x_{3,j})\right) \qquad (1)$$

with

$$\pi_{\mathrm{i}}^+(p_{i,j}) = W_i^+\left(p_{i,1},\dots,p_{i,j}\right) - W_i^+\left(p_{i,1},\dots,p_{i,j-1}\right), \qquad (2)$$

for $i = 1,2,3,$ and

$$\pi_{\mathrm{i}}^-(p_{i,j}) = W_i^-\left(p_{i,j},\dots,p_{i,n_i}\right) - W_i^-\left(p_{i,j+1},\dots,p_{i,n_i}\right), \qquad (3)$$

for $i = 1,2,3$. $\pi_{\mathrm{i}}^+(\quad)$ and $\pi_{\mathrm{i}}^-(\quad)$ are the decision weights for the i-th attribute's gains and losses, and W_i^+ and W_i^- are the i-th attribute's probability weighting functions for gains and losses. In practice, W_i^+ and W_i^- are often assumed to be the same and take an inverse-S shape that represents over-sensitivity to high and low probabilities (certainty and possibility effects) and under-sensitivity to probabilities near one half (likelihood insensitivity). U_1, $U_{2,}$ and U_3 are strictly increasing utility functions that are continuous and scaled on $[0,1]$ and k_1, k_2, and k_3 are weights that sum to one.

In this formulation, the terrorist's loss aversion and probability weighting function are incorporated on an attribute basis. Attribute level implementation of loss aversion and probability weighting function can be useful in applications. Rottenstreich and Hsee (2001) concluded that decision weighting is outcome dependent with people deviating from expected utility in outcomes having a strong emotional component. While attribute-specific loss aversion has not yet been tested empirically, we find it intuitive to utilize a model that can account for varying attribute-specific loss aversion values.

2.4 Case Study: Container Screening for Radioactive Dispersal Devices

Containers shipments are one method for terrorists to smuggle radioactive material into the United States. Analysis of the screening process must consider the attacker's decisions, defender's decisions, and uncertainty about outcomes. Several studies have focused on improving the container screening process given the significant costs of purchasing the screening equipment, continuing operations at each port, and delays to cargo entering the US. Lewis et al. (2003) consider problems encountered at prominent container ports, using algorithms to reduce delays incurred during the screening of containers. Bakir (2008) considers defending the US-Mexico border, evaluating the choice between new advanced spectroscopic portals and existing Radiation Portal Monitors on the US side of the border. Merrick and McLay (2010) consider a two-objective value function in examining whether the defender should screen cargo containers for nuclear material. Merrick and Parnell (2011) consider prescriptive intelligent adversary risk analysis, an adaptive decision making process where the defender adjusts the defense strategy based on the attacker's decisions.

Merrick and Leclerc (2016) layout the framework for the attacker and defender decisions, costs, and uncertain events for an adversary considering a single objective, maximizing economic damage (cost) suffered by the defender. The defender's decision is shown as a decision tree in Figure 2.1.

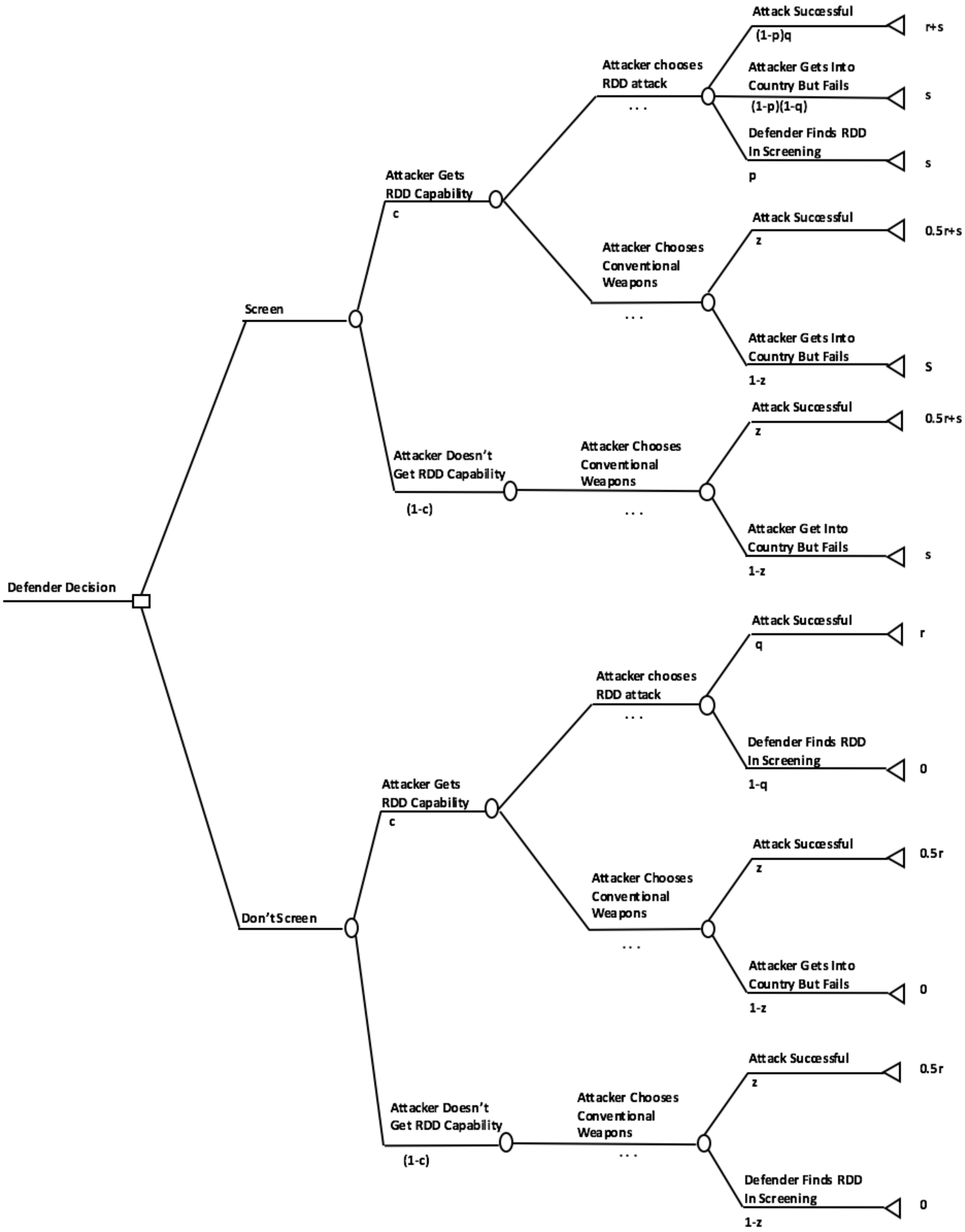

Defender Decision
Screen
Attacker Gets RDD Capability
c
Attacker chooses RDD attack
. . .
Attack Successful
(1-p)q
r+s
Attacker Gets Into Country But Fails
(1-p)(1-q)
s
Defender Finds RDD In Screening
p
s
Attacker Chooses Conventional Weapons
. . .
Attack Successful
z
0.5r+s
Attacker Gets Into Country But Fails
1-z
S
Attacker Doesn't Get RDD Capability
(1-c)
Attacker Chooses Conventional Weapons
. . .
Attack Successful
z
0.5r+s
Attacker Get Into Country But Fails
1-z
s
Don't Screen
Attacker Gets RDD Capability
c
Attacker chooses RDD attack
. . .
Attack Successful
q
r
Defender Finds RDD In Screening
1-q
0
Attacker Chooses Conventional Weapons
. . .
Attack Successful
z
0.5r
Attacker Gets Into Country But Fails
1-z
0
Attacker Doesn't Get RDD Capability
(1-c)
Attacker Chooses Conventional Weapons
. . .
Attack Successful
z
0.5r
Defender Finds RDD In Screening
1-z
0

Figure 2.1: The defender decision tree

The defender must first decide whether to install and operate the screening equipment. There is then a probability that the attacker can obtain the nuclear material and construct a Radiation

Dispersal Device (RDD) (or dirty bomb), followed by the attacker's choice between available attack options. The distribution on the attacker's choice is dependent on the defender's decision because the attacker can observe whether a screening operation is in place at a given port through surveillance. If the attacker chooses a dirty bomb attack, then there is a known probability that the device is found during screening and if not found then there is a probability of a successful or failed attack.

The defender can solve this decision tree by first determining the attacker's decision. Screening is an effective option if it is a deterrent, i.e. the attacker will choose an RDD attack if the defender is not screening, but will choose a CW attack if the defender is screening. If the attacker is going to choose an RDD attack whether the defender is screening or not, then clearly screening is not effective. Further, if the attacker will always choose a conventional weapons (CW) attack whether the defender is screening or not, then it is not worthwhile for the defender to screen. Therefore, the defender should choose to invest in screening if it will deter the attacker from an RDD attack, but should not choosing screening if it does not change the attacker's decision.

Merrick and Leclerc (2016) model the attacker's decisions using single-attribute prospect theory. They consider only the cost of the attack to the defender, with the attacker seeking to maximize the cost and the defender seeking to minimize it. Keeney (2007) develop an initial set of objectives for a generic terrorist. Keeney and von Winterfeldt (2010) interview intelligence experts and review public statements to determine Al Qaeda's objectives. Seibert et al. (2016) follow a similar process to analyze the objectives of both leaders and followers in ISIL.).

The objectives from Keeney (2007), Keeney and von Winterfeldt (2010), and Seibert et al. (2015) fall into three themes: maximizing economic impact, instilling fear in the target population, and ensuring the longevity of the terrorist organization (represented here as maximizing recruitment). We extend the work of Merrick and Leclerc (2016) to consider these three objectives. This is admittedly a small subset of the objectives determined for Al Qaeda and ISIL, but our objectives are generally representative of the type of objectives in the broader set and allow us to study the impact of the parameters of the behavioral multi-attribute model without becoming overly complex. The attacker seeks to achieve these objectives in their decision whether to attack using a RDD or a conventional weapons attack.

The recruitment attribute is standardized and ranges from 0 to 1. The best outcome from the attacker's perspective under recruitment is a successful attack, whether RDD or CW, as this will give positive publicity amongst the terrorist group's followers and increase recruitment, so both outcomes are assigned a value of 1. The worst outcome for recruitment is the smuggled nuclear material being found during screening as this will cause negative publicity and could reduce recruitment; this outcome is given the value 0. An attacker getting into the country but failing to carry out a CW attack is assumed to be neutral in terms of recruitment; this outcome is given the value 0.5. The effect on recruitment from the attacker getting an RDD into the country but failing to carry out the attack, denoted by f, is assumed to be 0.25.

The fear attribute is also standardized and ranges from 0 to 1. The worst outcome for instilling fear is a failed CW attack, which is given the value 0. The best outcome for instilling fear is again a successful RDD attack, which is given the value 1. The fear caused by a successful CW attack is assumed to be half that of an RDD attack. If the defender finds radioactive material during screening or an RDD attack is foiled after getting into the country, the publicity will still cause fear in the target population. This outcome is given the value t, which is assumed to be 0.2.

Following Merrick and Leclerc (2016), the cost of an RDD attack, denoted r, is assumed to be $40 billion. The cost of a conventional weapon attack, w, is assumed to be $20 billion. The cost of screening, denoted s, is assumed to be \$100 million. We linearly scale the cost attribute from 0 to 1 to match the other attribute so zero cost takes the value 0 and the maximum cost of $r + s$ takes the value 1. The attacker's decisions tree is shown in Figure 2.2 with the outcomes under these three objectives.

The reference point for the attributes of cost, recruitment, and fear are assumed to be \$15 billion, 0.5, and 0, respectively. Hence, an attack yielding an economic cost of less than \$15 billion is regarded as a loss by the attacker. An outcome that negatively affects recruitment, characterized by a recruitment value of less than 0.5, is regarded as a loss by the attacker. All attacks, failed and successful, are assumed to either instill fear or maintain fear at existing levels. However, as the attacker can reach a guaranteed minimum fear level of 0.2 by attempting an RDD attack, we assume that the attacker finds a fear value less than 0.2 to be a loss. We use a simple value function for each attribute to represent loss aversion, specifically

$$v_i(x_i) = \begin{cases} x_i - r_i, & x_i \geq r_i \\ \lambda_i(x_i - r_i) & x_i < r_i \end{cases}$$

where x_i is the i-th attribute, r_i is its reference point, and λ_i represents the attacker's loss aversion with respect to the attribute, for $i = 1,2,3$.

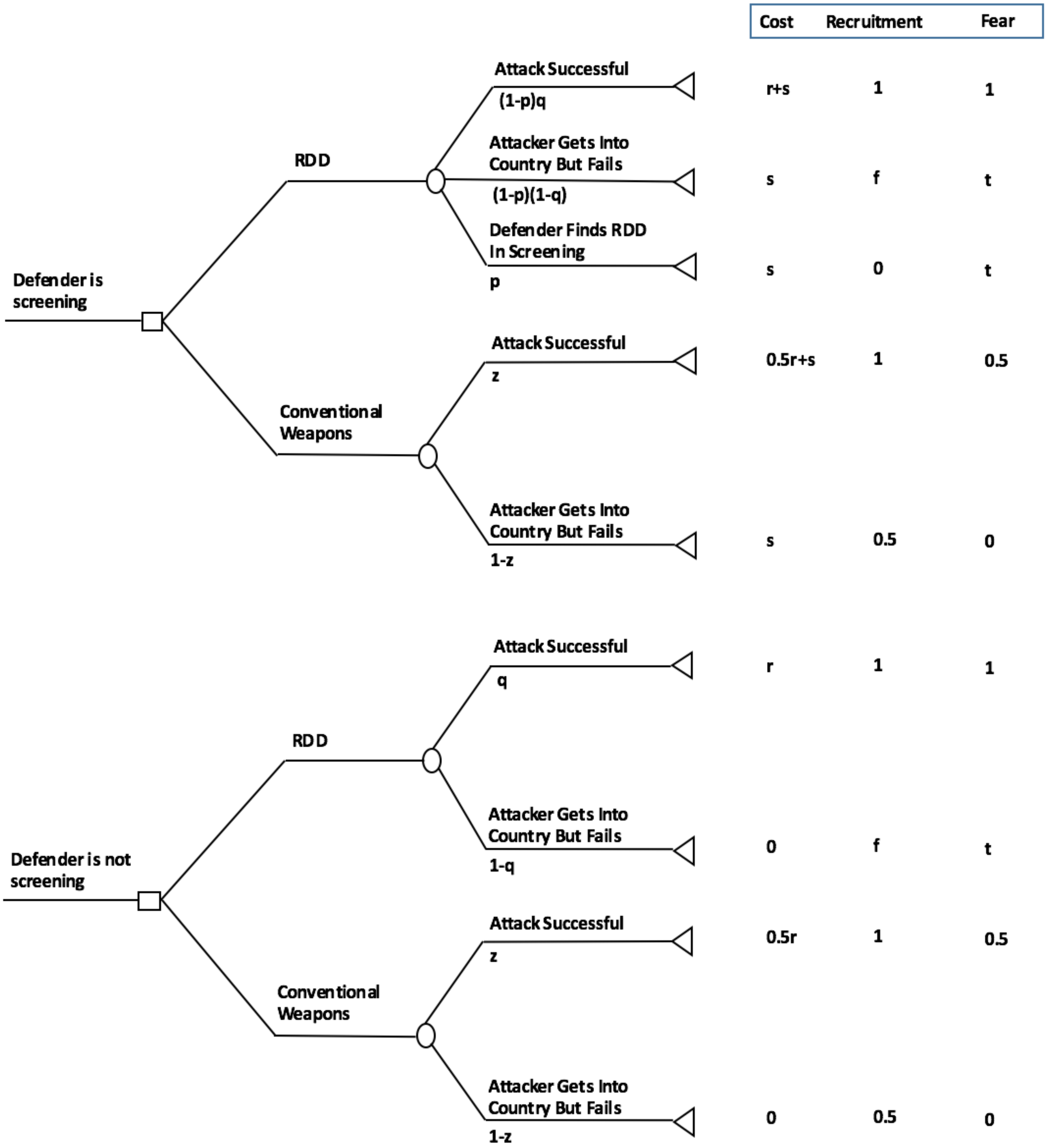


Figure 2.2: The attacker decision tree

Again, following Merrick and Leclerc (2016), the probability of the defender finding the radiological material during screening, denoted by p, is assumed to be 0.8. The probability of a successful RDD attack once the material has been smuggled into the country, denoted by q, is assumed to be 0.5. The probability of a successful CW attack, denoted by z, is assumed to be 0.9. We implement the neo-additive probability weighting function as discussed by Chateauneuf et al. (2007),

$$w(p) = \begin{cases} 1 & p = 1 \\ kp + \frac{1}{2}(1-k) & 0 < p < 1 \\ 0 & p = 0 \end{cases}$$

where $0 <$ k < 1. When k =1, this function is a straight line $(w(p) = p)$ and represents no probability weighting. When k <1, the function maintains the values $w(0) = 0$, $w(1) = 1$, and $w\left(\frac{1}{2}\right) = \frac{1}{2}$, but has a slope less than one for $0 < p < 1$ which represents likelihood insensitivity. This leads to a significant jump in probability weight from just below 1 to 1 (representing the certainty effect) and from 0 to just above 0 (representing the possibility effect). Baillon et al. (2014) consider values of k between 0.6 and 0.8 when evaluating likelihood insensitivity for the neo-additive probability weighting function. Thus, we use 0.7 as our base value and vary k between 0.6 and 0.8 to determine the sensitivity of the model to this parameter. Novemsky and Kahneman (2005) determined that values for loss aversion, here denoted by λ_i $(i = 1,2,3)$, range from 1 to 3 in experiments with 2 being the mode of observations. Hence, we consider the values of 1, 2, and 3 for each λ_i.

2.5 Results

2.5.1 Decisions for a Risk-Neutral Attacker Considering Only One Objective

We evaluate the optimal decision for a risk-neutral attacker under each of the three objectives independently (see Figure 2.2) before considering the implications of trade-offs amongst objectives. We assume that f, t, p, q, z > 0, and q ≤ z. We first consider the attacker's decisions when the defender is screening.

The expected cost for an RDD attack is $(1-\mathrm{p})\mathrm{q}(\mathrm{r}+\mathrm{s})+(1-\mathrm{p})(1-\mathrm{q})\mathrm{s}+\mathrm{ps}=-\mathrm{pqr}+\mathrm{qr}+\mathrm{s}=(1-\mathrm{p})\mathrm{qr}+\mathrm{s}$. The expected cost for a CW attack is $\frac{1}{2}\mathrm{rz}+\mathrm{s}$. Thus, a risk neutral attacker considering only cost will choose an RDD attack if $(1-\mathrm{p})\mathrm{qr}+\mathrm{s}>\frac{1}{2}\mathrm{zr}+\mathrm{s}$ or $(1-p)q>\frac{1}{2}z$. As $q\leq z$, this simplifies to $(1-p)>\frac{1}{2}$ or $p<\frac{1}{2}$. As p is the probability of finding an RDD during screening, this condition should not hold for any reasonable screening technology, so a risk neutral attacker should choose CW based on cost alone.

Considering only recruitment, CW stochastically dominates RDD if $z\geq(1-p)q$. As the probability of a CW attack ($z=0.9$) is high and the probability of getting through screening and then carrying a successful RDD attack ($(1-p)q=0.2\times 0.5=0.1$) is low, CW stochastically dominates RDD under recruitment for plausible parameter values.

The trade-offs under fear are more complex. The expected value for fear under an RDD attack and a CW attack are $(1-p)q+(1-(1-p)q)t$ and $\frac{1}{2}z$ respectively. Therefore, the preferences of a risk-neutral attacker considering only fear are dependent on the specific parameter choices. At the base case values, RDD is preferred. (Note that RDD would stochastically dominate CW if $(1-p)q\geq z$, but as this is the opposite condition to that mentioned above, this is not plausible)

We next consider the simpler case of the attacker's decisions when the defender is not screening. The expected value for economic damage under an RDD attack is qr compared to $\frac{1}{2}zr$ under a CW attack. Hence, the attacker prefers an RDD attack if $2q>z$; this condition is satisfied at base case values. Considering only recruitment, CW again stochastically dominates RDD as $f<\frac{1}{2}$ and q ≤ z at plausible parameter values. The expected value for fear under an RDD attack is $q + (1 - q)t$ compared to $\frac{1}{2}z$ under a CW attack, again leading to a more complex trade-off, but RDD is chosen at the base case parameter values.

In summary, a risk-neutral attacker considering only cost will choose CW if the defender screens and RDD if not. So this attacker is deterred by screening. However, the risk neutral attacker considering only recruitment will choose CW regardless of whether the defender is screening or not, while the risk-neutral attacker considering fear will choose RDD regardless. These results show the differences between attacker's who value outcomes differently.

2.5.2 Multi-objective Decision Making for the Rational versus Prospect Theoretic Attacker

Figure 2.3 shows strategy plots for varying attribute weights for the base case prospect-theoretic attacker and a rational (and risk-neutral) attacker. These plots show the full space of possible attribute weight combinations for the cost, recruitment, and fear attributes. Since the attribute weights must sum up to a 100%, a ternary plot is used to represent all of the possible attribute weight combinations for the three attributes (cost, recruitment, and fear) in two dimensional space.

The axes/edges of the ternary plot are labeled using arrows and the respective attribute names to display the direction in which the attributes increase in value. Since there are three attributes, there are seven types of attribute weight combinations. All of the possible attribute weight combination types and their respective location on the strategy plots are as follows:

- All weight placement on cost – top vertex
- All weight placement on fear – lower left vertex
- All weight placement on recruitment – lower right vertex

Excluding the vertices:

- Some weight placement on fear and cost– fear axis
- Some weight placement on fear and recruitment– recruitment axis

- Some weight placement on cost and recruitment– cost axis

Excluding the vertices and axes:

- Some weight placement on each of the attributes – triangle's interior

An attribute weight combination that falls on one of the three vertices corresponds to all weight placement on one attribute, hence no weight placement on the remaining two attributes. An attribute weight combination that falls on an axis, but not the vertex, corresponds to weight placement on only two of the three attributes (the axis that the point is on and the clockwise axis), hence no weight placement on the remaining attribute (the counter-clockwise axis). For example, if a point falls on the recruitment axis at the 40%, an attribute weight combination of 40% recruitment, 60% fear and 0% cost exists.

The attribute weight combinations where each attribute receives a weight are on the triangle's interior, where the triangle's interior is defined as any part of the triangle that is not an edge or a vertex. Three lines that run through the point of interest and extend to each axis are required to determine the attribute weights for cost, fear, and recruitment, respectively:

1) A line parallel to the recruitment axis that connects the point to the cost axis
2) A line parallel to the cost axis that connects the point to the fear axis
3) A line parallel to the fear axis that connects the point to the recruitment axis

The white gridlines on the triangle's interior provide the guidance required to determine the attribute weights for the points in the triangle's interior.

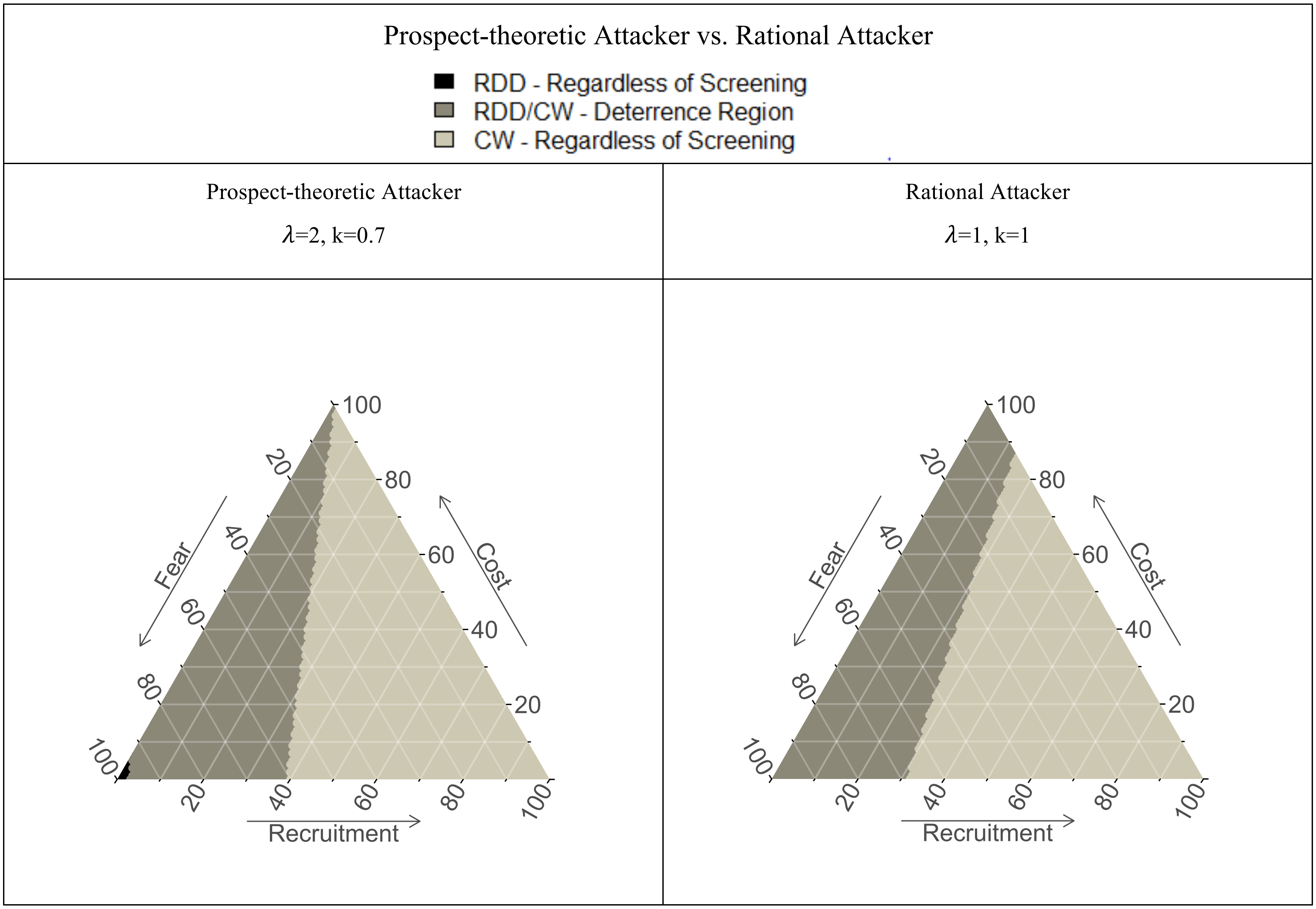


Figure 2.3: Evaluation of Prospect-theoretic Attacker vs. Rational Attacker

Three shaded regions are provided to display actions available to the attacker: (1) use an RDD attack regardless of whether the defender is screening (the black region), (2) use an RDD attack when the defender is not screening and a CW attack when the defender is screening (the dark brown region), and (3) use a CW attack regardless of whether the defender is screening (the light brown region). The three regions represent all possible actions for the attacker. Thus, moving from a black region to the dark brown region means the attacker has switched from choosing RDD to CW if the defender is screening, while the attacker's decision remains the same if the defender is not screening, namely RDD. Similarly, moving from the dark brown region to the light brown region means that the attacker has switched from choosing RDD to CW if the defender is not screening, while the attacker's decision remains the same if the defender is screening, namely CW.

Considering the defender's decision, as discussed in Merrick and Leclerc (2016) if the attacker uses an RDD attack regardless of screening or the attacker uses a CW attack regardless of screening then the defender should not screen, as screening is not a deterrent. On the other hand, if the attacker uses an RDD attack when the defender is not screening and a CW attack when the defender is screening, then the attacker is deterred by screening and screening is the best alternative for the defender. Thus, screening is a deterrent in the dark brown region, but not in the black or light brown regions.

Figure 2.3 shows the strategy plot for a prospect theoretic attacker on the left and a rational attacker on the right. The attacker is considered rational when $\lambda_i = 1$ $(i = 1,2,3)$ and $k = 1$, i.e. no loss aversion or likelihood insensitivity. The base case values for the prospect-theoretic attacker's loss aversion and likelihood sensitivity are $\lambda_i = 2$ for all attributes $(i = 1,2,3)$ and $k = 0.7$, as discussed in Section 2.4. As we found for the risk-neutral attacker considering only one objective, when all the weight is on cost, the rational attacker is deterred by screening; when all the weight is on recruitment, the rational attacker chooses CW regardless of screening; when all the weight is on fear, the rational attacker chooses RDD regardless of screening. In fact, the rational attacker does not use an RDD attack when the defender is screening, except when all the weight is on fear. The region where a prospect-theoretic attacker uses an RDD attack regardless of screening is larger than for a rational attacker, but the attribute weight on fear must still be close to 1.

Let us now consider the region where the attacker is deterred by screening. For the rational attacker, if no weight is applied to recruitment, the attacker will be deterred. If all the weight is on cost and we start to add weight to recruitment, then there is a region before screening stops being a deterrence and the attacker chooses CW regardless of screening. If all the weight is on fear and we start to add weight to recruitment, then there is a larger region before screening stops being a deterrence. Comparing this to the prospect theoretic attacker, if all the weight is on cost and we start to add weight to recruitment, then there is almost no region before screening stops being a deterrence and the attacker chooses CW regardless of screening. If all the weight is on fear and we start to add weight to recruitment, then there is a region before screening stops being a deterrence, but it is smaller than for the rational attacker. This relationship is complex, so in the following sections we analyze the effect of various parameters on the attacker's decision.

2.5.3 Sensitivity of the Prospect Theoretic Attacker to Changes in Probabilities

We vary the value of the probability of a successful RDD attack once the device is in the country, denoted by q, for a prospect theoretic attacker ($\lambda = 2, k = 0.7$) in Figure 2.4. As q increases, the regions corresponding with RDD attacks also increase. The attacker employs an RDD attack across more attribute weight combinations regardless of the screening status. The attacker does not use an RDD attack when the defender is screening with $q = 0.25$.

Evaluation of Varying Probability of a Successful RDD Attack Values (λ=2, k=0.7)

■ RDD - Regardless of Screening
▨ RDD/CW - Deterrence Region
□ CW - Regardless of Screening

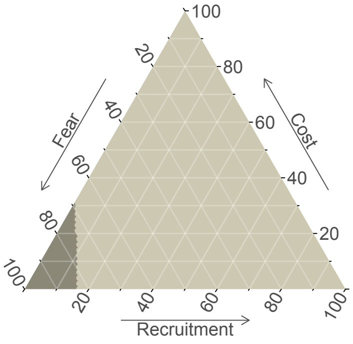
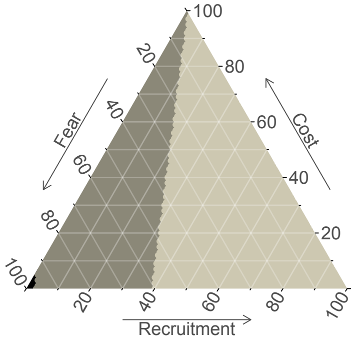
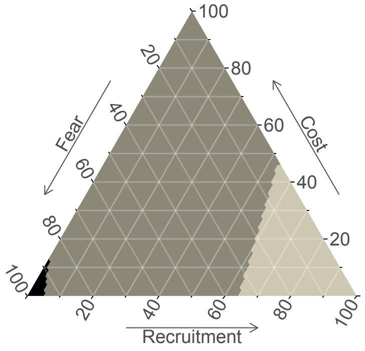
q=0.25	q=0.5	q=0.75
		

Figure 2.4: Sensitivity of the Attacker's Decisions to the Probability of Successful RDD Attack Values

The deterrence region, the area when screening is worthwhile, gets larger as q increases, thus rendering screening useful across more attribute weight combinations. Thus, if the chance of success of an RDD attack that gets into the country is low, the attacker will only be deterred by screening if they put significant weight on fear. Alternatively, if the chance of success once in the country is high, the attacker will be deterred unless almost all the weight is on fear or most of the weight is on recruitment.

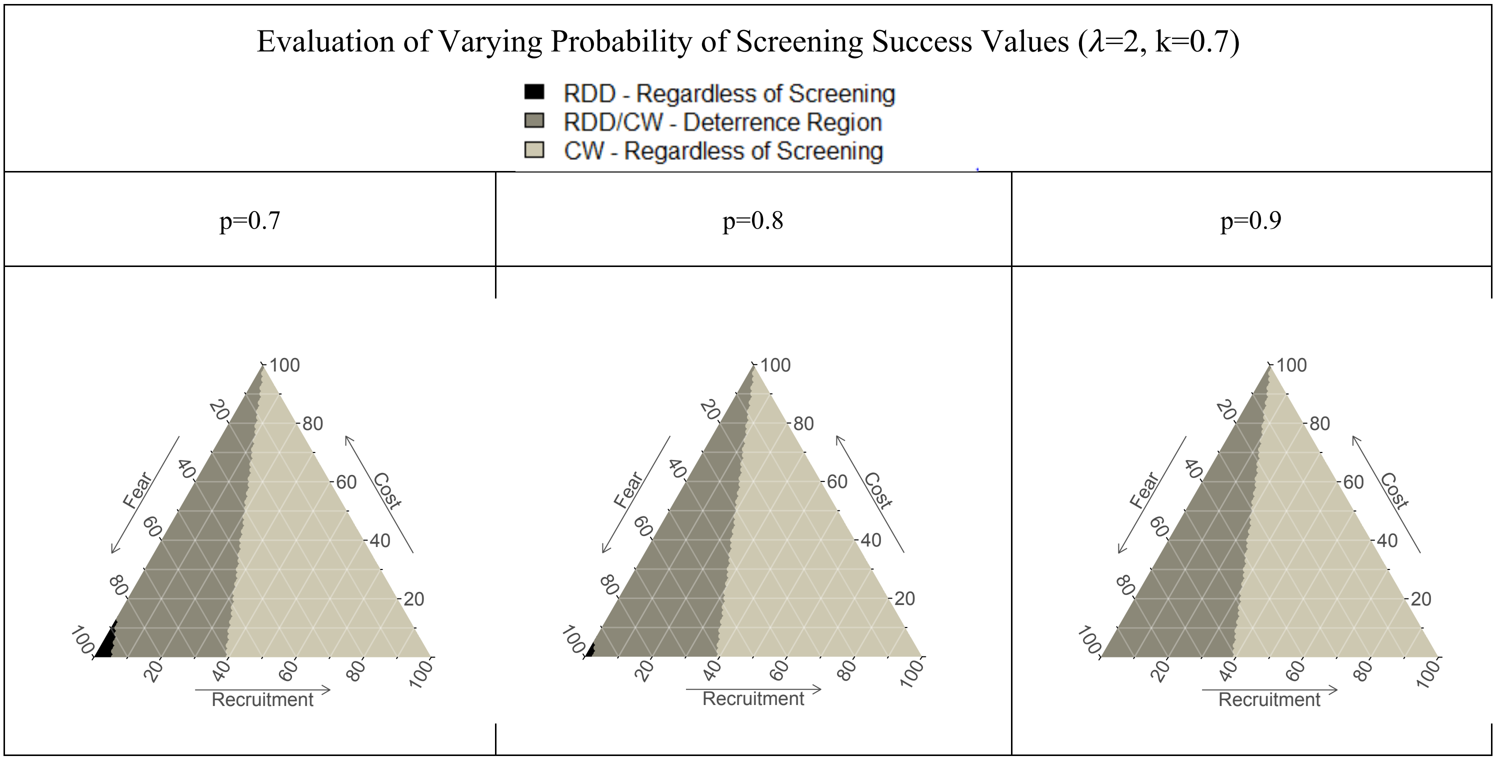


Figure 2.5. Sensitivity of the Attacker's Decisions to the Probability of Successful RDD Attack Values

The evaluation of varying probability of screening success values, p, using the benchmark values of (λ=2, k=0.7) is provided in Figure 2.5. As p increases, the area representing RDD attack regardless of screening decreases. The attacker is less likely to employ an RDD attack when the defender is screening with an increasing probability of screening success. At $p = 0.9$, the attacker does not use an RDD attack when the defender is screening. The deterrent area increases as p increases, since the attacker becomes less likely to employ an RDD attack regardless of screening. The difference between the dark brown and light brown regions is the attacker's decision when the defender is not screening, so the border between them is not affected by changes in p.

2.5.4 Sensitivity of the Prospect Theoretic Attacker's Decisions to Overall Loss Aversion and Likelihood Insensitivity

We vary the values of loss aversion and likelihood insensitivity while maintaining the benchmark values of other parameters in Figure 2.6. We consider the values of 1, 2, and 3 for loss aversion for all attributes ($\lambda_1 = \lambda_2 = \lambda_3$), and the values of 0.6, 0.7, and 0.8 for likelihood insensitivity (k) for the neo-additive probability weighting function. As the attacker becomes more probabilistically sensitive, the attacker is less likely to use an RDD attack. Across all values of loss aversion, the attacker does not use an RDD attack when the defender is screening with $k = 0.8$. This indicates that if the attacker was more sensitive to changes in probability then the attacker would not choose an RDD attack when the defender is screening present for any values of the attribute weights. They would also only choose the RDD attack when the defender is not screening if the weight on fear was significant. Since an RDD attack instills more fear than a CW attack, it is intuitive that the attacker would prefer an RDD attack when instilling fear is the main objective.

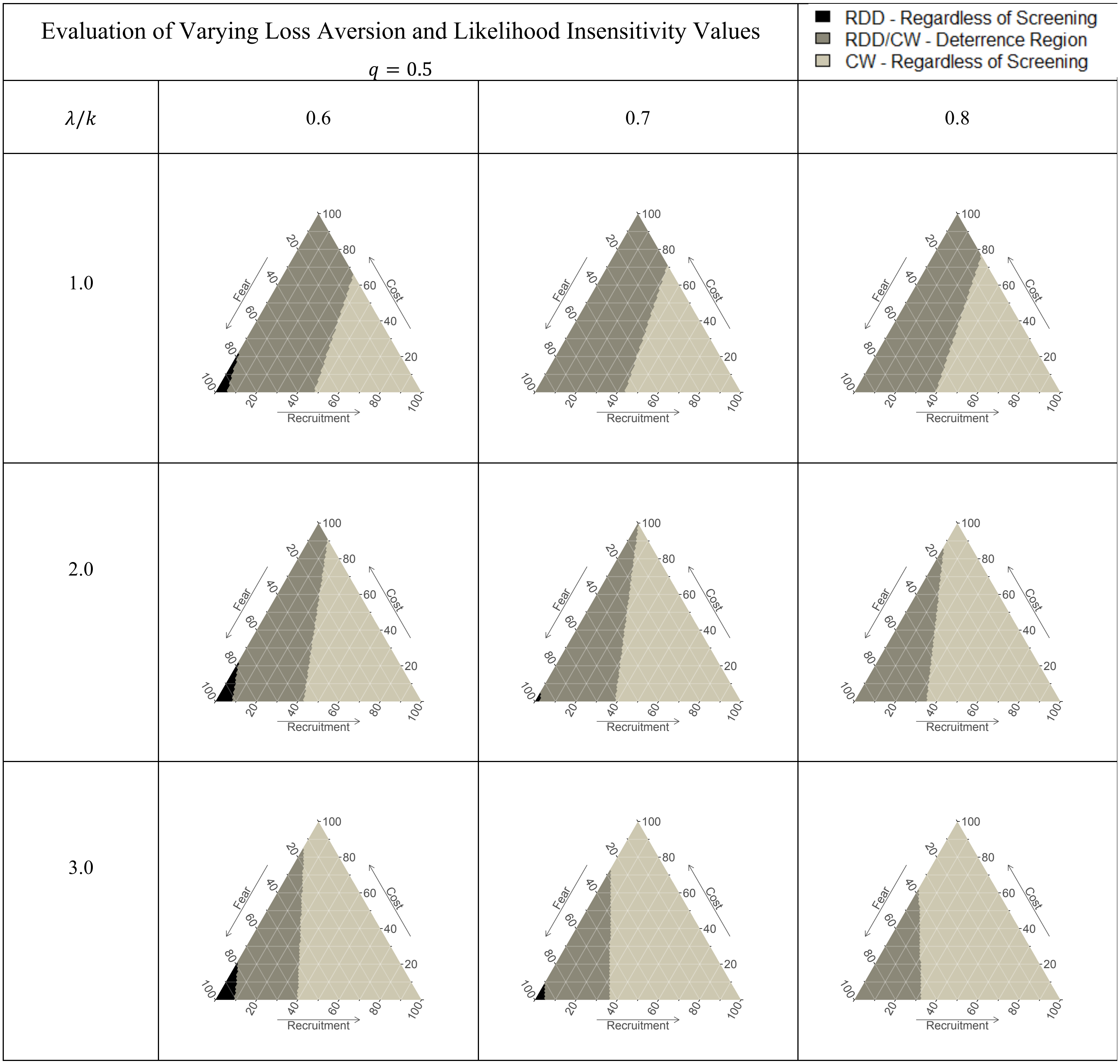


Figure 2.6: Sensitivity of the Attacker's Decisions to Varying Levels of Loss Aversion and Likelihood Insensitivity

As the probabilistic insensitivity increases and the attacker's decision weight increasingly deviates from actual probabilities, an RDD attack takes place across more attribute weight combinations. When all of the weight is placed on recruitment, the attacker chooses a CW attack in all scenarios. Since, the failure of an RDD attack is considered a loss under recruitment (failure including the RDD material being found in the screening or the RDD making it through screening but failing to execute), it makes sense that an attacker placing all the weight on recruitment would choose a CW attack. Furthermore, as the loss aversion of the attacker increases, while keeping likelihood insensitivity constant, the attacker chooses to employ a CW attack regardless of screening for more attribute weight combinations as a CW attack provides the attacker with a higher probability of avoiding loss at our benchmark reference points.

2.5.5 Sensitivity of the Prospect Theoretic Attacker's Decisions to Changes in Reference Points and Loss Aversion

To further examine the effect of loss aversion, we vary the reference point and loss aversion values one attribute at a time while keeping the rest of the attribute values at the benchmark setting. Loss aversion is varied between 1 and 3 for each attribute. The reference points are varied on an attribute basis.

The effects of the varying reference point and loss aversion values on cost are shown in Figure 2.7. We consider \$5 billion, \$15 billion, and \$25 billion for the reference point. The attacker's decision across the full space of attribute weight combinations when $\lambda = 1$ (irrespective of the reference point) is the same for each reference point level.

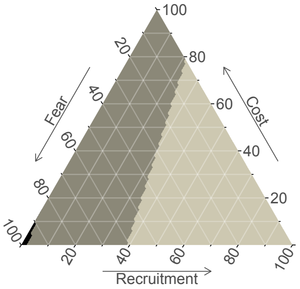
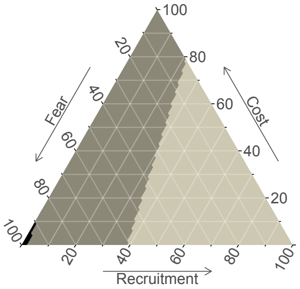
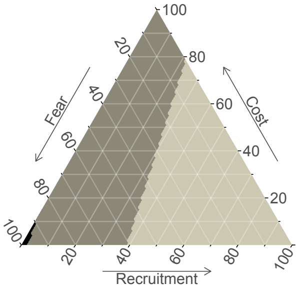
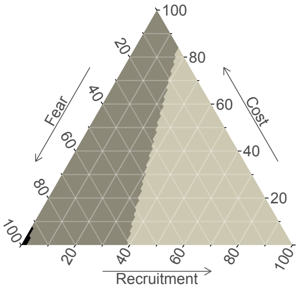
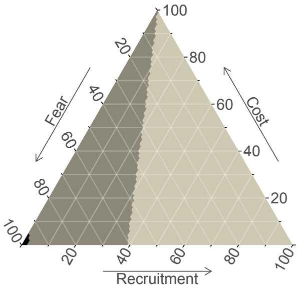
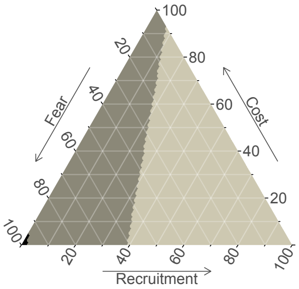
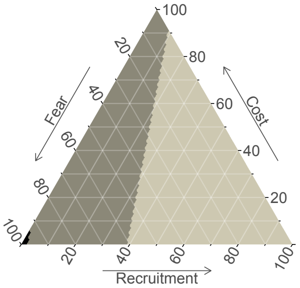
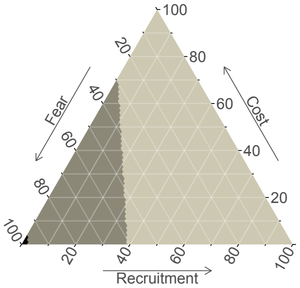
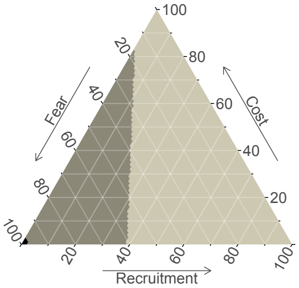
Evaluation of Cost			■ RDD - Regardless of Screening ▨ RDD/CW - Deterrence Region □ CW - Regardless of Screening
Lambda/ Reference Point	$5 Billion	$15 Billion	$25 Billion
1.0			
2.0			
3.0			

Figure 2.7: The Effect of Varying Loss Aversion and Reference Point Values on Cost

Thus, an attacker without loss aversion perceives the varying reference points the same. The necessity of loss aversion to alter the attacker’s decision is justified given that CW attack yields half the damage of an RDD attack. A loss aversion value greater than 1 modifies the attacker’s decision.

The attacker’s decision varies at different loss aversion values when the reference points for cost are \$15 billion and \$25 billion. As loss aversion increases for the reference points of \$15 billion and \$25 billion, the deterrence and RDD attack regardless of screening areas decrease. An attacker with most of the weight on instilling fear chooses an RDD attack irrespective of screening. As loss aversion increases, the attacker must place more weight on fear to use an RDD attack when the defender is screening because of the amplified losses of an RDD attack and the higher probabilities of such losses (when compared to CW). This is shown by the black region decreasing in size as loss aversion increases.

Similarly, the deterrence region gets smaller as loss aversion increases. When screening is not present, the attacker using an RDD attack has a 0.5 probability of getting \$40 billion and a 0.5 probability of getting \$0. Same attacker has a 0.9 probability of getting \$20 billion and 0.1 probability of getting \$0. As loss aversion increases, an RDD attack with an unfavorable outcome paired with a higher probability of loss results in significantly larger losses than a CW attack. Thus, an attacker chooses a CW attack regardless of screening for more attribute weight combinations as loss aversion increases. Lastly, the deterrence region increases as the reference point increases, keeping loss aversion constant. An attacker finds an RDD attack more attractive as the reference point increases, making the CW attack less worthwhile.

The effects of the varying reference point and loss aversion values on recruitment are shown in Figure 2.8. We consider 0.25, 0.5, and 0.75 for the varying reference points. When the reference point for recruitment is 0.25, the recruitment outcomes are effectively in an all gains frame, with the exception of the RDD being found during screening. This one outcome does not have a significant effect, so, the attacker's decision is the same across the varying loss aversion values. Further, the attacker's decision across the full space of attribute weight combinations when $\lambda = 1$ (irrespective of the reference point) is the same as what we observe when the reference point = 0.25 (irrespective of λ). The loss generated by recruitment at varying reference points is too small to have an impact on the overall decision in the absence of loss aversion.

The attacker's decision is not the same at varying loss aversion values when the reference points for recruitment are 0.5 and 0.75. As loss aversion increases for these reference points, the areas representing deterrence and an RDD attack regardless of screening decrease. An attacker whose primary objective is to instill fear uses an RDD attack regardless of screening for less attribute weight combinations for reference points of 0.5 and 0.75 as loss aversion increases. This is intuitive as an RDD attack getting into the country and failing poses a loss that is amplified by increasing loss aversion, causing the attacker to reduce RDD attacks regardless of screening. The deterrence region decreases with increasing loss aversion for the reference points of 0.5 and 0.75, because the loss on recruitment from an RDD attack failing once in the country is increasingly amplified. Thus, the attacker prefers a CW attack regardless of screening. Lastly, the deterrence region decreases as the value for the reference point increases, while keeping loss aversion constant.

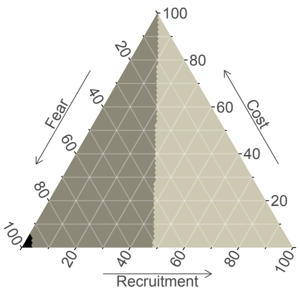
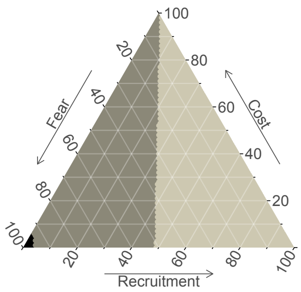
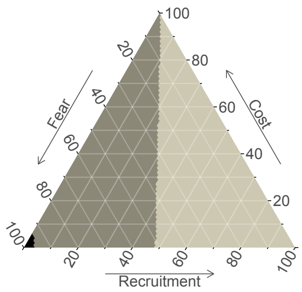
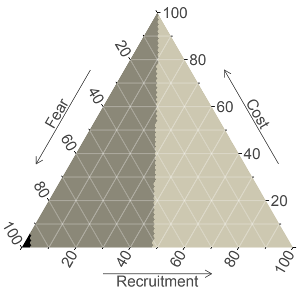
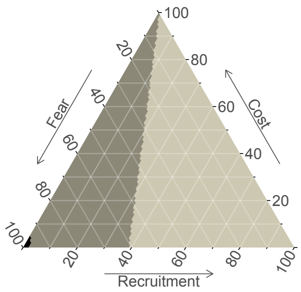
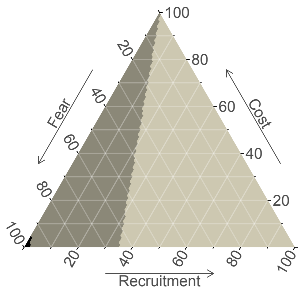
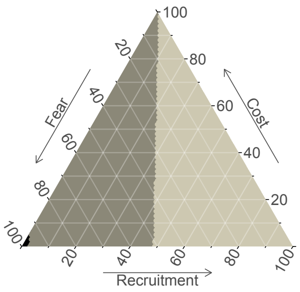
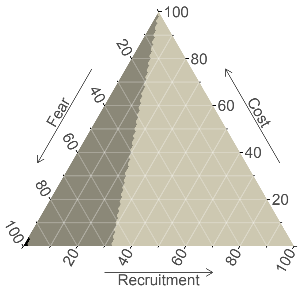
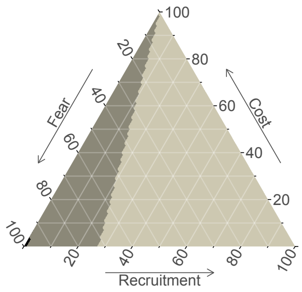
Evaluation of Recruitment			■ RDD - Regardless of Screening ▩ RDD/CW - Deterrence Region □ CW - Regardless of Screening
Lambda/ Reference Point	0.25	0.5	0.75
1.0			
2.0			
3.0			

Figure 2.8: The Effect of Varying Loss Aversion and Reference Point Values on Recruitment

The attacker suffers an increasing loss from an RDD attack getting into the country but failing as the reference point increases, making the CW attack regardless of screening a more attractive alternative.

The effects of the varying reference point and loss aversion values on fear are shown in Figure 2.9. We consider 0, 0.2, and 0.5 for the varying reference points. Fear is an all gains framework when the reference point is 0, since there are not any outcomes generating a fear value less than 0. Thus, the attacker's decision, influenced by an all gains fear frame, is the same across varying loss aversion values. Similarly, the attacker's decision across the full space of attribute weight combinations when $\lambda = 1$ (irrespective of the reference point) is the same as what we observe when the reference point = 0 (irrespective of λ). Thus, the attacker's decision in the absence of loss aversion placed on fear mirrors the actions of when there is an all gains fear framework. The loss generated by fear in the absence of loss aversion is too small to have an impact on the overall decision.

When the reference points for fear are 0.2 and 0.5, the attacker's decision changes as the loss aversion increases. An attacker whose primary objective is to instill fear uses an RDD attack regardless of screening for more attribute weight combinations when the reference point is 0.2 and loss aversion is increasing. This result is intuitive, since an RDD being intercepted during screening is not considered a loss with a 0.2 reference point. Conversely, a reference point of 0.5 renders RDD caught during screening a loss. Thus, the attacker entirely abandons an RDD attack regardless of screening for all loss aversion values when the reference point is 0.5.

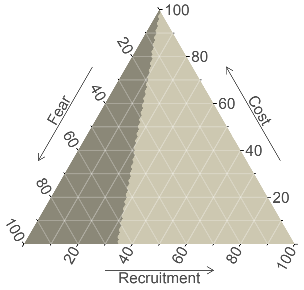
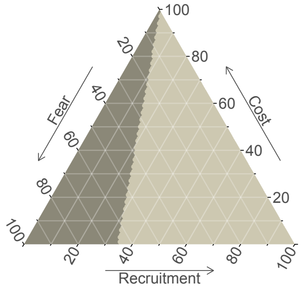
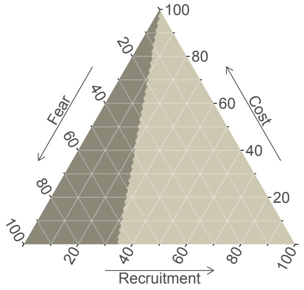
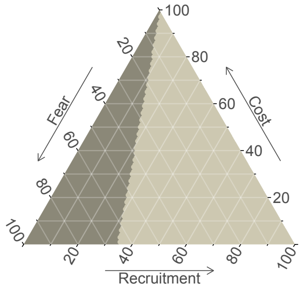
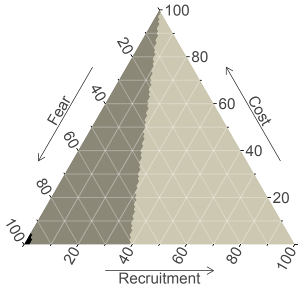
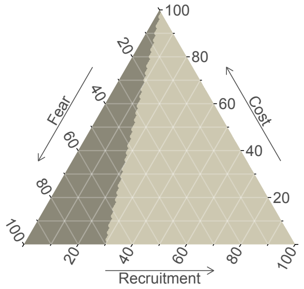
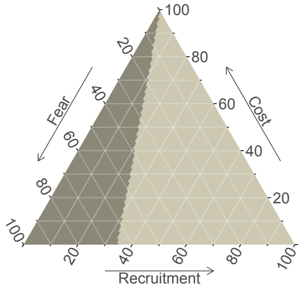
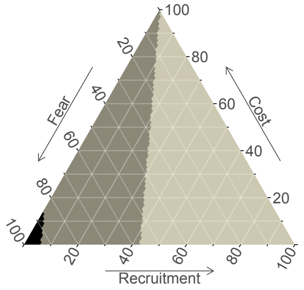
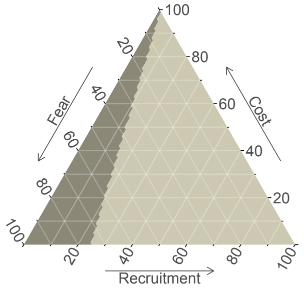
Evaluation of Fear			■ RDD - Regardless of Screening ▨ RDD/CW - Deterrence Region □ CW - Regardless of Screening
Lambda/ Reference Point	0	0.2	0.5
1.0			
2.0			
3.0			

Figure 2.9: The Effect of Varying Loss Aversion and Reference Point Values on Fear

The deterrence region, when the fear reference point is 0.2, shifts to encompass more attribute weight combination as loss aversion increases. The change of the size in deterrence regions when the reference point is 0.2 with varying loss aversion values is negligible since RDD attack regardless of screening region also increases. The attacker is equally deterred across loss aversion values when failing the attack is the only loss being considered. An RDD that is caught during screening or failing once in the country is considered a loss when the reference point is 0.5, and the increase in loss aversion amplifies the loss suffered by the attacker. Therefore, the deterrence region, when the fear reference point is 0.5, decreases in size as the loss aversion increases. Thus, attacker chooses a CW attack regardless of screening for more attribute weight combinations.

2.6 Discussion

Given our results and parameter choices, an attacker will only choose an RDD attack regardless of the defender's screening decision if almost all of their weight is on the fear attribute. In fact, a rational attacker will only choose this option if all the weight is on fear. The prospect theoretic attacker can have a small weight on cost or recruitment if:

- The probability of attack success once the RDD is in the country (q) is high
- The probability of the RDD being detected during screening (p) is low
- The attacker is quite likelihood insensitive (k is low)
- The attacker is more loss averse (λ is high)

However, it is unlikely that an attacker would choose an RDD regardless of screening unless they make multi-attribute decisions in a non-compensatory manner and fear is their most important (Tversky 1972) or most valid (Gigerenzer et al. 1999) attribute.

An attacker will choose a conventional weapons attack regardless of the defender's screening decision if most of their weight is on recruitment and/or cost, but not much weight is applied to fear. The prospect theoretic attacker can put more weight on fear (and hence less on recruitment and cost) if:

- The probability of attack success once the RDD is in the country (q) is low
- The attacker is more sensitive to likelihoods (k is high)
- The attacker is more loss averse (λ is high)

The probability of the RDD being detected during screening (p) does not affect the boundaries of the region for this option.

An attacker will be deterred by screening (choosing RDD if the defender doesn't screen and conventional weapons if the defender does screen) if they put significant weight on fear, but also have some weight on cost and recruitment. Thus, this pattern of choices is more prevalent in attackers who are making trade-offs between all three attributes. The region covers more combinations of weights (gets larger) if:

- The probability of attack success once the RDD is in the country (q) is high
- The attacker is quite likelihood insensitive (k is low)
- The attacker is less loss averse (λ is low)

The probability of the RDD being detected during screening (p) only affects the region as it increases the small area where the attacker chooses RDD regardless of screening if it is low.

These results show that the predictions of attacker behavior and the defender's optimal decision depend on the importance of the three attributes to the attacker and their level of likelihood insensitivity overall and loss aversion on each attribute. Experiments have shown that decision makers from numerous different backgrounds exhibit loss aversion and likelihood insensitivity. It would seem unwise to assume that terrorists are the one group that would not exhibit such effects and hence make purely rational decisions. Thus, assuming a single-attribute expected utility model for the attacker can lead to biased predictions of their decisions and hence non-optimal prescriptions for the defender.

However, to apply such methods we need to parametrize these models and it is not clear what parameter values we should use. In our example, we have used a wide range of values observed across a range of experimental populations. We cannot (or maybe choose not to) carry out behavioral decision experiments with active terrorists. So, at this point, assuming values from other experiments is the best we can do, but this is not sufficient for a fully accurate model that could be implemented by a practitioner. Nevertheless, this does not mean that this line of research should be ignored, and this work serves to demonstrate the need for more research in this area that will lead to important improvements in attacker/defender models.

2.7 Conclusions and Future Research

We have extended the work of Merrick and Leclerc (2016) on attacker modeling using single-attribute prospect theory to the multiple attribute case. We implemented the additive form from Bleichrodt et al. (2009) as it can represent different reference points for each attribute. The three attributes considered were maximizing economic impact (cost), instilling fear in the target population, and maximizing recruitment (i.e. ensuring the longevity of the terrorist organization), with our choices guided by the larger attribute sets from Keeney (2007), Keeney and von Winterfeldt (2010), and Seibert et al. (2016). In our analysis, we considered each attribute one at a time, before analyzing the impact of the weights applied to each attribute, the primary probabilities in the attacker decision trees, the level of likelihood insensitivity, the levels of loss aversion for each attribute, and the reference points that define the outcomes that are considered a loss in each attribute.

It is clear from our results that the impact of screening for nuclear material on the attacker's choice of attack is dependent on their preferences. Merrick and Leclerc (2016) found that the attacker's decision was sensitive to their level of loss aversion and likelihood insensitivity. We confirm this finding, but also add the additional dimension of the weight the attacker applies to each outcome attribute in their decision making. Our analysis has allowed a consideration of attackers who make explicit trade-offs, but we have also considered an attacker who makes non-compensatory decisions by applying all the weight to a single objective. The outcomes for cost, recruitment, and fear when considered alone make for different choices for the attacker. Using our parameter values, an attacker who considers only fear will choose an RDD attacker regardless of screening, while an attacker who considers only recruitment will choose conventional weapons regardless of screening. An attacker who considers only cost will be deterred by screening, an outcome that aligns with the results from Merrick and McLay (2010) and Merrick and Leclerc (2016) in their cost-only analysis.

The question remains whether terrorists make multi-attribute decisions in a compensatory or non-compensatory manner. If the former, then our methodology is appropriate. This would appear probable for a larger, more mature terrorist organization. However, when considering smaller terrorist groups or lone-wolf attacks then the findings of Einhorn (1970, 1971) and Tversky et al. (1988) suggest that non-compensatory models would be more appropriate. For this to be considered, the non-compensatory models (Simon 1955), Tversky 1972), Gigerenzer et al. 1999) must be extended to incorporate uncertainty and research questions remain about the interplay between likelihood insensitivity, loss aversion, and non-compensatory choices with multiple attributes. Further, as the prescriptive recommendations of attacker/defender models clearly depend on the choices of preference parameters for these descriptive decision models, we must estimate their values for specific attackers or terrorist organizations rather than use published experimental findings from very different populations.

Chapter 3: Adversarial Risk Matrices with an Expected Utility Framework

3.1 Introduction

A risk matrix is used to evaluate risk in many engineering applications. Risk matrices provide a decision maker with a risk ranking framework that evaluates the probability and consequence pairing of each risk scenario. The ease of use and ability to make quick decisions make risk matrices an attractive solution for industry use. Cox (2008a) states several shortcomings with the risk matrix approach, while concluding its discontinuation unlikely due to its widespread adoption. Ruan et al. (2015) improves risk matrices by integrating the risk attitudes of decision makers using expected utility theory. However, traditionally risk matrices have not considered adversarial risk and when used in counter-terrorism applications simply include the probability of a successful terrorist attack.

The Department of Homeland Security uses the Risk Analysis and Management for Critical Asset Protection (RAMCAP) framework when prioritizing the defense of US infrastructures against terrorist attacks (Cox 2008b). Risk = Threat X Vulnerability X Consequence formula is an important part of the RAMCAP framework. Cox (2008b) states potential limitations with the Risk = Threat X Vulnerability X Consequence formulation, as it is a single-level decision making process. Cox (2008b) concludes that an adversarial framework (a two-level decision making process) is necessary for counter-terrorism applications, where the defender predicts the attacker's action and then chooses the best response.

In this chapter, we bridge the work of Ruan et al. (2015) and Cox (2008b) to provide an adversarial risk matrix framework with an expected utility framework. The adversarial risk matrix approach starts with an attacker risk matrix, that considers the vulnerability of the target and its potential consequences. The attacker's risk ranking provides the threat level for the defender's risk matrix, which also includes the consequences of the attack. Thus, adversarial risk matrices are implemented by evaluating the Risk = Threat X Vulnerability X Consequence from an attacker/defender perspective.

3.2 Risk Matrices

A risk matrix is a two-dimensional table, with one axis representing the likelihood of an outcome, while the other representing its severity. Each probability and consequence combination that is represented by the risk matrix corresponds to a discretized risk level that represent the priority level of an outcome (Cox 2008a). Thus, the decision maker can prioritize risk posing events by assigning a discrete risk level to the probability and consequence pairs.

Ruan et al. (2015) consider the importance of incorporating risk attitudes of the decision makers by using their utility functions in determining the boundaries for regions representing the various risk ratings. Ruan et al. 2015 consider the utility function of the decision makers and introduce expected utility theory, EUT, to generate utility indifference curves to link the utility functions to the risk matrix. EUT is an integral part of decision making under uncertainty. EUT provides a prescriptive model for rational decision making. The utility function is used to quantify the risk attitudes. In this paper, we consider an attacker and a defender with varying risk attitudes.

Ruan et al. (2015) provides the following framework for the utility function. The consequence l is the independent variable and the utility value $u(l)$ is the dependent variable, where both l and $u(l)$ are negative. A transformation results in the scaling of the utility value, resulting in the range of $u(l)$ being $[-1,0]$. Ruan et al. (2015) next incorporate EUT and utility indifference curves to connect the utility functions to the risk matrix. The risk ratings of the decision makers are represented by the expected utility value of risk scenarios which correspond to probabilities and consequences on the risk matrix. The points appearing on the same utility indifference curves have the same expected utility, as they represent an equal amount of risk to the decision makers.

The discrete categories of risk ratings that are to be represented by the risk matrix are determined prior to establishing the utility indifference curves. A risk matrix with more risk ratings improves the successful categorization of multiple risk scenarios, since the ability to discriminate among outcomes are improved by providing additional constraints. The labeling of risk ratings as discrete categories, and the initial reference points for the risk ratings on the probability-consequence coordinate system are necessary in categorizing areas that represent the varying risk ratings.

We denote the initial reference point for the probability value that corresponds to a given risk level as p' and the consequence value as l'. We further assume an an arbitrary point, with a probability value p and a consequence value l, which coincides on the same utility indifference curve as the initial reference point. Since the initial reference point and the arbitrary point have the same expected utility value, the equality of the two utilities can be expressed as:

$$pu(l) + (1 - p)u(0) = p'u(l') + (1 - p')u(0) \tag{2}$$

where $u(0) = 0$; hence,

$$pu(l) = p'u(l') \tag{3}$$

or

$$p = \frac{u(l')}{u(l)}\, p' \tag{4}$$

The initial reference point probability value, p', is required to be 1, so that the decision maker only needs to determine the initial reference loss value, l', in generating the utility indifference curves. A risk matrix that has N risk ratings, requires $N - 1$ utility indifference curves. The N areas created by the $N-1$ indifference utility curves are then assigned to each risk level to establish the risk matrix.

3.3 Adversarial Risk Matrices

We provide an elementary example of the attacker/defender, adversarial, 3 by 3 risk matrices. The implementation of the decision makers' utility function is not considered in this section, as it will be presented in Section 3.4.

The Department of Homeland Security introduced a standard risk assessment formula in 2007 to determine the security risks encountered at chemical plants in the United States. The formula they used was

$$Risk = Threat\ x\ Vulnerability\ x\ Consequence \tag{5}$$

where risk is the potential for loss from an attack, threat is the probability that the attacker chooses to attack, vulnerability is the probability that the attacker is successful with the chosen attack, and consequence is quantifiable loss generated by the attack. Cox (2008b) considers the limitations of this formulation and discusses intelligent decision making, where the defender considers the attacker's decision prior to implementing a countermeasure, as a possible improvement. Thus, we consider equation (5) from the perspective of the attacker and defender in formulating an adversarial risk matrix with intelligent decision making.

We first consider the attacker's decision making process. The attacker wishes to maximize the damage suffered by the defender and focuses on the vulnerability of the defender and the consequence of the attack. We assume that vulnerability and consequence have three levels: low, medium, and high. We further assume that the attacker has three risk ratings, and that the attacker is considering five different attacks, of which one will be executed. The risk ratings are color designated, and they are light grey, medium grey, and dark grey, most preferred to least preferred, respectively. The various attacks are denoted a_1, a_2, a_3, a_4, and a_5. The attacker risk matrix is provided in Figure 3.1.

High	a_1	a_3	a_4
Medium			
Low	a_2		a_5
Vulnerability/Consequence	High	Medium	Low

Figure 3.1: The attacker risk matrix

The defender considers the attacker's risk matrix before implementing safeguards against potential attacks. The defender focuses on the probability that the attacker attacks (i.e. the threat) and the consequence of the attack. We assume that the defender, like the attacker, has three levels for threat and consequence: low, medium, and high. The defender risk matrix in response to the attacker risk matrix is provided in Figure 3.2. Like the attacker, the defenders' risk ratings are color designated, and they are light grey, medium grey, and dark grey, most preferred to least preferred, respectively.

High	a_1	a_3	
Medium	a_2		a_4
Low			a_5
Threat/Consequence	High	Medium	Low

Figure 3.2: The defender risk matrix

The attacker and defender risk matrices demonstrate the link between the vulnerability and the threat. For instance, the attacker prefers the attacks a_1 and a_3, color coded light grey, to all of the remaining attacks. Thus, it is reasonably assumed that the attacker would choose to carry out a_1 and a_3 with a higher probability than the remaining attacks, making the threat of such attacks a high risk-level in the defender risk matrix. Furthermore, the attacker prefers the attacks a_2 and a_4, color coded medium grey, to a_5, color coded dark grey. The defender; therefore, assumes that the attacker would carry out a_2 and a_4 with a probability lower than a_1 and a_3 but higher than a_5, thus assigning a_2 and a_4 a medium threat level. Finally, the attacker least prefers a_5, making it a low threat level to the defender.

The attacker's preference levels for the various attacks, the color-coded risk ratings in the attacker risk matrix, enables the defender to assign a threat level to each attack. The consequence is constant in both risk matrices. The defender's risk level assignment to the threat, consequence pairs places the various attacks in bins based on the defender's preferences, but the defender's risk rating assignments does not have an impact on the vulnerability/threat link between the adversarial risk matrices.

3.4 An EU Framework for Adversarial Risk Matrices

The discrete number of risk ratings, the boundary values for the consequence/probability pairs, and the consequence reference values to be used in equation (4) are determined by the decision maker in the risk matrix establishment process. The consequence values in the adversarial risk matrices are scaled to range from 0 to -5. An attacker and a defender with eight risk ratings are considered. The boundary values for risk loss ratings for the attacker and the defender are assumed to be (-0.25, -0.75, -1.25, -1.75, -2.25, -2.75, and -3.25). The consequence reference values to be used in the utility indifference curves are assumed to be the same as consequence boundary values.

The attacker and the defender are both assumed to be risk averse, and we parametrize their level of risk aversion, r_a and r_d respectively. The utility functions of the attacker and the defender for risk loss l are $u_a(l) = 1 - e^{\frac{l}{r_a}}$ and $u_d(l) = e^{\frac{l}{r_d}} - 1$, respectively. The utility indifference curves for the attacker, p$_a$, and the defender, p$_d$, are represented by using Equation 4:

$$p_a = \frac{1-e^{l'/r_a}}{1-e^{\frac{l}{r_a}}}\, p' \tag{6}$$

$$p_d = \frac{e^{l'/r_d-1}}{e^{l/r_d}1}\, p' \tag{7}$$

The utility indifference curves link the risk matrices and the utility functions. Each consequence and probability of success pair corresponds to a discretized risk rating of the attacker. The defender, consequently, models the defender risk matrix by assigning a threat level to the attacker's probability of success and consequence pairs. The defender achieves the assignment of threat levels by using the Luce model (Luce 1977) and considering all possible attacks by the attacker. The Luce model for choice, states that the probability of selecting an option, i, out of a pool of options is

$$P(i) = \frac{w_i}{\sum_i w_i} \tag{8}$$

where w_i is the weight assigned to the i-th item in the pool. We apply the Luce model to the expected utilities of the attacker to calculate the probability of an attacker choosing an attack from a pool of attacks – thus assigning a threat level and subsequently a risk rating to each attack on the defender matrix. The Luce model implies that an attacker will choose the i-th attack from a pool of attack options with probability:

$$P(i) = \frac{EU_a(i)}{\sum_i EU_a(i)} \tag{9}$$

We assume that the attacker will execute at least one of the attacks in consideration, and the attack that is most likely to be executed than any other attack warrants the highest threat level attainable with the other attacks scaled accordingly. Thus, the highest threat level (calculated using the Luce model) from the pool of attacks that the attacker is considering assigned a value of 1.0, and the remaining attacks are scaled accordingly to maintain proportionality.

3.5 Application

We consider an attacker, a terrorist organization, who has five different attack options, and a defender, a risk analyst working for the US government, with the objective of minimizing the damage inflicted by the attacker. The assumptions specified in Section 3.4 regarding the utility function of the attacker/defender and the settings of the risk matrix hold true. The damage of the attacks is evaluated on a monetary basis (in dollars). Short term damage inflicted on infrastructure and human life are the criterions considered in the valuation of the attacks. Long term impacts such as lost wages, reduction in commerce and tourism, psychological impacts on the masses are not considered to simplify the damage assessment process. To further simplify the damage assessment of the attacks, we assume that the attacks will result in only casualties, and not wounded survivors. We assign human life an economic value of \$5.8 million, as suggested by the U.S. Department of Transportation (2005). The list of attacks considered by the attacker, the economic damage and the probability of success (vulnerability) for each attack is provided in Table 3.1.

Table 3.1: Specifics of Each Attack Considered by the Attacker

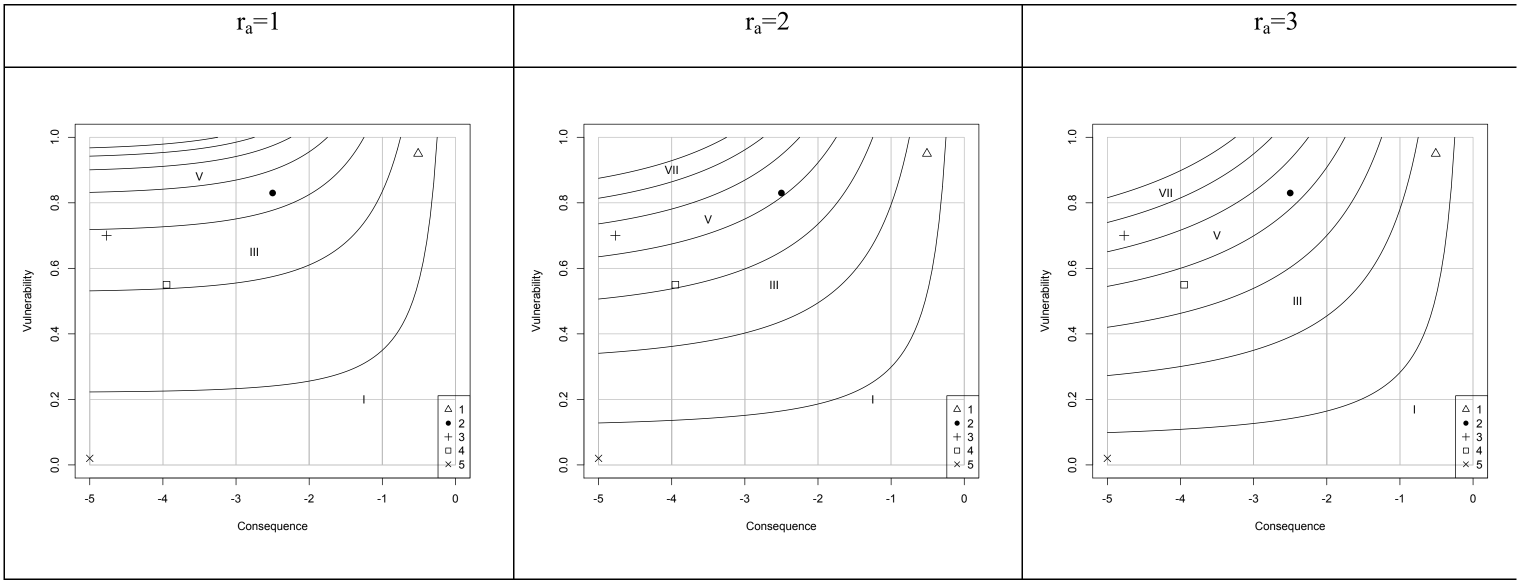
ID	Attack	Casualties (# of People)	Economic Damage (in Millions of $)	Vulnerability (%)	Consequence (Scaled Economic Damage)
1	Single person attack	50	300	95	-0.51
2	Terror cell attack (3-5 people)	225	1468	83	-2.50
3	Multi-terror cell attack (8-10 people)	430	2800	70	-4.77
4	Anthrax attack	400	2320	55	-3.95
5	Explosives on a plane (Boeing 777)	451	2936	2	-5.00

Economic damage in Table 3.1 does not consider the cost accrued by the attacker in generating the attack. The attacker the resources utilized in the attacks are sunk costs prior to the gains/loss evaluation process, thus they are not regarded as retroactive losses following an unsuccessful attack. An unsuccessful attack yields a gain of 0 for the attacker.

The probability that each attack is carried out successfully, i.e. the vulnerability, decreases as the number of people carrying out the attack increases and/or the type of attack becomes more complex. This is intuitive, since the probability of the communications between individual attackers being intercepted increases with more parties having to share information in planning the attack. Similarly, the complexity of the attack impacts the likelihood that the attack is carried out successfully. We assume that the single person attack, terror-cell attack, and multi-terror cell attacks share the same level of complexity, as we assume they will be carried out with assault rifles and homemade bombs. These three attacks do not have the same vulnerability level; however, since the attacks consist of varying number of people performing the attacks. We assume that the anthrax attack and placing explosives on a plane are highly complex attacks, as these attacks require higher level of preparation – thus impacting the probability of success for each attack.

3.5a The Attacker Risk Matrix

Given the consequence and vulnerability values of each attack in Table 3.1, and the risk matrix construction information along with the attacker/defender utility function in Section 3.4, we generate Figure 3.3, The Attacker Risk Matrix for Varying Risk Aversion Values of $r_a = 1, 2, and\ 3$. The risk ratings are labeled I through VIII, with VIII being the most preferable outcome for the attacker. The attacks are designated various shapes using the attack ids provided in Table 3.1.



r_a=1 | r_a=2 | r_a=3

Figure 3.3: The Attacker Risk Matrix for Varying Risk Aversion Values

An r_a of 1 for risk averseness implies risk neutrality, while values of 2 and 3 imply increasing risk averseness. The utility indifference curves for the risk neutral attacker are flatter than those of the risk averse attacker, since the risk averse attacker becomes increasingly sensitive over vulnerability, consequence pairings as the absolute magnitude of consequences increases. For example, the utility indifference curve that separates risk ratings II and III, is indifferent for the risk neutral attacker over high magnitude consequences, as the risk neutral attacker is indifferent between attacks with vulnerability, consequence pairings of 0.53, -5, and 0.56 and -3. As the attacker becomes more risk averse, the attacker exhibits increasing sensitivity to the vulnerability, consequence pairings. For the same utility indifference curve that separates risk ratings II and III, an attacker exhibiting risk aversion ($r_a = 3$), is indifferent between attacks with vulnerability, consequence pairings of 0.27, -5 and 0.35, -3 respectively.

Three of the five attacks: the terror cell attack, the multi-terror cell attack, and the anthrax attack correspond to varying risk ratings as the risk aversion parameter, r_a, changes. The highly probable outcome paired with a low consequence, the single person attack, and the highly improbable outcome paired with a high consequence, the explosives on a plane attack, are not impacted by the shifting in the utility indifference curves as the risk aversion varies. A risk neutral attacker prefers a terror cell attack over all attacks, while preferring multi-cell terror attack and an anthrax attack equally. As risk averseness increases, the attacker exhibits a more pronounced trade-off between probability, consequence pairings with consequence values being assigned more weight than the probability counterpart. For example, an attacker ($r_a = 3$), prefers a multi-cell terror attack over a cell terror attack – a switch of preference when compared to a risk neutral attacker. Thus, a risk averse attacker would rather choose an attack with a slightly lower probability of success (i.e. vulnerability) and a much higher consequence. The probability of attacks being successful is considered much more thoroughly with the respective consequence level when risk averseness increases.

3.5b The Defender Risk Matrix

The expected utility of each attack is used in the Luce model to calculate the probability that the attacker carries out each individual attack given the array of available attacks. The calculated probability is referred to as threat in the defender risk matrix. The consequence of each attack remains the same. The expected utility values for the attacker risk aversion values of 1, 2, and 3 and the corresponding scaled threat values are presented in Table 3.2.

Using the consequence and threat values of each attack in Table 3.2 for the risk neutral attacker ($r_d = 1$), and the risk matrix construction information along with the attacker/defender utility function in Section 3.4, we generate Figure 3.4, The Defenders Risk Matrix for Varying Risk Aversion Values of $r_d = 1, 2, 3$. The risk ratings are labeled I through VIII, with I being the most preferable outcome for the defender. The attacks are designated various shapes using the attack ids provided in Table 3.1.

Table 3.2: Specifics of Each Attack from the Defender Perspective

ID	Attack	Vulnerability (%)	Consequence (Scaled Economic Damage)	EU for the Attacker (r_a=1)	EU for the Attacker (r_a=2)	EU for the Attacker (r_a=3)	Scaled Threat (r_a=1)	Scaled Threat (r_a=2)	Scaled Threat (r_a=3)
1	Single person attack	95	-0.51	0.38	0.21	0.15	0.50	0.34	0.27
2	Terror cell attack (3-5 people)	83	-2.50	0.76	0.59	0.47	1.00	0.93	0.84
3	Multi-terror cell attack (8-10 people)	70	-4.77	0.69	0.64	0.56	0.91	1.00	1.00
4	Anthrax attack	55	-3.95	0.54	0.47	0.40	0.71	0.75	0.72
5	Explosives on a plane (Boeing 777)	2	-5.00	0.02	0.02	0.02	0.03	0.03	0.03

Like the attacker risk matrix, the defender risk matrix which considers a risk neutral attacker ($r_d = 1$ in Figure 3.4) exhibits increasing sensitivity to the threat levels paired with the consequence levels, as the absolute magnitude of consequence increases. For example, a risk neutral defender ($r_d = 1$) is indifferent between a terror cell attack (a threat of 1.0 and a consequence of 2.5) and multi-terror cell attack (a threat of 0.91 and a consequence of -4.77). A risk averse defender ($r_d = 3$); however, prioritizes the multi-terror cell attack over all attacks. The highly improbable but highly consequential event of explosives on a plane, and the highly probable but highly inconsequential event of a single person attack are assigned a risk rating of I across all risk aversion values for the defender.

The defender's prioritization of events using the risk rating system is highly impacted as the risk aversion increases and the events being considered are similar in threat but different in consequence. A risk neutral defender shows indifference to attacks sharing high treat level paired with significantly different consequence levels (i.e. -2.5 and -4.77). As the defender becomes more risk averse, the defender further differentiates risks associated with attacks sharing similar threat levels. Attacks with high threat and high consequence levels are most impacted by increasing risk aversion, with the assigned risk rating increasing as the risk aversion parameter, r_d, increases.

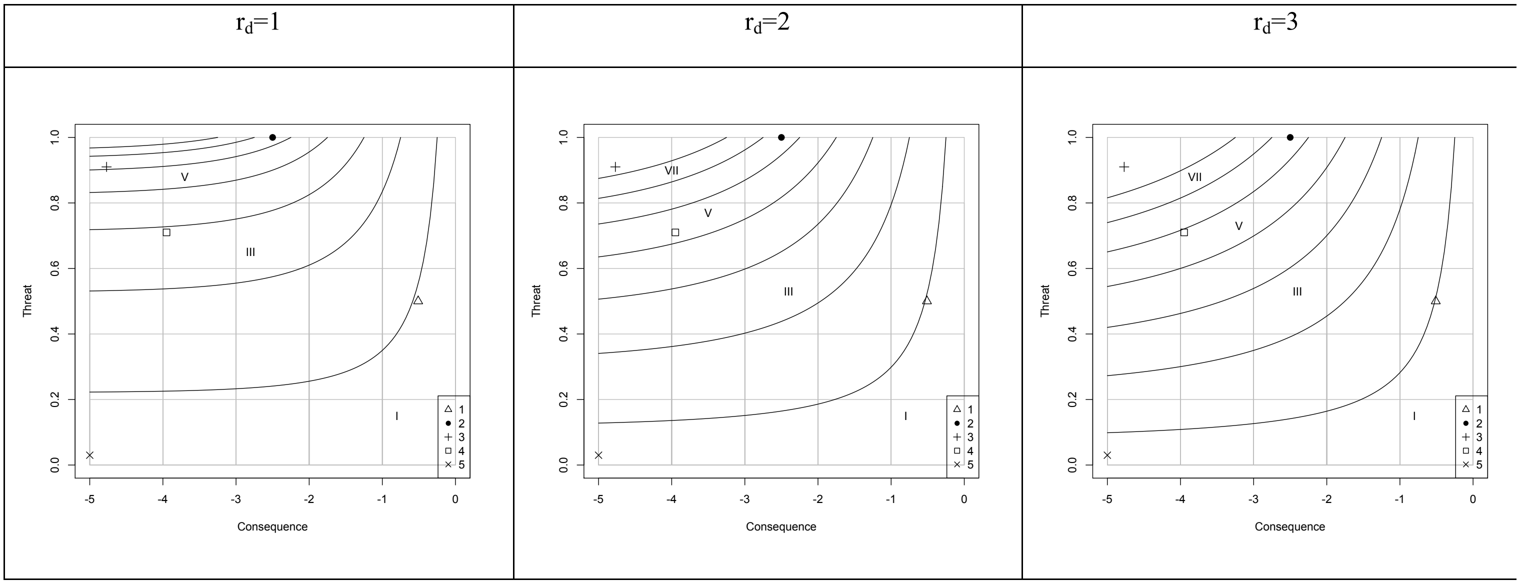


Figure 3.4: The Defender Risk Matrix for Varying Risk Aversion Values (r_a=1)

3.5c The Defender Risk Matrix for varying Attacker/Defender Risk Aversion Values

The defender risk matrix for varying attacker/defender risk aversion values are provided in Figure 3.5. The consequence and threat level of each attack, characterized by the risk aversion of the attacker in Table 3.2, are used to plot each threat, consequence pair. The risk aversion of the defender is then considered in generating the utility indifference curves. Since the utility indifference curves are generated solely by the defender's risk aversion, the utility indifference curves are the same when the defender risk aversion is the same. Hence, all the plots in Figure 3.5 that fall in the same column share the same utility indifference curves. Similarly, all the plots in Figure 3.5 that fall in the same row share the same threat level, since the threat levels are solely dependent on the attacker's risk aversion. The consequence values of the attacks remain constant across all defender risk matrices.

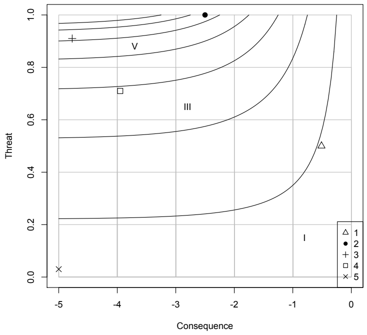
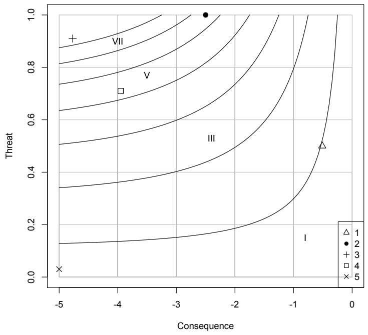
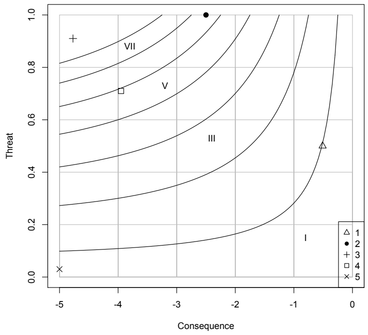
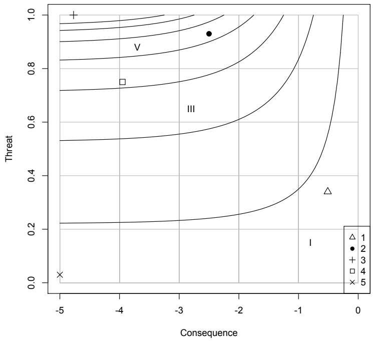
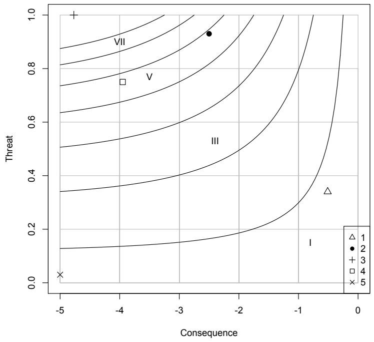
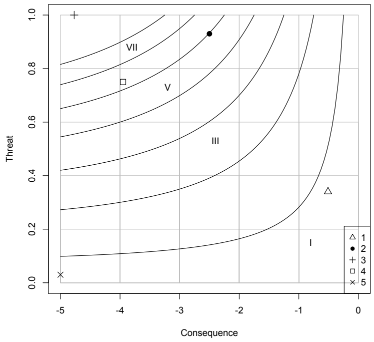
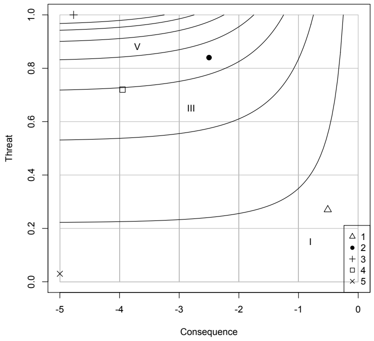
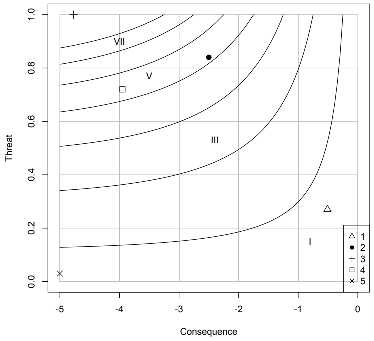
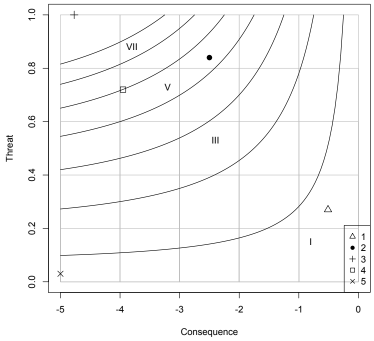
Attacker/ Defender Risk Aversion	r_d =1.0	r_d =2.0	r_d =3.0
r_a =1.0			
r_a =2.0			
r_a =3.0			

Figure 3.5: The Defender Risk Matrix for Varying Risk Aversion Values

The threat level for a given attack is calculated by dividing the attacker's EU for that attack by the sum of the attacker's EUs for all available attacks, then scaling the result so that the attack with the highest threat from all available attacks is assigned a value of 1.0. Since the vulnerability of the attacks remain constant, the attacker's risk aversion is the parameter that impacts both the numerator, attacker's EU for a given attack, and the denominator, the sum of attacker's EU for all available attacks. Since the threat value requires considering a given attack on both a proportional and a scaling basis to the other attacks, a rule of thumb for how the threat will vary for a given risk aversion of the attacker is not feasible. Contrarily, the impact of the defender's risk aversion values on a given risk aversion value for the attacker are predictable. As the risk aversion of the defender increases, the defender's utility curves become increasingly sensitive over high threat, high consequence attacks. The defender's sensitivity in considering trade-offs between attacks sharing similar threat levels with varying consequence levels increases as the defender's risk aversion parameter, r_d, increases. Thus, for a given r_a where the threat levels care constant, increases in r_d results in risk rating assignments that are equally or more severe than those with lower r_d.

Figure 3.6, The Defender's Risk Ratings for Varying Attacker/Defender Risk Aversion Values, provides a condensed version of Figure 3.5 – eliminating the utility indifference curves to display the defender's risk rating of each attack for all r_a, r_d combinations. In the case of the defender, excess risk aversion can be detrimental since an unnecessarily high risk rating could mean a costly misappropriation of resources. In Figure 3.6, we observe that all attacks for varying risk aversion and likelihood insensitivity values for the attacker correspond to a risk rating equal to or higher than those assigned to a less risk averse defender as the defender becomes more risk averse. Thus, for any give value of r_a and k, if $r_d > r_d'$ then the defender's $EU_d > EU_d'$, thus the risk rating is also higher.

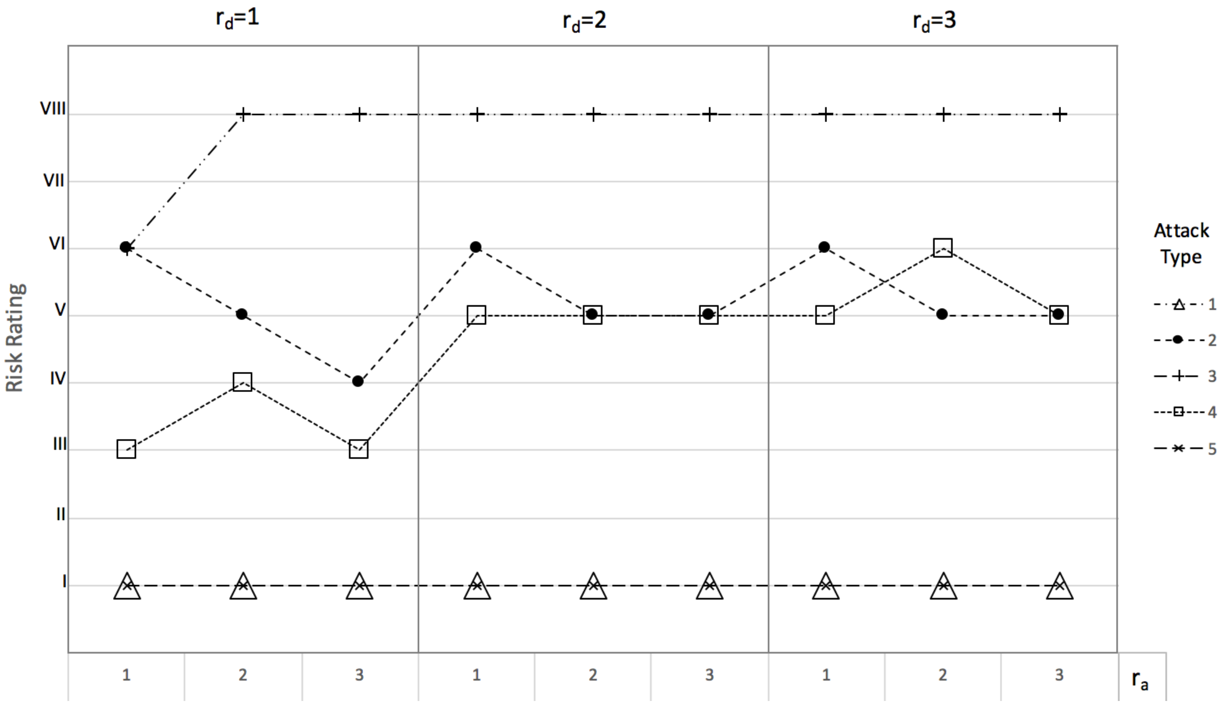


Figure 3.6: The Defender's Risk Ratings for Varying Attacker/Defender Risk Aversion Values

We observe that a single person attack and explosives on a plane attack yield a risk rating of I for every r_a, r_d combination. The threat level of the explosives on a plane attack (3%) and the consequence level of a single wolf attack (-0.51) are extremely inconsequential that the shifts in the utility indifference curves caused by the changes in r_d and the shifts in the threat levels caused by the changes in r_a do not result in a more severe risk rating. Thus, the defender should only allot resources that are consistent with a risk rating of I in regards to the single person and explosives on a plane attacks.

The risk aversion of the defender has a significant impact on attacks with high consequence and high threat levels, since the utility indifference curves on the defender risk matrix (Figure 3.5) becomes increasingly refined as r_d increases. In addition to r_d, r_a impacts risk rating assignments, as the threat level for each attack is a function of the attacker's EU as calculated in the Luce model – which is directly impacted by the attacker's risk aversion. For example, a multi-terror cell attack and a terror cell attack share a risk rating of VI when both the attacker and the defender are risk neutral. As the risk aversion increases for either the attacker, the defender or both, the multi-cell terror attack gets assigned a constant risk rating of VIII. If the defender understands that being risk averse ($r_d > 1$) is detrimental, the defender must determine whether the attacker is risk averse, as the attacker's risk aversion will mandate the resources allotted in defending the multi-terror cell attack.

Having established the benefits of the defender's risk neutrality ($r_d = 1$), we observe that misappropriation of resources can also result from incorrectly evaluating the attacker's risk aversion, r_a. For example, the risk ratings are at the lowest value for all attacks except for the terror cell attack when $r_d = 1$ and $r_a = 1$. Similarly, the risk ratings are at the lowest value for all attacks except for the multi-terror cell attack when $r_d = 1$ and $r_a = 3$. Hence, if a defender believes that the attacker is risk neutral when the attacker is in fact severely risk averse ($r_a = 3$), the multi-cell terror attack would be inadequately defended while the single-cell terror attack would be excessively defended. Thus, the defender's understanding of the attacker is instrumental in ensuring the appropriate assignment of resources for a given attack. Using Figure 3.6, a defender at the very least would know not to assign a risk rating more than 1, 6, 8, 4, and 1 for the single person attack, terror cell attack, multi-terror cell attack, anthrax attack and explosives on a plane attack, respectively – as these values are the most extreme risk rating assignments when the risk aversion of the attacker is unknown and the defender is risk neutral.

3.6 Conclusions

We have introduced adversarial risk matrices to provide a two-level decision making process where a defender evaluates decisions available to an attacker then translates the attacker's actions into a discretized risk rating system using a defender risk matrix. The Risk=Threat X Vulnerability X Consequence formula, which is widely used by the Department of Homeland Security in assessing risk, was broken down to vulnerability/consequence and threat/consequence pairings to generate the attacker risk matrix and the defender risk matrix, respectively. The Luce model was used to link the attacker and defender risk matrices.

The adversarial risk matrices accounts for the risk aversion exhibited by the attacker and the defender, since the risk aversion parameter was used in the attacker/defender utility functions. The risk aversion of the defender impacts the shape of the utility indifference curves while the risk aversion of the attacker impacts the threat level for each attack on the defender risk matrix. As the defender becomes more risk averse for a given risk aversion of the attacker, the attacks receive a risk rating equal to or higher than that of a less risk averse defender. The defender provides more than or equal to what is necessary to defend against an attack as the risk aversion of the defender increases. Therefore, maintaining risk neutrality is the optimal solution for a defender guarding against an attacker.

In addition to the risk aversion of the defender, the defender's perception of the attacker, namely the risk aversion of the attacker, is instrumental in appropriate resource allocation for a given attack. For example, if the defender believes the attacker is risk neutral when the attacker is in fact severely risk averse, the defender could allocate resources to excessively defend against some attacks while others are inadequately defended. The defender risk matrix can assist the defender in performing scenario analysis, thus uncovering hidden trade-offs among various defense strategies.

Chapter 4: Adversarial Risk Matrices with a Prospect Theoretic Framework

4.1 Introduction

In Chapter 3, we introduced adversarial risk matrices to model sequential counter-terrorism decisions and provided an EUT framework to model the risk preferences of both the attacker and defender. Specifically, we assumed utility functions for the attacker and defender, $u_a(l) = 1 - e^{\frac{l}{r_a}}$ and $u_d(l) = e^{\frac{l}{r_d}} - 1$, respectively. The utility indifference curves for the attacker, p_a, and the defender, p_d, were represented by:

$$p_a = \frac{1 - e^{l'/r_a}}{1 - e^{\frac{l}{r_a}}}\, p' \tag{1}$$

$$p_d = \frac{e^{l'/r_{d-1}}}{e^{l/r_d 1}}\, p' \tag{2}$$

The utility indifference curves link the risk matrices and the utility functions. We used the Luce model to find the probability or frequency that the attacker would choose a given attack given the attacker's expected utility for each attack:

$$P(i) = \frac{EU_a(i)}{\sum_i EU_a(i)} \tag{3}$$

We then scaled the probabilities of each attack so that the attack with the highest likelihood being executed corresponds to the highest threat possible, 1.0, with the other attacks being scaled accordingly to maintain proportionality among attacks.

In Chapter 3, we concluded that an accurate description of the attacker has a direct impact on the appropriate allocation of resources, thus it is ideal to provide the defender the best descriptive model of the attacker. As previously mentioned in Chapter 2, expected utility theory (EUT) fails to adequately work as a descriptive model and, therefore, EUT could be questioned in applicability for modeling the attacker's preferences. Three common areas where EUT falls short in describing human decision making include: reference dependence, probabilistic insensitivity, and rank dependence. The prospect theoretic approach by Kahneman and Tversky provides a framework that can address systematic deviation of human behavior from EUT (Kahneman and Tversky 1979). Thus, a more appropriate approach to adversarial risk matrices would be to model the attacker decisions using prospect theory, and prescribe a prioritization of attacks to the defender using EUT.

The issue of reference dependence does not need to be considered in regards to the attacker in our prospect theoretic approach. The defender deems the resources utilized in the attacks as sunk costs; therefore, the resources spent towards generating the attacks are not regarded as retroactive losses following an unsuccessful attack. Thus, a gain/loss framework is not necessary.

As for issues surrounding probabilistic insensitivity, the prospect theoretic approach in describing the attacker is justified. Likelihood insensitivity is when a decision maker overweighs probabilities that are small, underweights probabilities that are large, while remaining insensitive to probabilities that are near the middle (50%). Empirical research suggests that likelihood insensitivity is common and that prospect theoretic modeling is appropriate to use in descriptive modeling (Kahneman and Tversky 1979).

4.2 Methodology

Similar to our approach in Chapter 2, we implement the neo-additive probability weighting function used by Chateauneuf et al. (2007),

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

where $0 < k < 1$. A $k = 1$ indicates the absence of probability weighting. Baillon et al. (2014) suggests that a value of k between 0.6 and 0.8 should be adopted when evaluating the impact of likelihood insensitivity for the neo-additive probability weighting function. Thus, we consider values of 0.6, 0.7, and 0.8 in modeling the attacker's likelihood insensitivity.

The reference dependence of the attacker is not considered within our prospect-theoretic descriptive framework, since we assume that the attacker evaluates each attack as a part of a gains framework regardless of whether the attack was successful. The attacker considers the resources used to generate each attack to be an investment, and an unsuccessful attack yields a gain of 0. The resources utilized in the attacks are "spent" prior to the gains/loss evaluation process, thus they are not regarded as retroactive losses following an unsuccessful attack.

The prospect-theoretic approach in modeling the attacker does not have an impact on the utility indifference curves used in the defender risk matrix. However, there is an impact on the expected utility of each attack, since the utility of each attack is multiplied by the weighted probability when calculating expected utility. Since, the expected utility of each attack is used in the Luce model to calculate the proportional EU of the attacks, the prospect-theoretic attacker yields varying threat levels on the defender risk matrix. The varying combinations of k, risk aversion of the attacker (r_a), the risk aversion of the defender ($\mathrm{r_d}$), and the corresponding risk rating for each attack are explored in the Results section.

4.3 Application

We mirror the example used in Chapter 3 and extend our work by describing the attacker in a prospect-theoretic manner. We consider an attacker, a terrorist organization, who has five different attack options, and a defender, a risk analyst working for the US government, with the objective of minimizing the damage inflicted by the attacker. The assumptions specified in Chapter 3 regarding the utility function of the attacker/defender and the settings of the risk matrix hold true. The damage of the attacks is evaluated on a monetary basis (in dollars). Short term damage inflicted on infrastructure and human life are the criterions considered in the valuation of the attacks. Long term impacts such as lost wages, reduction in commerce and tourism, psychological impacts on the masses are not considered to simplify the damage assessment process. To further simplify the damage assessment of the attacks, we assume that the attacks will result in only casualties, and not wounded survivors. We assign human life an economic value of $5.8 million, as suggested by the U.S. Department of Transportation (2005).

The Risk = Threat X Vulnerability X Consequence formula, which is widely used by the Department of Homeland Security in assessing risk, provides the foundation of the attacker and the defender risk matrices with vulnerability/consequence and threat/consequence pairings, respectively. The Luce model translates the vulnerability/consequence pairings to a threat level on on the defender risk matrix. The utility of each attack for the attacker that is used in the Luce model is multiplied by the weighted probability of the respective attack, with the likelihood insensitivity depending on the neo-additive probability weighting function parameter, k. We consider the values of 0.6, 0.7, 0.8, and 1.0 for k. A k of 1.0 describes an attacker without likelihood insensitivity. As the value of k decreases, the likelihood insensitivity of the attacker increases. The list of attacks considered by the attacker, the probability of success for each attack (the vulnerability), the expected utility of each attack for a given risk aversion by the attacker, and the corresponding probability of the attacker carrying out each attack (the threat) are provided for the various values of the probability weighting parameter (k) in Table 4.1.

Table 4.1: The Specifics of Each Attack from the Defender Perspective

ID	Probability Weighting (k)	Attack	Vulnerability (%)	Consequence (Scaled Economic Damage)	EU for the Attacker (r_a=1)	EU for the Attacker (r_a=2)	EU for the Attacker (r_a=3)	Threat (r_a=1)	Threat (r_a=2)	Threat (r_a=3)
1	0.6	Single person attack	77.0	-0.51	0.31	0.17	0.12	0.48	0.30	0.24
2	0.6	Terror cell attack (3-5 people)	69.8	-2.50	0.64	0.50	0.39	1.00	0.89	0.80
3	0.6	Multi-terror cell attack (8-10 people)	62.0	-4.77	0.61	0.56	0.49	0.95	1.00	1.00
4	0.6	Anthrax attack	53.0	-3.95	0.52	0.46	0.39	0.81	0.82	0.80
5	0.6	Explosives on a plane (Boeing 777)	21.2	-5.00	0.21	0.19	0.17	0.33	0.34	0.35
1	0.7	Single person attack	81.5	-0.51	0.33	0.18	0.13	0.49	0.31	0.25
2	0.7	Terror cell attack (3-5 people)	73.1	-2.50	0.67	0.52	0.41	1.00	0.90	0.80
3	0.7	Multi-terror cell attack (8-10 people)	64.0	-4.77	0.63	0.58	0.51	0.94	1.00	1.00
4	0.7	Anthrax attack	53.5	-3.95	0.52	0.46	0.39	0.78	0.79	0.76
5	0.7	Explosives on a plane (Boeing 777)	16.4	-5.00	0.16	0.15	0.13	0.24	0.26	0.25
1	0.8	Single person attack	86.0	-0.51	0.34	0.19	0.13	0.49	0.32	0.25
2	0.8	Terror cell attack (3-5 people)	76.4	-2.50	0.70	0.55	0.43	1.00	0.92	0.81
3	0.8	Multi-terror cell attack (8-10 people)	66.0	-4.77	0.65	0.60	0.53	0.93	1.00	1.00
4	0.8	Anthrax attack	54.0	-3.95	0.53	0.47	0.40	0.76	0.78	0.75
5	0.8	Explosives on a plane (Boeing 777)	11.6	-5.00	0.12	0.11	0.09	0.17	0.18	0.17
1	1.0	Single person attack	95	-0.51	0.38	0.21	0.15	0.50	0.34	0.27
2	1.0	Terror cell attack (3-5 people)	83	-2.50	0.76	0.59	0.47	1.00	0.93	0.84
3	1.0	Multi-terror cell attack (8-10 people)	70	-4.77	0.69	0.64	0.56	0.91	1.00	1.00
4	1.0	Anthrax attack	55	-3.95	0.54	0.47	0.40	0.71	0.75	0.72
5	1.0	Explosives on a plane (Boeing 777)	2	-5.00	0.02	0.02	0.02	0.03	0.03	0.03

4.4 Results

The defender risk matrices for the various combination of r_a (risk aversion of the attacker), r_d (risk aversion of the defender), and k (likelihood insensitivity parameter) of 0.6, 0.7, 0.8, and 1.0 are provided in Figures 4.1, 4.2, 4.3, and 4.4, respectively. The various values for k which highlights the likelihood insensitivity exhibited by a prospect-theoretic attacker does not impact the defender's utility indifference curves. The utility indifferences curves are function of r_d, and the boundaries that define each risk rating which the defender uses to assign risk ratings are identical for a given r_d, hence the utility indifference curves for the defender are identical to our results in Chapter 3.

While the defender risk matrices' utility indifference curves are identical for the prospect-theoretic and the EU attacker for any given r_d, the varying values of k provide insight on how a prospect-theoretic attacker fundamentally differs from a (EU) attacker in respect to probability of carrying out each attack, i.e. the threat. The threat level is a scaled measure of the proportion of an attack's expected utility in comparison to the summation of the expected utility of all attacks being considered. The expected utility is a multiplicative calculation involving the utility of each attack and the probability of success for that respective attack (i.e. the vulnerability). Vulnerability measurements are directly impacted by the likelihood insensitivity of the attacker. Likelihood insensitivity causes an attacker to overweigh low probabilities and underweight high probabilities, thus significantly distorting the probability weighting of the explosives on a plane (vulnerability=2%) attack and the single person attack (vulnerability=95%). An attacker described by a lowest value of k (i.e. 0.6) exhibits the highest level of likelihood insensitivity while an attacker described by the highest value of k (i.e. 1.0) does not exhibit likelihood insensitivity.

Table 4.1, The Specifics of Each Attack from the Defender Perspective, provides a numerical insight of likelihood insensitivity's impact on the threat levels, while Figures 4.1, 4.2, 4.3, and 4.4 provide a visual insight of defender's prioritization of attacks using risk matrices. The attack that is most impacted by attacker's likelihood insensitivity is the explosives on the plane attack, since this attack poses the highest level of consequence (-5) while being paired with the lowest level of vulnerability (2%) in the absence of likelihood insensitivity. As the likelihood insensitivity changes from the lack of likelihood insensitivity ($k = 1.0$) to 0.8, 0.7, and 0.6, the attacker's impression of a successful explosives on a plane attack increases from 2% to 11.6%, 16.4%, and 21.2%, respectively. The increase in vulnerability increases the EU of this attack, thus increasing the threat level associated with the explosives on the plane attack.

In Chapter 3, prior the implementing a prospect-theoretic attacker, the threat level of attacks changed with shifts in the risk aversion of the attacker, r_a, and the utility indifferences curves changed with shifts in the risk aversion of the defender, r_d. With the implementation of a prospect-theoretic attacker, we observe that changes in likelihood insensitivity significantly impacts the threat of attacks with high and low probabilities for vulnerability. In the case of the explosives on a plane attack, the threat level from an EU attacker ($k = 1$) is a constant measure of 3%, yielding a defender's risk rating of I for all r_a, r_d combinations in the defender risk matrix of Figure 4.4. The same attack paired with the most extreme likelihood insensitivity ($k = 0.6$) yields a risk rating of II, III, and III for a risk averse attacker ($r_a = 3$) across r_d=1, 2, and 3, respectively. The attacker's perception of increased success with respect to an attack with a high consequence increases the likelihood of the attacker carrying out the respective attack, thus increasing the threat level and subsequently the attack's risk rating on the defender risk matrix.

Contrarily, an attack with high probability of success, namely the single person attack, likelihood insensitivity does not have as pronounced impact as the low probability success attack, namely the explosives on a plane attack. The single person attack has a consequence of -0.51 and the highest level of vulnerability (95%) in the absence of likelihood insensitivity. As the likelihood insensitivity changes from the lack of likelihood insensitivity ($k = 1.0$) to 0.8, 0.7, and 0.6, the attacker's impression of a successful single person attack decreases from 95% to 86%, 81.5%, and 77%, respectively. Since the threat for a given attack is a measure of the proportional EU of that attack in respect to the summation of the EU of all attacks, the threat level of a single person attack is not greatly impacted across various likelihood insensitivity values for k. The utility of the single person attack in the absence of likelihood insensitivity is sufficiently low that the defender risk matrix prioritizes the attack with a risk rating of I across all $r_{a,} r_d$ combinations in Figure 4.4. As the attacker's perception of the single person attack's success decreases, the expected utility of that attack decreases, thus decreasing the threat of the attack in the presence of likelihood insensitivity. Since the single person attack is assigned a risk rating of I in the absence of likelihood for every combination of $r_{a,} r_d$ and likelihood insensitivity further reduces the threat level of the attack, the single person attack is assigned a risk rating of I across every combination of k, $r_{a,}$ and $r_{d.}$

While Figures 4.1, 4.2, 4.3, and 4.4 provide a visual depiction of the defender risk matrices for various likelihood insensitivity values, k =0.6, 0.7, 0.8 and 1.0, respectively, Figure 4.5 provides a condensed visualization of varying levels of risk aversion for the attacker and defender, likelihood insensitivity values and the respective risk ratings. Figure 4.5 is designed to be a prescriptive tool to help the defender evaluate attackers with varying likelihood insensitivity and risk aversion values while understanding the impact of a given risk aversion for the defender.

Research shows that decision makers tend to be overly risk averse and that being less risk neutral over time would be beneficial to the decision maker.

In the case of our defender, excess risk aversion can be detrimental since an unnecessarily high risk rating could mean a costly misappropriation of resources. In Figure 4.5, we observe that all attacks for varying risk aversion and likelihood insensitivity values for the attacker correspond to a risk rating equal to or higher than those assigned to a less risk averse defender as the defender becomes more risk averse. Like Chapter 3, for any given value of r_a and k, if $r_d > r_d'$ then the defender's $EU > EU'$, thus the risk rating is also higher. A costly misappropriation of resources can result when a defender does not exhibit risk neutrality, consequently assigning a risk rating for an attack higher than what the attack requires.

Given that we have justified the defender's risk neutrality over risk averseness, we now evaluate prioritization schemes available to the defender when $r_d = 1$ based on the defender's characterization of the attacker in regards to likelihood insensitivity and risk aversion. Since the single person attack's threat level results in a risk rating of I in the absence of likelihood insensitivity ($k = 1$) across all risk aversion values for the attacker, and the attack's threat level is further reduced in the presence of likelihood insensitivity ($k > 1$), the single person attack is assigned a risk rating of I across all risk aversion and likelihood insensitivity values for the attacker. Like our findings in Chapter 3, the defender is prescribed to appropriate resources aligned with a risk rating of I in respect to a single person attack, regardless of the defender's perception of the attacker.

Contrarily, the explosives on a plane attack receives a risk rating of I only if the attacker's likelihood insensitivity (k) is greater than or equal to 0.8. An attacker's perception of the explosives on a plane attack's success heavily depends on the attacker's likelihood insensitivity, as the risk aversion of the attacker does not play a role. A reduction in likelihood insensitivity causes the attacker to overweight the probability of a successful explosives on a plane attack, thus raising the threat level for the defender and subsequently raising the attack's risk rating. The defender's mischaracterization of the attacker's likelihood insensitivity in regards to the explosives on plane attack can result in excessive defense if the defender believes that the attacker is more likelihood insensitive than reality (a risk rating assignment of II when I is required), or inadequate defense if the defender believes that the attacker is less likelihood insensitive than reality (a risk rating assignment of I when II is required). Thus, the attacker's likelihood insensitivity plays a significant role in the appropriate resource allocation of the explosives on a plane attack's defense.

The multi-terror cell attack and the terror cell attack receive a risk rating of VI when the attacker exhibits risk neutrality ($r_a = 1$) while not displaying severe likelihood insensitivity ($k \neq 0.6$). When the attacker exhibits risk neutrality ($r_a = 1$) and severe likelihood insensitivity ($k = 6$), the multi-terror cell attack receives a risk rating of VII while the terror cell attack receives a risk rating of VI. As the attacker becomes more risk averse ($r_a > 1$), the multi terror cell attack receives the most severe risk rating of VIII for all r_a, k combinations. Contrarily, the terror cell attack receives a risk rating of IV for all r_a, k combinations as the attacker becomes more risk averse, except for an attacker with no/little likelihood insensitivity ($k = 0.8$ & $k = 1.0$) when $r_a = 2$. Determining the risk aversion of the attacker is critical in the case of the multi-terror cell attack and the terror cell attack, since the existence of the attacker's risk aversion would require the most severe risk rating for the multi-terror attack for all r_a, k combinations, while requiring a risk rating of either IV or V for the terror cell attack. The likelihood insensitivity impacts the risk rating assignment of the multi-terror cell attack if the attacker is risk neutral and severely likelihood insensitive. The terror cell attack's risk rating assignment; on the other hand, is only impacted by likelihood insensitivity if the attacker is moderately risk averse with no/little likelihood insensitivity.

As previously mentioned, likelihood insensitivity causes the attacker to overweight low probabilities and underweight high probabilities. Likelihood insensitivity also causes the attacker to be insensitive to probabilities near 50%. Since the anthrax attack has a probability of success, vulnerability, value of 53% in the absence of likelihood insensitivity, the anthrax attack is the least impacted by the existence of likelihood insensitivity. As the likelihood insensitivity changes from the lack of likelihood insensitivity ($k = 1.0$) to 0.8, 0.7, and 0.6, the attacker's impression of a successful anthrax attack increases from 53% to 53.5%, 54%, and 55%, respectively. However, since the anthrax attack is accompanied by a large consequence value of -3.95 and the attack is in a region on the defender risk matrix that is heavily impacted by changes in r_d, small changes in the threat level causes a change in the risk rating of the attack. Hence, a risk neutral attacker ($r_a = 1$) without likelihood insensitivity ($k = 1$) and a severely risk averse attacker ($r_a = 3$) without likelihood insensitivity ($k = 1$) requires a risk rating of III, while the same risk having attackers with likelihood insensitivity is assigned a risk rating of IV. Thus, we conclude that small changes on the threat level caused by likelihood insensitivity can impact risk rating assignments if the attack is of high consequence and moderate threat.

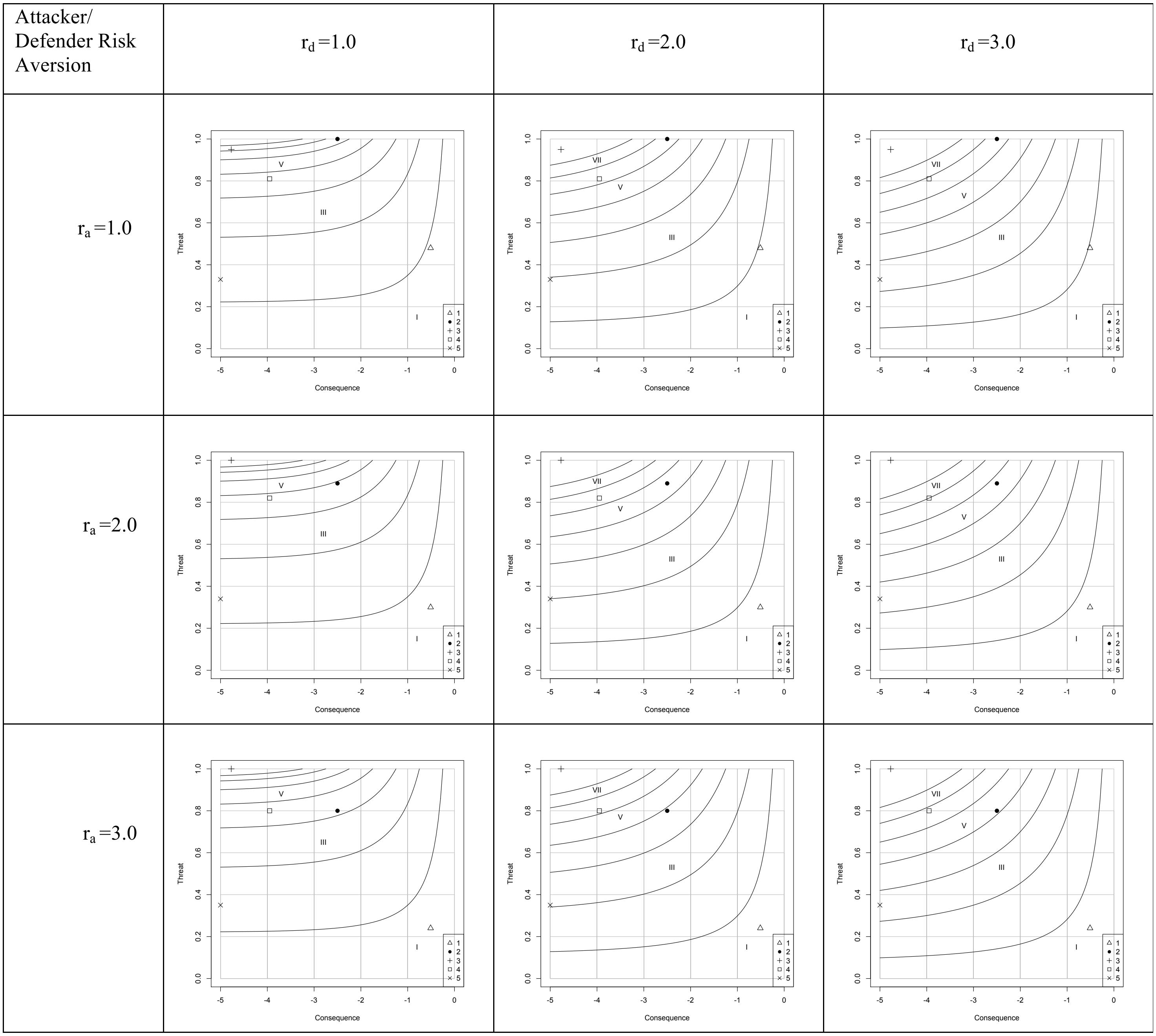


Figure 4.1: The Defender Risk Matrix for Varying Risk Aversion Values when k=0.6

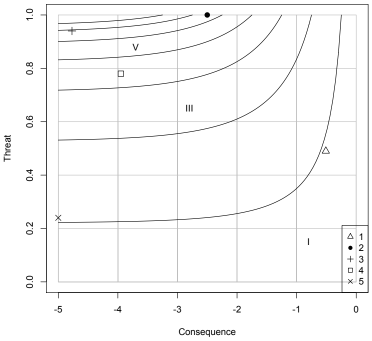
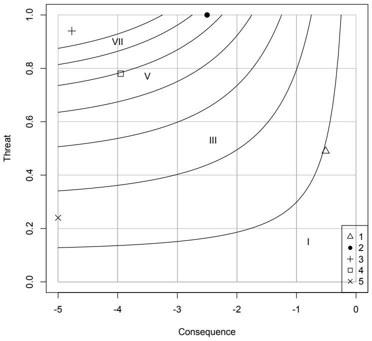
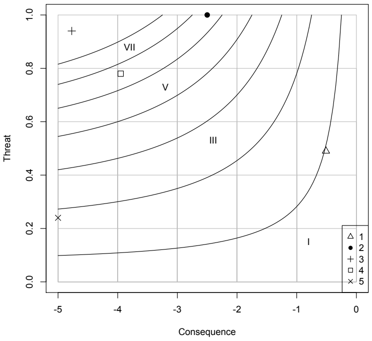
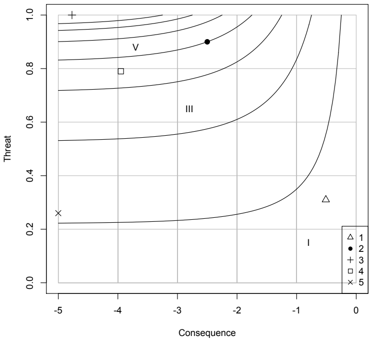
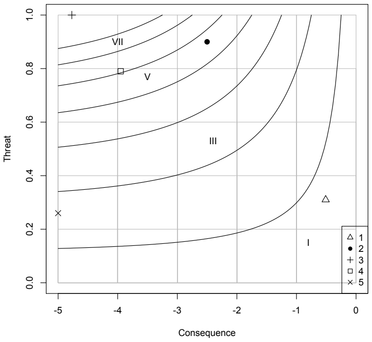
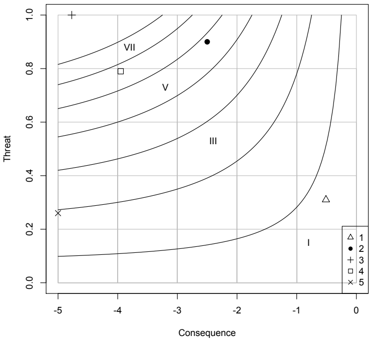
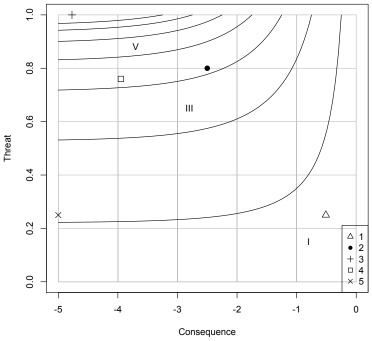
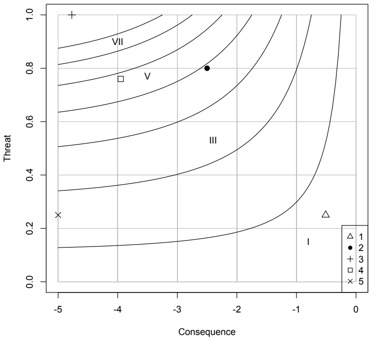
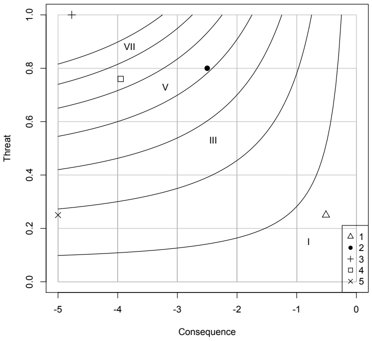
Attacker/ Defender Risk Aversion	r_d=1.0	r_d=2.0	r_d=3.0
r_a=1.0			
r_a=2.0			
r_a=3.0			

Figure 4.2: The Defender Risk Matrix for Varying Risk Aversion Values when k=0.7

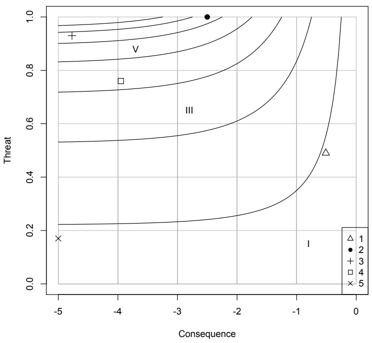
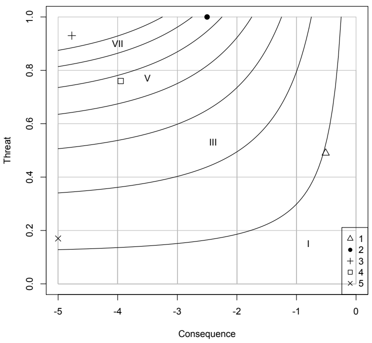
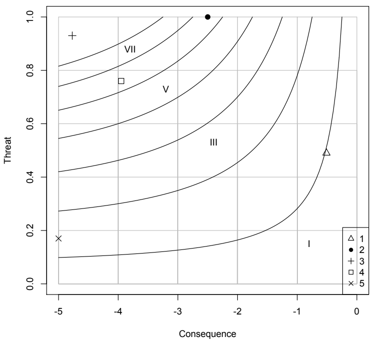
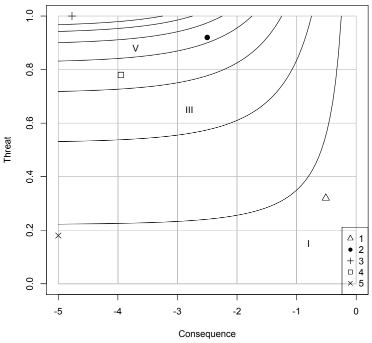
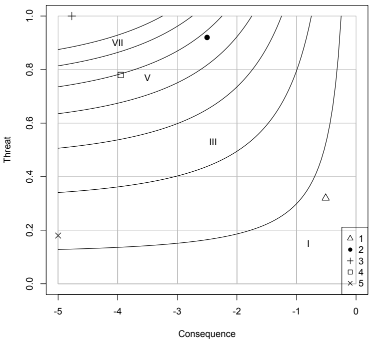
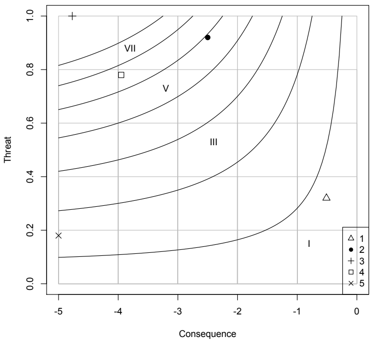
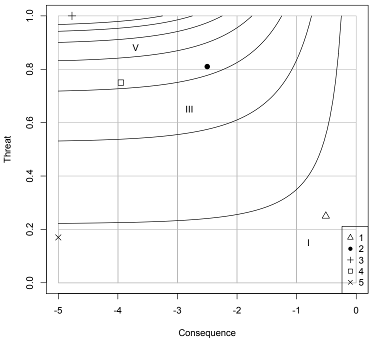
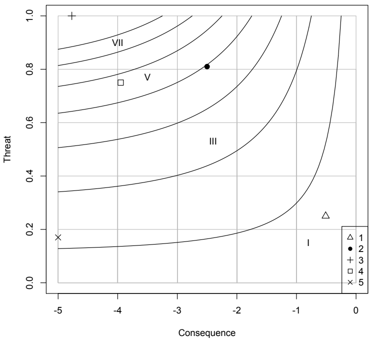
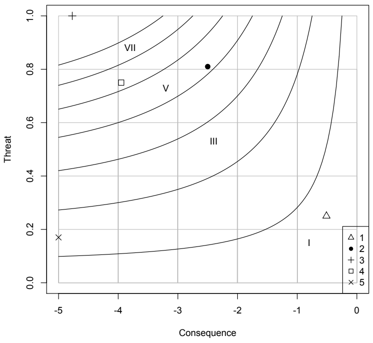
Attacker/ Defender Risk Aversion	r_d =1.0	r_d =2.0	r_d =3.0
r_a =1.0			
r_a =2.0			
r_a =3.0			

Figure 4.3: The Defender Risk Matrix for Varying Risk Aversion Values when k=0.8

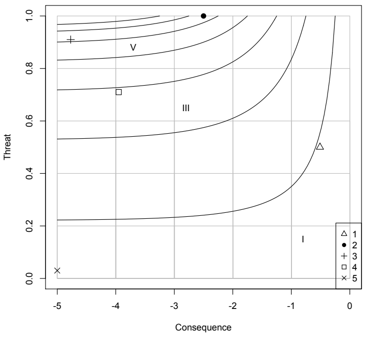
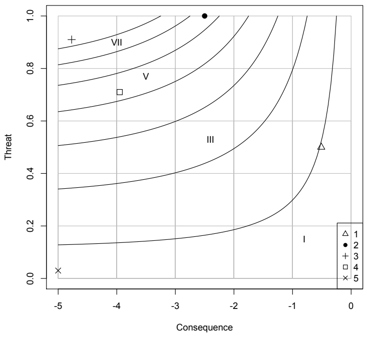
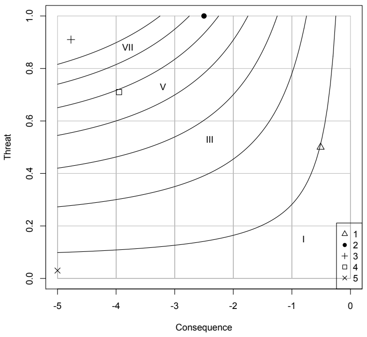
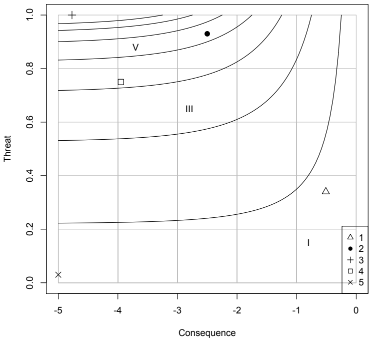
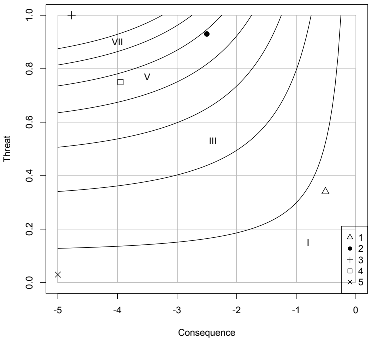
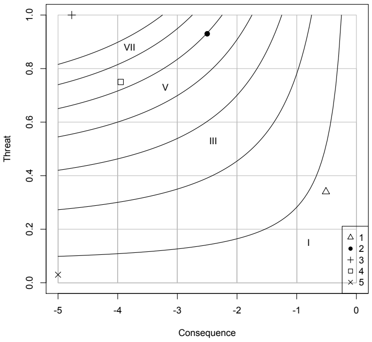
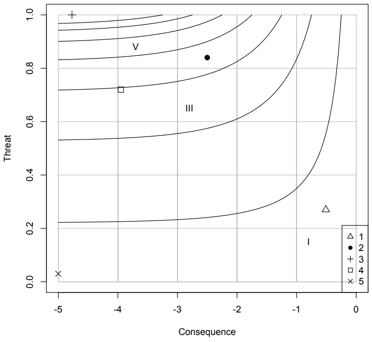
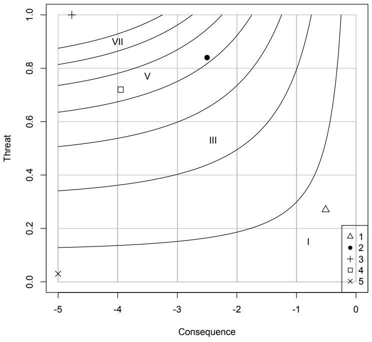
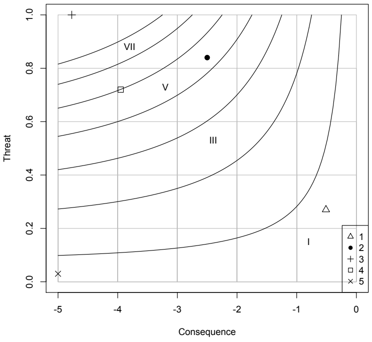
Attacker/ Defender Risk Aversion	r_d =1.0	r_d =2.0	r_d =3.0
r_a =1.0			
r_a =2.0			
r_a =3.0			

Figure 4.4: The Defender Risk Matrix for Varying Risk Aversion Values when k=1.0

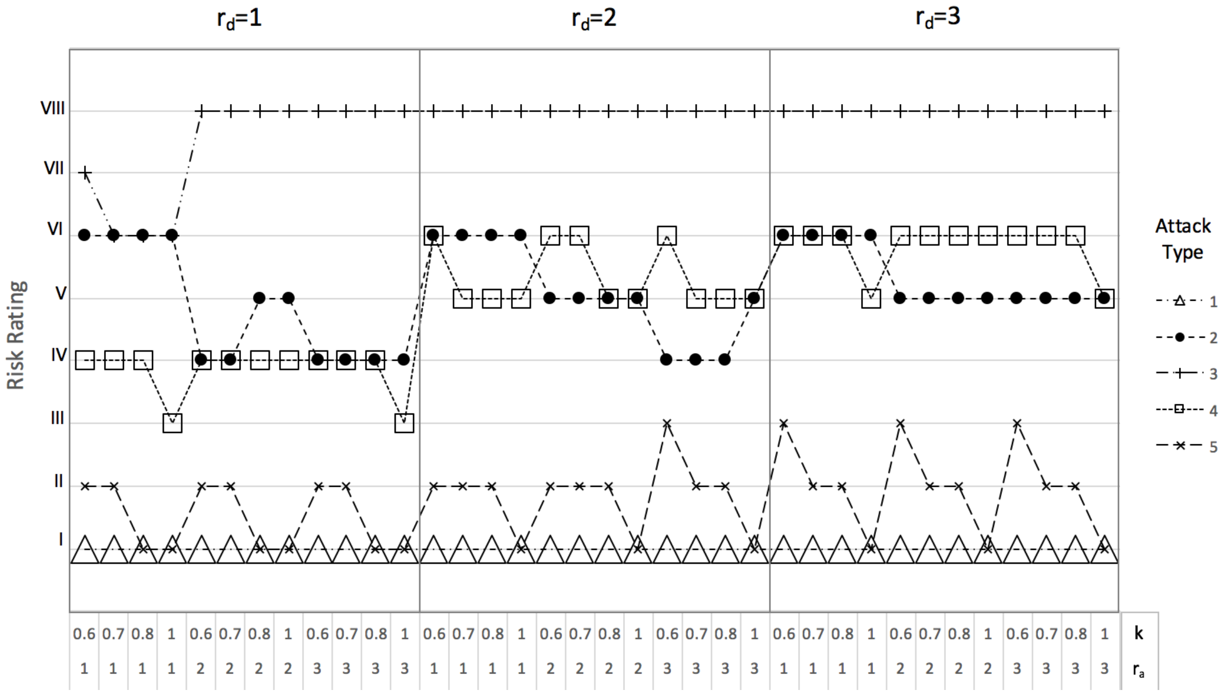


Figure 4.5: Risk Ratings for Varying Risk Aversion and Likelihood Insensitivity Values

4.5 Conclusions

In Chapter 3, we introduced adversarial risk matrices that utilize an expected utility framework for the attacker and the defender, and concluded that the defender's description of the attacker is essential in proper resource allocation. In Chapter 4, we improved the defender's description of the attacker by utilizing a prospect-theoretic framework. Adversarial risk matrices that describe the attacker in the prospect-theoretic sense and prescribe actions to the defender using Expected Utility Theory enhance the defender's risk categorization capabilities by providing the defender a more realistic description of the attacker, one that is more aligned with empirical research.

Chapter 5: Conclusions

Decision making in counterterrorism applications have traditionally considered a rational attacker with a single objective of maximizing financial damage and a defender with the objective of minimizing the damage inflicted by the attacker. Empirical research suggests that decision makers deviate from rationality, thus it is more realistic to provide a descriptive model of the attacker using prospect theory, while prescribing actions to the defender using expected utility theory. Using this as a foundation of our research, in Chapter 2, we modeled the attacker's behavior using multi-attribute prospect theory to account for an attacker with multiple objectives and deviations from rationality, while using expected utility theory to prescribe the appropriate actions to the defender. We modeled our approach by considering an attacker who wishes to smuggle radioactive material into the United States and a defender who has the option of implementing a screening process to hinder the attacker. In Chapter 3, we incorporated an Expected Utility framework to model adversarial risk matrices where risk matrices have not been historically considered as a two level decision making process. In Chapter 4, we enhanced our work in Chapter 3 by considering a prospect theoretic approach to describe the attacker, while using expected utility to prescribe actions to the defender. Our findings show that the descriptive validity of the defender's characterization of the attacker is critical in appropriate resource allocation in defense of the attacks, thus the defender should carefully consider the impacts of the attacker's deviations from rationality when considering defense options.

Bibliography

Bibliography

Abbas, A. E., & Bell, D. E. (2012). METHODS—One-Switch Conditions for Multiattribute Utility Functions. *Operations Research*, *60*(5), 1199-1212.

Abbas, A. E., & Bell, D. E. (2011). One-switch independence for multiattribute utility functions. *Operations Research*, *59*(3), 764-771.

Abbas, A. E., & Matheson, J. E. (2005). Normative target-based decision making. *Managerial and Decision Economics*, *26*(6), 373-385.

Abbas, A. E. (2011). The multiattribute utility tree. *Decision Analysis*, *8*(3), 180-205.

Abbas, A. E. (2009). Multiattribute utility copulas. *Operations Research*, *57*(6), 1367-1383.

Anthony Tony Cox, L. (2008a). What's wrong with risk matrices?. *Risk analysis*, *28*(2), 497-512.

Baillon A, Bleichrodt H, Keskina U, L'Haridonb O, Lia C. Learning under ambiguity: An experiment using initial public offerings on a stock market. Economics Working Paper Archive (University of Rennes 1 & University of Caen), Center for Research in Economics and Management (CREM), University of Rennes 1, University of Caen and CNRS. Available at: http://EconPapers.repec.org/RePEc:tut:cremwp:201331, Accessed July 1, 2014.

Bakır, N. O. (2008). A decision tree model for evaluating countermeasures to secure cargo at United States southwestern ports of entry. *Decision Analysis*, *5*(4), 230-248.

Ball, D. J., & Watt, J. (2013). Further thoughts on the utility of risk matrices. *Risk analysis*, *33*(11), 2068-2078.

Banks, D. L., & Anderson, S. (2006). Combining game theory and risk analysis in counterterrorism: A smallpox example. *Statistical methods in counterterrorism*, 9-22.

Berejikian, J. D. (2002a). A cognitive theory of deterrence. *journal of peace research*, *39*(2), 165-183.

Berejikian, J. D. (2002b). Model building with prospect theory: A cognitive approach to international relations. *Political Psychology*, *23*(4), 759-786.

Bier V. M. Game-theoretic and reliability methods in counter-terrorism and security. In *Mathematical and Statistical Methods in Reliability, Series on Quality, Reliability and Engineering Statistics*, A. Wilson, N. Limnios, S. Keller-McNulty, Y. Armijo, eds., pp. 17–28, World Scientific, Singapore, 2005.

Bier, V. M., Gratz, E. R., Haphuriwat, N. J., Magua, W., & Wierzbicki, K. R. (2007a). Methodology for identifying near-optimal interdiction strategies for a power transmission system. *Reliability Engineering & System Safety*, *92*(9), 1155-1161.

Bier, V., Oliveros, S., & Samuelson, L. (2007b). Choosing what to protect: Strategic defensive allocation against an unknown attacker. *Journal of Public Economic Theory*, *9*(4), 563-587.

Bleichrodt, H., & Miyamoto, J. (2003). A characterization of quality-adjusted life-years under cumulative prospect theory. *Mathematics of Operations Research*, *28*(1), 181-193.

Bleichrodt, H., Schmidt, U., & Zank, H. (2009). Additive utility in prospect theory. *Management Science*, *55*(5), 863-873.

Bond, S. D., Carlson, K. A., & Keeney, R. L. (2008). Generating objectives: Can decision makers articulate what they want?. *Management Science*, *54*(1), 56-70.

Bond, S. D., Carlson, K. A., & Keeney, R. L. (2010). Improving the generation of decision objectives. *Decision Analysis*, *7*(3), 238-255.

Brown, G. G., & Cox Jr, L. A. T. (2011). How probabilistic risk assessment can mislead terrorism risk analysts. *Risk Analysis*, *31*(2), 196-204.

Butler, C. K. (2007). Prospect theory and coercive bargaining. *Journal of Conflict Resolution*, *51*(2), 227-250.

Camerer C. *Behavioral Game Theory*. Princeton University Press, Princeton PA, 2003.

Chateauneuf, A., Eichberger, J., & Grant, S. (2007). Choice under uncertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory*, *137*(1), 538-567.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and brain sciences*, *26*(2), 139-198.

.

Cox Jr, L. A. T. (2008b). Some limitations of "Risk= Threat× Vulnerability× Consequence" for risk analysis of terrorist attacks. *Risk Analysis*, *28*(6), 1749-1761.

Cox Jr, L. A. T. (2009). What's Wrong with Hazard-Ranking Systems? An Expository Note. *Risk Analysis*, *29*(7), 940-948.

Cox Jr, L. A. (2009). Improving Risk-Based Decision Making for Terrorism Applications. *Risk Analysis*, *29*(3), 336-341.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological bulletin*, *73*(3), 221.

Einhorn, H. J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance*, *6*(1), 1-27.

Feng, T., & Keller, L. R. (2006). A multiple-objective decision analysis for terrorism protection: Potassium iodide distribution in nuclear incidents. *Decision Analysis*, *3*(2), 76-93.

Fishburn, P. C. (1984). Multiattribute nonlinear utility theory. *Management Science*, *30*(11), 1301-1310.

Gigerenzer, G., Todd, P. M., & ABC Research Group, T. (1999). *Simple heuristics that make us smart*. Oxford University Press.

Heal G, Kunreuther H. IDS models of airline security. *Journal of Conflict Resolution,* 2005; 49(2) 201–217.

Heal, G., & Kunreuther, H. (2005). IDS models of airline security. *Journal of conflict resolution*, *49*(2), 201-217.

John R S, Rosoff H. Modeling effects of counterterrorism initiatives for reducing adversary threats to transportation systems. In *Proceedings of the 2011 DHS Science Conference – 5th Annual University Network Summit, focused on Catastrophes and Complex Systems: Transportation*, 2011.

John R S, Rosoff H. Validation of proxy random utility models for adaptive adversaries. In *Proceedings of the Probabilistic Safety Assessment and Management (PSAM) Bi-Annual Conference*, 2014.

Kadane, J. B., & Larkey, P. D. (1982a). Subjective probability and the theory of games. *Management Science*, *28*(2), 365-1379.

Kadane, J. B., & Larkey, P. D. (1982b). The Confusion of Is and Ought in Game Theoretic Contexts. *Management Science*, *29*(12), 113-120.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263-291.

Kar, D., Fang, F., Delle Fave, F., Sintov, N., & Tambe, M. (2015, May). A game of thrones: when human behavior models compete in repeated Stackelberg security games. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1381-1390). International Foundation for Autonomous Agents and Multiagent Systems.

Keeney, G. L., & Von Winterfeldt, D. (2010). Identifying and structuring the objectives of terrorists. *Risk Analysis*, *30*(12), 1803-1816.

Keeney R, Raiffa H. *Decisions with Multiple Objectives, Preferences, and Value Tradeoffs*. John Wiley and Sons: New York, 1976.

Keeney, R. L. (2007). Modeling Values for Anti-Terrorism Analysis. *Risk Analysis*, *27*(3), 585-596.

Keeney R. *Value-focused thinking: A path to create decision-making*. Princeton University Press: Princeton, N.J., 1992.

Kirkwood C W. *S*trategic Decision Making, Multiobjective Decision Analysis with Spreadsheets. Duxbury Press: Belmont, CA, 1997.

Kunreuther H, Heal G. Interdependent security. *Journal of Risk & Uncertainty,* 2003; 26(2–3) 231–249.

Kunreuther, H., & Heal, G. (2003). Interdependent security. *Journal of risk and uncertainty*, *26*(2-3), 231-249.

Leung, M., Lambert, J. H., & Mosenthal, A. (2004). A Risk-Based Approach to Setting Priorities in Protecting Bridges Against Terrorist Attacks. *Risk Analysis*, *24*(4), 963-984.

Lewis, B., Erera, A., & White III, C. (2003). Optimization approaches for efficient container security operations at transshipment seaports. *Transportation Research Record: Journal of the Transportation Research Board*, (1822), 1-8.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of mathematical psychology*, *15*(3), 215-233.

Markowski, A. S., & Mannan, M. S. (2008). Fuzzy risk matrix. *Journal of hazardous materials*, *159*(1), 152-157.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, *10*(1), 6-38.

Merrick, J. R., & Leclerc, P. (2016). Modeling Adversaries in Counterterrorism Decisions Using Prospect Theory. *Risk Analysis*, *36*(4), 681-693.

Merrick, J. R., & McLay, L. A. (2010). Is screening cargo containers for smuggled nuclear threats worthwhile?. *Decision Analysis*, *7*(2), 155-171.

Merrick, J., & Parnell, G. S. (2011). A comparative analysis of PRA and intelligent adversary methods for counterterrorism risk management. *Risk Analysis*, *31*(9), 1488-1510.

Metzger L P, Rieger M O. Equilibria in Games with Prospect Theory Preferences. Finrisk Working Paper Series, Nos. 598, 2009.

Miyamoto, J. M., & Wakker, P. (1996). Multiattribute utility theory without expected utility foundations. *Operations Research*, *44*(2), 313-326.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, *36*(1), 48-49.

Nguyen, T. H., Yang, R., Azaria, A., Kraus, S., & Tambe, M. (2013, July). Analyzing the Effectiveness of Adversary Modeling in Security Games. In *AAAI*.

Novemsky, N., & Kahneman, D. (2005). The boundaries of loss aversion. *Journal of Marketing research*, *42*(2), 119-128.

Parnell, G. S., Smith, C. M., & Moxley, F. I. (2010). Intelligent adversary risk analysis: A bioterrorism risk management model. *Risk Analysis*, *30*(1), 32-48.

Parnell, G. S., Smith, C. M., & Moxley, F. I. (2010). Intelligent adversary risk analysis: A bioterrorism risk management model. *Risk Analysis*, *30*(1), 32-48.

Paté-Cornell, E., & Guikema, S. (2002). Probabilistic modeling of terrorist threats: A systems analysis approach to setting priorities among countermeasures. *Military Operations Research*, *7*(4), 5-23.

Raiffa, H. (1982). *The art and science of negotiation*. Harvard University Press.

Ramsey F. *Foundations Essays in Philosophy, Logic, Mathematics and Economics*. Humanities Press, 1931.

Rios Insua, D., Rios, J., & Banks, D. (2009). Adversarial risk analysis. *Journal of the American Statistical Association*, *104*(486), 841-854.

Ritzberger, K. (1996). On games under expected utility with rank dependent probabilities. *Theory and Decision*, *40*(1), 1-27.

Rosoff, H., & John, R. (2009, June). Decision analysis by proxy for the rational terrorist. In *Quantitative risk analysis for security applications workshop (QRASA) held in conjunction with the International Joint Conference on AI* (pp. 25-32).

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological science*, *12*(3), 185-190.

Ruan, X., Yin, Z., & Frangopol, D. M. (2015). Risk matrix integrating risk attitudes based on utility theory. *Risk Analysis*, *35*(8), 1437-1447.

Savage L J. *Foundations of Statistics*. John Wiley: New York, N.Y, 1954.

Shan, X., & Zhuang, J. (2013). Hybrid defensive resource allocations in the face of partially strategic attackers in a sequential defender–attacker game. *European Journal of Operational Research*, *228*(1), 262-272.

Siebert, J., von Winterfeldt, D., & John, R. S. (2015). Identifying and structuring the objectives of the Islamic State of Iraq and the Levant (ISIL) and its followers. *Decision Analysis*, *13*(1), 26-50.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, *69*(1), 99-118.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, *5*(4), 297-323.

Tversky A, Sattath S, Slovic P. Contingent weighting in judgment and choice. *Psychological Review,* 1988; 95(37): 1-84

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, *79*(4), 281.

U.S. Department of Transportation (2005). Treatment of the Economic Value of a Statistical Life in Departmental Analyses. http://ostpxweb.dot.gov/policy/reports/080205.htm.

von Neumann J, Morgenstern O. *The Theory of Games and Economic Behaviour*. Princeton University Press: Princeton, NJ, 1944.

Von Winterfeldt, D., & O'Sullivan, T. M. (2006). Should we protect commercial airplanes against surface-to-air missile attacks by terrorists?. *Decision Analysis*, *3*(2), 63-75.

Yang, R., Kiekintveld, C., OrdóñEz, F., Tambe, M., & John, R. (2013). Improving resource allocation strategies against human adversaries in security games: An extended study. *Artificial Intelligence*, *195*, 440-469.

Zank, H. (2001). Cumulative prospect theory for parametric and multiattribute utilities. *Mathematics of Operations Research*, *26*(1), 67-81.

Zhuang, J., Bier, V. M., & Gupta, A. (2007). Subsidies in interdependent security with heterogeneous discount rates. *The Engineering Economist*, *52*(1), 1-19.

Zhuang, J., & Bier, V. M. (2007). Balancing terrorism and natural disasters—Defensive strategy with endogenous attacker effort. *Operations Research*, *55*(5), 976-991.