

More Options for Quantifying Deterrence and Reducing Critical Infrastructure Risk: Cognitive Biases

by Eric F. Taquechel, Ted G. Lewis

Abstract

We expand on the application of quantifiable deterrence to critical infrastructure/key resource protection by considering cognitive biases. These biases include what we call “information obfuscation bias” and “prospect bias”, the latter inspired by Kahneman and Tversky’s Prospect Theory. We show how quantifiable deterrence effectiveness and resulting critical infrastructure risk change when we obfuscate notional Port Security Grant investment information from a prospective attacker, and we also explore whether these metrics change if we assume Prospect Theory is a more accurate explanation of decision making than classical Subjective Expected Utility Theory. Importantly, we do not advocate for policy changes but rather expand on a previously published methodology that might support such decisions in the future.

Suggested Citation

Taquechel, Eric F. & Lewis, Ted G. “More Options for Quantifying Deterrence and Reducing Critical Infrastructure Risk: Cognitive Biases.” *Homeland Security Affairs* 12, Article 3 (September 2016). <https://www.hsaj.org/articles/12007>

Executive Summary

The goal of this article is to illustrate a process to support decisions on whether to publicize information about CIKR security investments intended to deter attacks, or whether to obfuscate those investments, by considering cognitive biases. Importantly, we are not advocating for publicizing or obfuscating details of federal grant investments in general. We simply offer a methodology to support such decisions.

To set the context for this proposed process, we claim that the notion that people make completely rational, fully informed decisions is debatable. Expected utility theory (EUT) assumes that people act according to their preferences, and preferences are consistent regardless of how options are presented. However, Kahneman and Tversky developed prospect theory (PT) showing experimentally that decisions may be inconsistent with EUT, depending on how options are presented or framed.

The concept of cognitive bias follows from this alternative to EUT. Other biases may be due to limited information availability such as imperfect or incomplete information. These biases become relevant to critical infrastructure/key resource (CIKR) risk reduction and attack deterrence when we consider adversarial decision making, as attacker intent is one component of critical infrastructure risk.

Deterrence is the process of influencing decision making; we want to manipulate our adversary’s assessment of their interests. Expanding on previous published work, we focus on quantifiable deterrence as we evaluate the effect of cognitive biases. More specifically, we want to evaluate the effects of cognitive biases upon CIKR attack desirability and deterrence, hypothesize attacker preferences, estimate the resulting risk, and suggest advantages of publicizing or obfuscating CIKR security investments.

Relevance and Background

We know that CIKR security practitioners already publicize some information and obfuscate other information in defense of their facilities. But, can we predict how additional security investments might measurably deter and buy down additional risk if an adversary has biased perceptions of those investments? Furthermore, we claim it is possible that the effectiveness of obscuring or publicizing certain information may not be robust across different utility theory assumptions. Thus, the intended audience for this research is broad: CIKR owners/operators, policymakers, academics, and risk analysts.

Before we illustrate our process, we provide an extensive literature review in the article, covering work on game theory, utility theories including EUT and PT, information availability, optimization, and deterrence quantification and portfolio development.

Our Process

Previous research has explained a generic process to quantify deterrence and show how that influences CIKR risk, yielding a “deterrence portfolio” of metrics to support decision making. We here extend this process and show how notional deterrence portfolios might change if we apply various cognitive biases. We also examine the robustness of this approach across different ways of approximating attacker intent. We leverage basic principles of game theory in “deterrence games.”

We explore this methodology in a case study of notional CIKR, and notional defender budgets such as those that might be available from FEMA’s Port Security Grant Program (See Appendix I).

Importantly, all data and maritime facilities are notional in this case study, but real threat, facility vulnerability, facility consequence, and budget data would be used for a real analysis.

Results of Case Study

We learned that assumptions about how our opponents perceive and rank their options dictate the relative advantages of obfuscating or publicizing our deterrence actions as CIKR defenders. For example, if an attacker formulates intent to attack only one target, rather than “ranking” multiple targets in order of desirability, that may mean our decision to obfuscate or publicize information on CIKR defensive investments does not change expected risk. However, more research is needed to generalize such findings.

Options for Future Research

We encourage future research along the lines of modifying assumptions about information availability, modifying the expected utility functions used in deterrence games, and applying principles of PT in different ways. We also advocate consideration of how investments to improve CIKR resilience and/or mitigate attacker capabilities to launch attacks might influence the outcome of deterrence games and inform decisions on whether to obfuscate or publicize deterrence investment information. In our approach here, we have only focused on notional investments to protect CIKR against attacks. Perhaps most importantly to practitioners, we suggest efforts to incorporate theoretical findings into real-world risk models.

The “Big Picture”

Research has suggested that deterrence theory is applicable to many of the 21st century threats the U.S. will face, but more work needs to be done to determine how that theory is put into practice, or “operationalized.” We believe this extension of an earlier published deterrence quantification approach helps advance “operationalization” of deterrence theory to a very relevant 21st century threat: terrorist attacks against CIKR.

Introduction

Our objective in this article is to propose enhancements to an existing process to quantify the deterrent effects of investments to secure critical infrastructure. To set the context for these enhancements, we claim that the notion that people make completely rational, fully informed decisions is debatable. Expected utility theory (EUT) assumes that people find the best possible solution from among all known options, choosing a solution that will maximize their expected utility. They act according to their preferences, and preferences are consistent regardless of how options are presented. As one alternative to EUT, Simon proposed satisficing theory, which predicts a decision maker will find a minimum acceptability threshold, instead of necessarily maximizing their utility to attain the optimal solution.¹ As a second alternative, Kahneman and Tversky developed prospect theory (PT) showing experimentally that decisions may be inconsistent with EUT, depending on how options are presented or framed.

The concept of *cognitive bias* follows from these alternatives to rational, fully informed decision making. Some biases, such as those associated with PT, are influenced by how possible outcomes or “prospects” are presented relative to a “reference point” of desired utility. This introduces inconsistency into decision making. Other biases may be due to limited information availability² such as imperfect or incomplete information, which may lead to satisficing.

Does this matter to critical infrastructure/key resource (CIKR) risk reduction and attack deterrence? It matters if we consider adversarial decision making. CIKR risk is the expected loss resulting from an attack. Expected can mean probabilistic, or the likelihood that an attack will be successful. In DHS terms, likelihood can be a combination of threat * vulnerability.³ Threat is intent * capability, capability is the probability an adversary *can* attack a CIKR, and intent is the probability that an adversary *wants* to attack.⁴ We consider adversarial decision making when we focus on intent.

Adversarial decision making entails how a would-be CIKR attacker considers information and decides whether to attack, or to refrain from attack altogether. Their anticipated decision reflects their intent. We may not know their intent in advance. However, we might instead speculate about:

1. what they do know about our CIKR,
2. how they might evaluate options, and
3. what decisions they might arrive at, subject to cognitive biases and different ways to evaluate prospects.

More specifically, we want to evaluate the effects of cognitive biases upon CIKR attack desirability and deterrence, hypothesize attacker preferences, estimate the resulting risk, and suggest advantages of publicizing or obfuscating security investments. Deterrence is the process of influencing decision making; we want to manipulate our adversary's assessment of their interests.⁵ We focus on *measurable* or *quantifiable* deterrence as we evaluate the effect of cognitive biases. Recent work proposed a way to quantify the effects of deterrence upon CIKR risk⁶, but did not explore cognitive biases. Nikhil Dighe et al. urge that future work on analyzing deterrence investment should consider alternatives to EUT.⁷ With respect to imperfect/incomplete information, attacker uncertainties are critical to understanding deterrence, but are allegedly rarely leveraged in game-theoretic analysis of counterterrorism.⁸ In this work we will call biases resulting from information imperfection and information incompleteness "information obfuscation biases (IOB)." We will call biases proposed by PT "prospect biases."

Intended Audiences

Some CIKR security information is obvious to the public, such as the presence of armed guards at facility entrances. Some information is obfuscated: for example, what does the facility's proprietary security plan say about law enforcement response to a security threat? Thus, we know that practitioners already publicize some information and obfuscate other information. But, can we predict how additional security investments might measurably deter and buy down additional risk if an adversary has biased perceptions of those investments?

IOB might seem more relevant to a CIKR owner than prospect bias: owners and regulators may decide whether to make their security measures overt or covert.⁹ But, it makes sense to also consider biases predicted by different utility theories. What if our adversaries are motivated in ways that EUT does not account for? Do they evaluate the possible gains from successful CIKR attacks on their face value? Or, do they evaluate them relative to some desired reference point?

A CIKR owner may not consider these factors on a daily basis, but risk analysts and policymakers might consider them, given the evidence of inconsistent decision making. Sensitivities of deterrence effectiveness and risk reduction to different utility theories may have real implications for CIKR owners and practitioners. It is possible that the effectiveness of obscuring or publicizing certain information may not be robust across different utility theory assumptions. Thus, the intended audience for this research is broad: CIKR owners/operators, policymakers, academics, and risk analysts.

Relevant Work

Taquechel and Lewis give a brief overview of literature on risk analysis and deterrence that is relevant to the present work.¹⁰ They then explain basic principles of game theory and offer a simple approach to information availability. They also discuss the concepts of EUT and PT. But, the emphasis in that work is on the basic process to quantify deterrence and create "deterrence portfolios." Thus, the present work will elaborate on the game theoretical approach and cognitive biases introduced in Taquechel and Lewis (2012). It will also show additional findings to support decisions on whether to publicize or obfuscate deterrence investment information.

Game Theory

Game theory helps us model the possible effects of our adversary's perceptions. Those perceptions may be influenced by what information we publicize or obfuscate in a deterrence game. For example, if we communicate details of our ability to defeat an attack, we may deter terrorism, regardless of whether we actually can defeat an attack. Moran claims that deterrence works most convincingly between known adversaries who share a common estimate of each other's hostile intentions.¹¹ In contrast, Chilton and Weaver claim that under some circumstances, ambiguity will enhance deterrence.¹² Which is true?

With regard to a "common estimate" vs. ambiguity, we now discuss two dyads of information availability commonly applied in game theoretic analyses: perfect vs. imperfect information, and complete vs. incomplete information.

Perfect or Imperfect Information?

Games can assume either perfect or imperfect information. Perfect information means a sequential game, wherein the attacker could observe the defender's investment courses of action (COA), and could analyze the implications of different CIKR attacks. A sequential game is "one in which players make decisions following a certain predefined order, and in which at least some players can observe the moves of players who preceded them."¹³ Often sequential games are known as Stackelberg games, and in the context of CIKR protection, they are called attacker-defender games. For example, see Brown et al.¹⁴ Some claim that Stackelberg games are appropriate for real word scenarios because attackers can observe CIKR defenses before making decisions.¹⁵ However, a game to determine where to invest *additional* resources and deter potential attackers could simulate a defender obfuscating their actual investment COA. This would result in a game of imperfect information.

Imperfect information means a simultaneous game, wherein all players select a COA without knowledge of the COAs that other players select. This is even if the decisions are made at different points in time.¹⁶ For CIKR attack deterrence, a simultaneous game would mean:

1. that the defender would not know in advance which CIKR an attacker had decided to attack, and
2. the attacker would not know in advance where the defender had invested to deter, although the attacker may know the investment amount if the defender protected a specific CIKR.

This may mean that an attacker cannot observe the effects of deterrence investment at a CIKR.

The effects of imperfect and perfect information on CIKR protection game results have been studied. For example, Hausken et al. compare the outcomes of both simultaneous and sequential games. In these games, a defender is considering investment to protect CIKR from both terrorism and natural hazards, and an attacker is considering what CIKR to attack.¹⁷ For sequential games, Yin et.al. propose the idea of a Strong Stackelberg Equilibrium (SSE) in a game between an attacker and defender in the CREATE PROTECT model.¹⁸ The PROTECT approach, created for the U.S. Coast Guard, leverages a Stackelberg algorithm to produce a

randomized Coast Guard boat patrol schedule. When executed over a long period of time, this schedule theoretically minimizes an attacker's ability to plan an attack on a maritime CIKR.

Complete or Incomplete Information?

Incomplete information means that players do not know some of the elements which define the rules of the game.¹⁹ For CIKR deterrence games, this may mean the attacker does not know how much a CIKR defender would invest at a specific CIKR target. If an attacker does not know dollar amounts of deterrence investments, and dollar amounts are used to create the expected utility functions that determine payoffs, then the attacker will have incomplete information.

In contrast, a game of complete information means players know all elements of the game. If the attacker knows the dollar amounts invested to deter, they would have complete information, assuming they know all of the other elements. This information would be publicized by the defender, or easily attainable by the attacker.

The effects of complete and incomplete information on CIKR protection game results have been studied. For example, Jenelius et al.²⁰ examine how adversarial "observation error" influences deterrence, risk, and optimal resource allocation to defend CIKR.²¹ Azaiez proposes three characterizations of attacker confidence when CIKR vulnerabilities are uncertain: optimistic, neutral, and pessimistic.²²

Utility Functions and COAs For Deterrence Games

Utility is the value or payoff of the outcome (prospect) of a COA.²³ For this research, attacker utility will just be the payoff of defender deaths and immediate economic consequences from a successful attack. From the defender's perspective, this loss of life and economic damage would be risk, but any retained life and economic productivity is defender utility.

Morrall and Jackson advocate reducing probability of achieving a payoff, as well as the value of the payoff itself, in order to deter attacks.²⁴ Subjective expected utility (SEU) is thus utility modified by a subjective probability of attaining that utility.²⁵ This is similar to EUT except that now a subjective probability governs the expected outcome. These concepts are interchangeable and we will use SEU from here on out.

The present work uses a game theoretical approach to measure deterrence, and claims that the effect of deterrence is a component of CIKR risk. Therefore, we translate the language of game theory into the language of risk analysis. We have determined how expected utility functions, used in game theory, can be converted into probabilistic risk equations. This paper will leverage utility functions and risk equations that include probabilities of attack success. Also, we will model probabilities as functions of investment to reduce CIKR vulnerability.

What is an appropriate way to model this function? Lewis argues that a linear cost model is unrealistic. In practice, a target's security may be increased by 50% for 10% of the budget, but only by another 20% by investing twice as much.²⁶ Al-Mannai and Lewis propose the following²⁷:

$$v_i(C_i) = e^{-\alpha_i C_i}, \quad \alpha_i = \frac{-\ln(EF_i)}{EC_i}$$

Equation 1. General exponential vulnerability-investment relationship

where:

1. $v_i(C_i)$ is the vulnerability of the i -th CIKR target, and is a function of defender investment C_i , and
2. α_i is the slope of the exponential curve, a function of the elimination cost EC_i to reduce vulnerability to some elimination fraction EF_i .

For exponential relationships, vulnerability cannot be completely eliminated, so an arbitrary elimination fraction such as 5% can be used. Bier et al. also assess the probability of an attack as an exponential function of budget invested to defend that target.²⁸ It may make sense to use a nonlinear vulnerability-investment term if we believe our adversaries are adaptive.

Optimization COA

Since we model expected utility functions as functions of deterrence investments, we can calculate the optimal investment to maximize expected utility, either formally or via simulation. Lewis calculates optimal investments for CIKR protection using both methods.²⁹ The formal techniques include Lagrange multipliers.³⁰ We use the optimal CIKR deterrence investment as input for one of the defender's COAs in our deterrence games.

Other studies propose optimization methods for critical infrastructure protection, using game theoretic context. For example, Levitin offers techniques for calculating optimal strategies for both attacker and defender with respect to complex infrastructure systems, considering possibilities of a single or multiple simultaneous or sequential attacks.³¹

Utility Theories

Lebow and Stein³² claim that utility can be calculated differently by different rational actors. Extending that claim, the present work claims that the application of measurable deterrence to CIKR risk analysis and protection must also account for different utility calculations.

Subjective Expected Utility Theory

With SEU, the expected utility from a COA is relative to a net asset position.³³ SEU is traditionally thought to govern attitudes toward decision-making and so is not a “cognitive bias.” Rather, we treat it here as a baseline against which the influence of PT biases can be evaluated.

With respect to SEU, Schoemaker explained that risk aversion meant a gamble would be less preferred than its expected value for certain, and that risk seeking meant a gamble would be more preferred than its expected value for certain.³⁴ This reflects the concept of certainty equivalent, the maximum amount someone would pay for some expected utility.³⁵ Importantly, Schoemaker argued that the certainty equivalent is invariant under different conditions of wealth, when assuming SEU.³⁶ If we assume one’s current wealth is one’s reference point, or desired utility, then under SEU assumptions the framing of options relative to a reference point would not influence one’s risk aversion or risk-seeking preferences.

To illustrate, suppose one has \$10 and faces the prospect to gain \$10 with 100% certainty, or gain \$30 with 33% chance. The expected utility of this “gains-framed” prospect is \$20 either way. A risk averse actor would prefer the certain bet, whereas a risk seeking actor would prefer the gamble. Alternatively, suppose one has \$10 but stands to lose \$10 with 100% certainty, or lose \$30 with 33% chance. The expected utility of this “losses-framed” transaction is \$0 either way. Again, the risk averse actor would choose the certain bet, whereas the risk seeking actor would choose the gamble and risk incurring a debt of \$20. Attitude towards risk determines preferences, and preferences would be consistent regardless of how options are “framed.” However, this is not the case with Prospect Theory.

Evaluating Prospects per SEU – “Ordinary Prospect”

Before we review Prospect Theory (PT), let us define a “prospect” to mean the aggregation of possible future outcomes from a COA. This does not necessarily require an assumption that PT holds rather than SEU. SEU treats an individual’s expected utility of an “ordinary” prospect as the sum of expected utilities (expectation) of that prospect’s possible outcomes. “Ordinary” here notes that the prospect is not in the context of a game theoretical scenario. For a prospect with two possible outcomes, we have the following example:

$$U(x,p;y,q) = pu(x) + qu(y)$$

Equation 2. One individual’s expected utility of “ordinary” prospect (SEU)³⁷

- where the prospect is represented by $(x, p; y, q)$; and
- p is the probability of attaining outcome x ;
- $u(x)$ is the utility of attaining outcome x ;
- q is the probability of attaining outcome y ; and
- $u(y)$ is the utility of attaining outcome y .

However, game theoretical scenarios introduce multiple prospects because interactions between *multiple players with multiple COAs* influence each player’s respective expected utilities. We propose a concept of “equilibrium prospect” in future discussion.

Prospect theory

In contrast to SEU, prospect theory (PT) claims that the expected utility from a COA is not relative to a net asset position, but instead it follows from the *change in value* relative to a reference point.³⁸ This theory was developed by Kahneman and Tversky who determined that people make choices differently depending on how options are presented, or “framed” relative to reference points. Kahneman and Tversky explained that “the reference outcome is usually a state to which one has adapted; it is sometimes set by social norms and expectations; it sometimes corresponds to a level of aspiration, which may or may not be realistic.”³⁹

If the utilities from the COAs of a game are framed for a player as losses relative to a reference point, then PT predicts the player will “over-weight” the utilities from those COAs, as compared to their utility assuming SEU. That player is therefore likely to take greater risks to avoid those losses if currently at their reference point, than he would take for an equivalent amount of gain that would put them ahead of their reference point.⁴⁰

Returning to Schoemaker’s claim, risk-seekers would pay larger certainty equivalent for the possibility of some expected utility than would risk avoiders. However, under PT assumptions, framing of prospects would govern attitude toward risk. Certainty equivalents would vary under different conditions of wealth, unlike under SEU assumptions. Those faced with prospective losses from a reference point would be risk seeking, and those faced with prospective gains beyond a reference point would be risk averse. Figure 1 summarizes how SEU and PT influence risk propensity or risk attitude.

Utility Theory

		SEU: Asset Position	PU: Reference Point
Risk Propensity	Risk Seeking	Low % of high utility	Overweight utility when presented as loss relative to reference point
		High Certainty Equivalent (CE)	High CE
		Gamble more preferred than expected value for certain	Losses predict risk-seeking
	Risk Averse	High % of low utility	Underweight utility when presented as gain relative to reference point
		Low CE	Low CE
		Gamble less preferred than expected value for certain	Gains predict risk aversion

Figure 1. How SEU and PT Influence Risk Propensity

To illustrate, returning to the SEU example, suppose one has \$10 (which is also coincidentally her reference point) and stands to gain \$10 with 100% certainty, or gain \$30 with 33% chance. The expected utility of this prospect (including the player’s original net asset position) is still \$20 either way. However, PT predicts framing will determine attitude towards risk. When presented with gains relative to their reference point, players will be risk averse. In this case a player would prefer the certain bet of +\$10, even though he stood to possibly make a lot

more money (+\$30). Alternatively, suppose one has \$10 but stands to lose \$10 with 100% certainty, or lose \$30 with 33% chance. The expected utility of this transaction is still \$0 either way. However, since the options are now presented as losses relative to their reference point, players will be risk seeking and choose the gamble of -\$30, even though they stand to possibly lose more money. Framing determines attitude towards risk, which then determines preferences. Preferences are now *inconsistent* depending on how prospects are framed. Kahneman and Tversky posed similar prospects to numerous respondents in their research.

These findings suggested inconsistent decision making, which Kahneman and Tversky proposed is governed by a nonlinear value function with a reference point at the intersection of the axes in Figure 2. Notice that a specific amount of loss holds more value than an equivalent amount of gain:

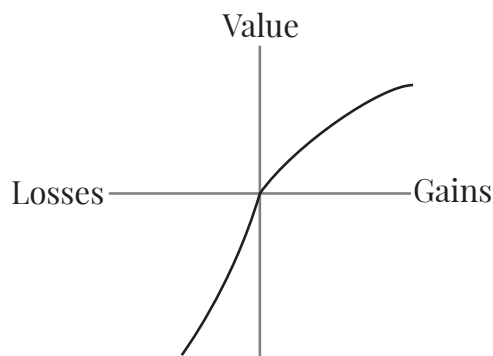


Figure 2. Value function for PT⁴¹

Kahneman and Tversky suggested that in PT, the expected utilities from COAs are modified by probability weights and utility values.⁴²

Evaluating Prospects per PT – “Ordinary Prospect”

PT treats the expected utility of a prospect as:

$$U(x,p;y,q) = \pi(p)v(x) + \pi(q)v(y)$$

Equation 3. One individual’s expected utility of “ordinary” prospect (PT)

-where $(x,p;y,q)$ is a prospect with at most two non-zero outcomes⁴³;

-where $\pi(p)$, $\pi(q)$ represent decision weights, which reflect the influence of probabilities of outcomes on the overall prospect, but are not themselves probabilities;

- $v(x)$, $v(y)$ represent subjective values of outcomes (utilities) x and y , based on whether those outcomes are gains or losses relative to a reference point;

-and one stands to receive nothing with probability $1 - p - q$, where $p+q \leq 1$.

The decision weights modify probabilities such that low probabilities tend to be over-weighted. Also, medium to high probabilities tend to be underweighted, much more so than lower probabilities are over-weighted. The utility values modify utility such that consequences are over-valued if presented as losses relative to a reference point, and undervalued if presented as gains relative to a reference point. Given these weights and values, Kahneman and Tversky discussed the “certainty effect.” This means people tend to overweight outcomes that are certain over outcomes that are merely probable. For gains this generally means sure gains are preferred to probabilistic gains. In contrast, for losses, probabilistic (and possibly larger) losses are generally preferred to sure losses. Thus, the certainty effect became the “reflection effect” for losses.⁴⁴

Therefore, to compare deterrence quantification results under SEU assumptions to results under PT assumptions, initially we found it appealing simply to modify numerical probability estimates in our expected utility functions with probability weights, and similarly to modify utilities with utility values.

However, in our review of PT literature, we could find no evidence that any of Kahneman and Tversky’s surveyed respondents literally multiplied their estimates of probabilities by probability weights, or substituted weights for probability estimates altogether. Nor could we find evidence that subjects multiplied given utilities by utility values. Therefore, Kahneman and Tversky’s modified expected utility equations seem to be equations “fitted to the data” to explain their findings, rather than equations explicitly used by their respondents.

Unfortunately, we also know of no elicitation in game theoretical context that helped formulate PT, nor do we know of any terrorist elicitation that yielded data on terrorist preferences for different prospects from critical infrastructure attacks. Therefore, we will keep the same structure of the expected utility functions in our deterrence games when we assume PT, but instead predict what COA the attacker might prefer based on Kahneman and Tversky’s findings.

Berejikian urges that theories of politics should be “based on models of the individual consistent with empirical evidence about how individuals made decisions.”⁴⁵ Because Kahneman and Tversky showed evidence that people make decisions inconsistently, Berejikian then showed how to apply PT to deterrence games. This work yielded useful insights but did not discuss CIKR risk analysis explicitly.

In addition to Berejikian’s analysis of deterrence and PT, there is some application of PT to international relations and other areas of study in the literature, but not much specific application to deterrence of terrorist attacks on CIKR. For example, Schaub writes about how PT affects the effort necessary to sustain a strategy of deterrence, but focuses more on traditional nation-state conflict.⁴⁶ Yang describes how PT compares to other decision-making theories in her dissertation on modeling bounded rationality for protecting CIKR, but does not explicitly discuss deterrence.⁴⁷

Other Biases

An et al. applied the concept of quantal response (QR) to the PROTECT model referenced earlier.⁴⁸ QR suggests that people will select non-optimal COAs with probability inversely proportional to costs of making an error.⁴⁹ This error might be observation error, similar to error described by Jenelius et al.⁵⁰, or payoff error or attacker inability to accurately estimate the probability/utility from a COA.

An et al. compared the effectiveness of their model that leveraged QR bias to the effectiveness of models that leveraged other theories of biased decision making. One such model leveraged PT, another leveraged PT and incomplete information, and various others leveraged additional different theories.⁵¹ However, these models did not explicitly quantify deterrence.

There are also efforts to indirectly model effects of biases upon defender results, as an alternative to modeling adversary decision making. Pita et al. propose a method to bound the extent of defender “sacrifice” of expected utility based on a constrained attacker deviation from the attacker’s optimal solution.⁵²

Major Takeaways – Relevant Work

In sum, previous work on topics relevant to our proposed approach captures several key points. First, there are two claims in the literature: one that full knowledge of an opponent’s capabilities and intentions is more efficacious for deterring adversaries, and one that ambiguity will be more efficacious for deterrence.

Second, games between adversaries can be modeled using complete or incomplete information, wherein players either know all elements that define the rules of the game or that knowledge is limited. Furthermore, games can also be modeled using perfect or imperfect information, wherein players either know what previous actions their opponents have taken, or they do not know all previous actions.

Third, expected utility functions reflecting value retained for protecting CIKR have been modeled as a function of budget expended to protect those CIKR. Fourth, previous work has figured out ways to model optimal investment in CIKR protection, when available resources have been insufficient to eliminate *all* risk.

Fifth, literature exploring the tenets of SEU has shown that attitudes toward risk determine preferences, and preferences are consistent regardless of how choices are presented or framed. In contrast, PT has shown that framing of options determines a player’s attitude toward risk, which then influences player preferences such that choices are made differently than expected under SEU. PT has yielded insights that decision makers may tend to “overweight” or place more emphasis on outcomes that are certain, than on outcomes that are merely probable. However, efforts to apply principles from PT to expected utility functions in CIKR deterrence games have been limited. Sixth, there have been other efforts to incorporate decision biases into decision making models.

Putting It All Together

How can we leverage these ideas to support decisions on whether to publicize or obfuscate deterrence investment information? Taquechel and Lewis explain the generic process to quantify deterrence and show how that influences CIKR risk, yielding a “deterrence portfolio” of metrics to support decision making.⁵³ We here extend this process and leverage the simultaneous game, expected utility function, exponential probability-investment relationship, and optimization concepts discussed earlier.

We also show how this methodology and resulting deterrence portfolios might change if we apply IOB and prospect biases. In doing so we extend Berejikian's concept of applying PT to deterrence, by showing how PT may affect the *quantification* of deterrence and resulting change in risk. We also examine robustness across different ways of approximating attacker intent. We explore this methodology in a case study of notional CIKR, and notional defender budgets (FEMA's Port Security Grant Program grants). We first synthesize existing ideas about cognitive biases to formulate our own ideas on how to change the deterrence quantification methodology.

Our Proposed Concept of CIKR Deterrence Cognitive Biases

Taquechel and Lewis used a "modified game approach" in 2012 in that defender risk was calculated from attacker expected utility functions from a "game." This game only explicitly showed attacker expected utility functions as outcomes, even though these outcomes depended on different defender investments.

In other words, the game did not calculate a Nash Equilibrium, for which more than one player's utility functions are needed.⁵⁴ In the current approach we include both attacker and defender expected utility functions when analyzing the deterrent effects of different defender investment options. This allows us to explore different equilibria and their implications for deterrence and risk reduction.

Information Obfuscation Bias

Previous work on deterrence quantification has discussed the concepts of credibility and signaling.⁵⁵ The present work assumes that anything signaled or publicized by the defender is credible to the attacker, but the defender may obfuscate some things an attacker might want to know.

Deterrence games may reflect four possible permutations of information the defender publicizes to (or obfuscates from) the attacker. These permutations are:

1. imperfect and complete information,
2. imperfect and incomplete information,
3. perfect and complete information, and
4. perfect and incomplete information.

We focus on the first two permutations in the present work. Thus, the attacker will not know the defender's selected COA in all deterrence games, but they may or may not know the details of possible COAs.⁵⁶

We modify the expected utility functions in incomplete information games based on an attacker's information obfuscation biases (IOB). More specifically, the present work focuses on one kind of IOB, what we call organizational obfuscation bias (OOB). This bias reflects

“organizational tendencies” – what an attacker would estimate for the impact of defender investment when they do not know the actual investment amounts. Inspired by Azaiez (2009)⁵⁷, we use three discrete levels of OOB.

An attacker’s OOB level is *neutral* if they would attribute the CIKR defender credit for a reasonable amount of defensive effort. This means that the attacker modifies their estimate of target vulnerability⁵⁸ that would result from defender optimal and suboptimal investments in a game. A *neutral* attacker assumes that if the defender were to invest *optimally* to deter, the resulting target vulnerability would be 5% for all targets in the game.⁵⁹ Also, a neutral attacker assumes that if the defender were to invest *suboptimally* to deter, the resulting target vulnerability for all targets would be 50% lower than what it was pre-deterrence.

An attacker’s OOB level is *optimistic* if they would attribute the CIKR defender too little credit for their defensive efforts. An *optimistic* attacker believes resulting target vulnerability would be 10% for all targets if the defender invested optimally,⁶⁰ and believes it would be only 25% lower than what it was pre-deterrence if the defender invested suboptimally. Finally, an attacker’s OOB level is *pessimistic* if they would attribute the CIKR defender too much credit for their defensive efforts. A *pessimistic* attacker believes resulting target vulnerability would be 1% for all targets if the defender invested optimally, and believes it would be 75% lower than what it was pre-deterrence if the defender invested suboptimally.

Prospect Bias

Prospect bias will mean that the attacker will prefer outcomes or prospects based on what Kahneman and Tversky respondents chose under similar circumstances. We will make assumptions on attacker reference points.

Importantly, Kahneman and Tversky developed Prospect Theory from eliciting preferences on prospects from survey groups, where prospects were posed as “sure gain vs probabilistic gain” or “probabilistic gain vs probabilistic gain.” Alternatively, prospects were posed as “sure loss vs probabilistic loss” or “probabilistic loss vs probabilistic loss.” One can imagine the difficulty of interviewing a terrorist to derive their preferences for prospects. Instead, we simply apply the principles of PT that resulted from analysis of Kahneman and Tversky survey results, rather than the survey methodology itself.

Approximating Attacker Intent

There are various ways to approximate an attacker’s intent in our approach. We will explore whether it makes sense to create deterrence portfolios for each of these proxies, while varying cognitive bias assumptions.

Pure or Mixed Strategy NE

If our deterrence games result in a pure Nash Equilibrium (NE)⁶¹, this means we can approximate an attacker’s intent to choose their equilibrium COA as 100%. A mixed strategy NE⁶² would mean the attacker might prefer a probabilistic distribution of different COAs, but this is conceptually problematic for our approach. Rasmussen explains why mixed strategy results can be problematic:

The number of players needed so that mixed strategies can be interpreted as pure strategies in this way depends on the equilibrium probability, since we cannot speak of a fraction of a player. For the interpretation to apply no matter how we vary the parameters of a model we would need a continuum of players.⁶³

Since we do not play our deterrence games with a “continuum of attackers,” we save for future work the study of mixed strategies as proxies for attacker intent.

Intent Ratios – Individual COAs

We explore intent ratios of individual attacker COAs, hereafter referred to as “intent ratios,” given the defender’s equilibrium solution in games with pure NE, as proxies for attacker intent.

Prospects

We consider single maximum-value prospects as proxies for attacker intent, meaning the attacker’s intent to execute that COA is 100%. This is similar to assuming the attacker’s intent is 100% for a pure NE COA. We now explore how we would do this under both SEU and PT assumptions, in game theoretical context.

Prospects and Game Theory –SEU

An alternative to our “ordinary prospect” is a prospect that reflects the result of a game theoretic equilibrium solution. An example imperfect game is:

		Defender COA	
		defend target 1	defend target 2
Attacker COA	attack target 1	(Payoff A1, Payoff D1)	(Payoff A2, Payoff D2)
	attack target 2	(Payoff A3, Payoff D3)	(Payoff A4, Payoff D4)

Figure 3. Imperfect Game

The equilibrium solution here is either a pure or mixed strategy Nash Equilibrium (NE).⁶⁴ Using the example of Figure 3, a pure NE might be (A1, D1), where the attacker selects COA “attack target 1” and the defender selects COA “defend target 1.”

Equilibrium Prospect

Thus, the attacker's equilibrium COA could be called an "equilibrium prospect." The attacker's expected utility of such a prospect could be shown:

$$U^a(b,s) = su(b)$$

Equation 4. "Equilibrium prospect": expected utility of attacker's pure NE prospect (SEU)

-where the attacker's prospect is represented by $U^a(b,s)$ whose outcome is numerically equal in this case to "Payoff A1";

-s is the "success probability" of attaining outcome b (where b = result of a successful attack against target 1); and

- $u(b)$ is the utility of attaining outcome b (the defender's loss from target 1 –which equals the attacker's gain)

Thus, in an imperfect game with a pure NE outcome, the expected utility of a prospect would be represented differently than that of an ordinary prospect as shown in Equation 2.

Ordinary Prospect Resulting From Game: "Aggregate Prospect"

We need an alternate way to represent attacker prospects if the attacker does not necessarily evaluate their prospects based on an equilibrium solution. For example, we could consider prospects *with aggregated attacker outcomes that each depend on what the defender chooses for their COA*:

$$U^a(b,s;c,t) = su(b) + tu(c)$$

Equation 5. Expected utility of attacker's prospect, for one COA

-where the attacker's prospect is represented by $U^a(b,s;c,t)$ whose outcome is numerically equal in this case to "Payoff A1" + "Payoff A2";

-s is the "success probability" of attaining outcome b (where b = result of a successful attack against target 1 given the defender was defending target 1);

- $u(b)$ is the utility of attaining outcome b (the defender's loss from target 1 given the defender was defending it –which equals the attacker's gain);

- t is the “success probability” of attaining outcome c (where c = result of a successful attack against target 1 given the defender was defending target 2);

- $u(c)$ is the utility of attaining outcome c (the defender’s loss from target 1, given the defender was defending target 2 –which equals the attacker’s gain);

The difference between Equation 2 and Equation 5 is that in the latter, outcomes are influenced by *another player* in a game theoretical scenario. In our deterrence quantification methodology, we will explore how deterrence might be quantified under assumptions of both equilibrium prospects and aggregate prospects. This may have implications for how we apply the attacker’s intent to create unconditional defender risk in our deterrence portfolios.

Prospects and Game Theory -PT

Absent a defensible way to directly incorporate PT principles into utility functions, the equilibrium prospect and aggregate prospect are the same under PT assumptions as they are under SEU assumptions in our approach. However, Kahneman and Tversky’s principles may help predict which equilibrium or aggregate prospects attackers may prefer, based on certainty and reflection effects, as opposed to net asset position.

Prospect Intent Ratios

Finally, we explore prospect intent ratios, as proxies for attacker intent. Prospect intent ratios reflect the relationship between “aggregate prospects.”

Game Types

We can now apply these biases to analyze four deterrence game types, and we will elucidate each type’s details in the case study found in the appendix. Importantly, all data and the two maritime facilities, a chemical facility and a ferry terminal, are notional in this case study, but real threat, facility vulnerability, facility consequence, and budget data would be used for a real analysis. In general, Types 1 and 2 are games of complete information. Therefore the attacker and defender are “playing the same simultaneous game.” Even though they do not know each other’s moves, they would agree with each other what the outcomes would be *if* each player made certain moves.

In contrast, Types 3 and 4 are games of incomplete information. Thus the attacker’s OOB would influence their estimates of the expected utility functions. However, the defender’s estimate of their own expected utility would be the true, unbiased value, as they obviously know their own investments. Thus, we claim this game type requires the analyst to consider the results of two different simultaneous games: the “attacker’s game”, and the “defender’s game.” We created a heuristic for producing the deterrence portfolios for this game type. This heuristic considers whether the equilibrium results of the two games predict the same COAs, or different COAs.

For game types that assume PT, we define the attacker's reference point as the maximum monetized death/injury plus total economic loss from destruction of both CIKR in the game. Thus, all prospects are technically losses relative to the reference point given our notional CIKR data, even though they are gains relative to the attacker's initial asset position. The exception is the prospect of attacker restraint, in which case there is no gain.

Regardless of utility theory, the COA chosen is obfuscated for games of imperfect information. However, if that game is also one of complete information, then the following details are publicized:

1. specific optimal and suboptimal investment amounts for each possible COA,
2. resulting post-deterrence CIKR vulnerability, and
3. available deterrence budget.

In contrast, in games of incomplete information, these details are obfuscated. Furthermore, the precise calculation of vulnerability, as an exponential function of investment, must be obfuscated from an attacker.⁶⁵

Major Takeaways – Our Approach to Applying Decision Biases to Deterrence Quantification and Risk Reduction

In sum, our approach to applying decision biases to games of CIKR deterrence quantification does the following. In modifying the approach to quantifying deterrence first introduced in Taquechel and Lewis (2012), we first focus on games that assume either imperfect and complete information, or imperfect and incomplete information. Second, we create proxies for attacker OOB, based on criteria for what they would assume about our defensive investments at various CIKR.

Third, we develop different proxies for attacker intent, which is used to estimate the desirability of various attack options, to estimate the quantification of deterrence, and to create "unconditional risk" for each CIKR attack option. Taquechel and Lewis (2012) introduced a basic proxy for attacker intent that supports the quantification of deterrence. Here, we elaborate on that concept. These proxies now vary based on whether we assume equilibrium game results will predict a single attacker COA is 100% desirable and all others are completely undesirable, or instead attackers "rack and stack" all possible COAs according to their relative attractiveness. These proxies also may change if we think an attacker will aggregate possible game outcomes into "prospects," and preferences for COAs may change if we assume PT controls attacker decision making rather than SEU. When we assume PT controls rather than SEU, we apply principles of PT to our case study, rather than Kahneman and Tversky's exact methodological approach underpinning the development of their theory. These principles include the "certainty effect", where certain outcomes are disproportionately valued over probable outcomes.

Case Study- Summary of Results and Implications

SEU

Under SEU assumptions, we evaluated deterrence portfolios across different information availability circumstances in the case study. Based on this evaluation, we now propose an approach for communicating port security grant investment decisions. We do this by proposing a question and answering it by summarizing findings from our different deterrence portfolios.

“For individual options to proxy attacker intent, is any advantage of obfuscating information (over publicizing information) consistent across attacker OOBs?”

For the attacker intent proxy of intent ratios: we saw that the defender’s unconditional risk was less when we obfuscated information than when we publicized information, for a pessimistic attacker. We also saw the same result for the other two OOBs. Thus the advantage of obfuscating information was consistent across all attacker OOBs for this attacker intent proxy.

For the attacker intent proxy of prospect intent ratios: we saw that the defender’s unconditional risk was less when we obfuscated information than when we publicized information, for a pessimistic attacker. We also saw the same result for the other two OOBs. Thus the advantage of obfuscating information was consistent across all attacker OOBs for this attacker intent proxy.

For the other two attacker intent proxy options (pure NE results and single prospects), we did not show any advantage of information obfuscation across attacker OOBs. This is because the same deterrence portfolio resulted regardless of OOB. This suggests that if we believe an attacker actually does choose an equilibrium solution or chooses the maximum value prospect with 100% certainty, we would suffer the greatest unconditional risk. There was thus in these instances *no* quantifiable advantage of obfuscating information over publicizing information.

Therefore, assuming SEU, our advantage from obfuscating port security grant allocation information seemed to be robust against two different assumptions about how the attacker evaluated COAs. However, this illustrated the potential value of collecting intelligence on attacker decision making processes, to inform decisions on how we communicate port security grant distributions. If we had confidence in knowledge of how attackers evaluate COAs, we might be more fully convinced to obfuscate all details of grant allocations and evidence of their implementation. Some of these details are restricted from public release, but inadvertent release of the restricted material would negate attempts to obfuscate that information. Other information is currently public knowledge and may be observable to adversaries.

What might stakeholders do?

The government and CIKR owners/operators would be well advised to obfuscate the details of grant investments that reduce CIKR vulnerability, rather than publicize those details, based on our notional data (see appendix).

For this example, the optimal investment at the chemical facility was \$664,566.67. This might be used to add more cameras clandestinely to monitor the maritime approaches to the chemical facility's perimeter. This would increase the likelihood that the CIKR security would detect an attacker with a backpack bomb, and would lower the overall facility vulnerability. The optimal investment at the ferry terminal was \$1,335,433.33. This might be used to clandestinely train and equip additional security guards to protect vulnerable ferry passenger crowds against a similar attacker during peak transit times.

In order for the defender to gain any advantage in this case, not only must the above investments be clandestine, but the following must be obfuscated:

1. the decision to invest optimally,
2. the exact dollar amounts of the optimal and suboptimal investments, and
3. estimates of resulting target vulnerability.

Equally importantly, the dollar amount of the port security grant must be obfuscated⁶⁶, thus creating a business case to make this specific port security grant award restricted information.

Prospect Theory

We then evaluated additional deterrence portfolios across different information availability circumstances, but under assumptions of prospect theory. We revisit the question originally posed when we assumed SEU:

(Assuming PT) "For individual options to proxy attacker intent, is any advantage of obfuscating information (over publicizing information) consistent across attacker OOBs?"

For PT, we only chose one option to proxy attacker intent, and there was no advantage of obfuscating information over publicizing information across attacker OOBs. Importantly, an optimistic attacker's preference for attacking both targets might be greater to them if we obfuscate information, than if we publicize information, under PT conditions. However, since we do not create intent ratios under PT conditions, our unconditional risk will be the same from an attack on both targets. This is regardless of whether we obfuscate or publicize grant investment information.

What might stakeholders do?

If we assume PT, it makes no difference whether the government and CIKR owners/operators publicize or obfuscate the details of grant investments that reduce CIKR vulnerability. This is based on our notional data.

However, if we focus solely on attacker expected utilities under PT assumptions rather than risk, then our comparisons between complete and incomplete information results depend on assumptions on attacker OOB. For a pessimistic or neutral attacker, we predict lower attacker expected utility if we obfuscate information. In contrast, for an optimistic attacker, we predict lower attacker expected utility if we publicize information.

Major Takeaways – Results of Case Study

Our case study with notional data yielded several insights (see appendix). First, assuming SEU controlled decision making, any defender advantage gained from obfuscating information about CIKR deterrence investments was consistent across all attacker OOBs when we assumed the attacker developed “intent ratios” for all their possible COAs, rather than preferring one COA with 100% certainty. This “defender advantage” meant that the unconditional post-deterrence risk was lower when CIKR investment information was obfuscated than it was when it was publicized to a prospective attacker.

Furthermore, we learned that defender advantage gained from obfuscating information was consistent across all attacker OOBs when we assumed the attacker aggregated game outcomes into “prospects” and prioritized those prospects proportional to their values. However, we also learned that obfuscating information yields no advantage to a defender if we assume an attacker will choose a single equilibrium solution COA, or will choose the maximum value prospect as a COA. Thus, assumptions about how our opponents perceive and rank their options dictate the relative advantages of obfuscating or publicizing our deterrence actions as CIKR defenders.

Finally, changing the utility theory assumption to PT, we found that obfuscation or publicization of information made no difference on results.

Future Research

Information Obfuscation Bias Options

We have only studied games of imperfect information, although we have studied the difference between complete information and incomplete information in the present work. Future work should examine perfect or sequential games. This may show how deterrence and risk change if we publicize or obfuscate which CIKR we are defending, versus how deterrence and risk change if we publicize or obfuscate the details of how we might defend our CIKR. Also, future work might add a factor for diminished credibility of defender signaling.

Future work might alter the OOB parameters. For example, we have set parameters that require our CIKR to have pre-deterrence vulnerability $>10\%$; otherwise our logic could not be used. Also, future work might leverage continuous OOB functions, rather than discrete OOB levels.

We have proposed the heuristic that the attacker and defender play two separate games when information is both imperfect AND incomplete. Expected utility represents outcomes of what would happen *if* the attacker attacked after grant implementation. However, the

deterrence portfolios are intended to inform decisions to publicize or obfuscate information *before* an attack, and the attacker has different information from the defender *if the latter obfuscates information*. Thus, we think this is an appropriate initial heuristic, but future work may explore alternative approaches.

We have assumed that all pre-deterrence information is known to both parties. Future work might explore outcomes when pre-deterrence information is partially or fully obfuscated from the attacker.

Utility Function Options

We have composed utility functions of monetized death and economic losses, multiplied by attacker capabilities and CIKR vulnerabilities. Future work may incorporate attacker and defender budget and expenditures into those functions.

Future work might change the elimination fraction in the vulnerability-investment term in the expected utility function. We assumed 5%; future work may test results' sensitivity to changes in this input. It also might vary the slopes of the vulnerability-investment curves to reflect nuances of how certain investments reduce CIKR vulnerability to attack. Future work also might incorporate the attacker's option to invest to improve their attack capabilities, during the same time period as the defender is implementing their grant investments.

Future work might explore results when we treat the attacker's expected utility of attacking both targets simultaneously as the combination of individual expected utilities, rather than a joint probability of attacking both targets multiplied by the sum of both targets' human life and economic consequence. We chose the latter approach to model an assumption that attacks would be launched simultaneously, but attackers may prefer to "stagger" individual target attacks that are part of a larger, coordinated port-wide effort.

Finally, future work may modify the consequence portion of the utility functions to incorporate secondary economic effects of attacks. Existing risk models incorporate such factors; we chose not to incorporate secondary effects in this approach, but there is no reason not to consider incorporating them in future work.

CIKR Data Options

Future work may explore the assumption that attacker capabilities to attack CIKR targets differ; here we have simplified and assumed capabilities are the same.

Furthermore, we assume we do not have sufficient grant funding to reduce vulnerability of all CIKR under consideration to the elimination fraction (here 5%). Thus, future work may alter this assumption and so we would not necessarily need to optimize.

Deterrence Through Capability Reduction, Security, Resilience?

The present work assumed that only vulnerability reduction investments were made to deter attacks as part of the Port Security Grant program. However, future work might examine the deterrence effects of resilience investments upon deterrence portfolios, across information availability circumstances and utility theories.⁶⁷ Such investments might reduce the likelihood of loss *given a successful attack*. Future work also might model deterrence effects of reducing attacker capabilities.

Optimization Options

We have used mathematically optimal investments as one defender deterrence COA and have also used an arbitrary suboptimal investment.⁶⁸ However, the assumption is that suboptimal investment means the defender spends their entire available deterrence budget. Future work may assume that the defender spends less than their entire budget when they allocate suboptimally.

Furthermore, we could treat the defender's objective function as minimization of the attacker's expected utility, rather than maximization of their own utility, and compare results.

We have formally solved the optimization for a two target game. Future work that considers more than two targets may require computer programming to solve for optimal (or near-optimal) solutions.

Prospect Bias Options: Different Reference Point

We have assumed the attacker only overweights expected utility under PT assumptions. What if the attacker stands to exceed their reference point instead? Then, under PT assumptions, the attacker would normally underweight these expected utilities.

Biasing Pre-Deterrence Expected Utilities

We could bias the attacker's pre-deterrence expected utilities as well as their post-deterrence utilities. The former are still prospects (albeit with only one outcome each) *prior* to defender deterrence investments.

Reference Point vs Status Quo: Does Actor "Domain" Influence How Prospect Framing Changes Preferences?

A review of Kahneman and Tversky's work on PT shows that the respondents' reference points are usually equal to the status quo, or what assets the respondents have when evaluating prospects. However, they also acknowledge that the reference point might not always be equal to the status quo.⁶⁹

We have proposed a notional game in which the attacker will evaluate prospects, but has NOT adapted their status quo as their reference point. However, we have no data to support preferences in real world situations. We hypothesize that such data may test the current theory that prospective gains exceeding a reference point are always underweighted, or that prospective losses falling behind a reference point are always overweighted. In their formulation of Cumulative Prospect Theory, which revised the original PT, Kahneman and Tversky identified scenarios where gains are actually overweighted (an actor becomes risk seeking) and where losses are underweighted (an actor becomes risk averse). However, like their original PT analysis, this analysis was not done in game theoretical context.

Future analysis might incorporate the concept of actor domain that Linnington discusses.⁷⁰ Linnington claims that actor domain is “a state in which the actor resides, that of losses or gains. If the actor feels he is in a position of strength, he is in the domain of gains, and conversely, if his position is weak, he is in the domain of losses.”⁷¹ We might adopt this to mean a position of strength reflects a status quo that exceeds a reference point, when evaluating prospects. In contrast, a position of weakness might mean a status quo that falls short of a reference point.

Kahneman and Tversky also give examples which suggest that an actor is in a certain domain.⁷² However, actor domain relative to a reference point is not explicitly distinguished from how a prospect relates to a reference point. They discuss “shifting” of reference points and how that leads to inconsistency of preferences, but without explicitly accounting for the relationship between actor domain and reference point.

For example, in a case where someone’s status quo is less than their reference point, their risk seeking or overweighting of utility increases.⁷³ This is our attacker in our example. And, they should normally underweight the expected utility of any prospective gains beyond their reference point. However, this seems to beg the question: if an attacker is in the “domain of loss” relative to their reference point when evaluating prospects, should they actually *overweight* prospects that surpassed their reference point, but perhaps to a lesser extent than prospects that approached but did not exceed the reference point? This would be a “diminishing returns” effect. Intuition suggests any prospects that increase one’s assets will be valuable if one starts from a position of weakness, but that value may attrite past a certain point.

Future experimentation could show whether changing the status quo relative to the reference point influences the change in preferences for prospects. Kahneman and Tversky claim, “an essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states.”⁷⁴ But, actor domain could be an “initial” state; perhaps it could influence how changes are perceived. Kahneman and Tversky acknowledge that there is evidence that “initial entitlements” do matter in the evaluation of prospects.⁷⁵

We could infer player reference points from the results of Kahneman and Tversky’s elicitation⁷⁶, but they did not do the inverse: structuring the elicitation so as to front load the respondents with pre-determined reference points. Thus, future work could *elicit* preferences for differently framed prospects, in the context of explicitly given reference points, to see if the original Kahneman and Tversky preference distributions hold.

Temporal Options – Opportunistic vs Methodical Attackers?

There may be opportunities to introduce temporal complexities into this analysis. Das and Teng point out that experimental psychologists, including Tversky, have argued that situational factors influence risk taking more than do dispositional traits.⁷⁷ Das and Teng claim that longer-term decision-making is less constrained by situation-specific decision factors such as those predicted by PT, and dispositional or context-neutral utility calculations are more prevalent. Port security grant implementation can take time. So, depending on the assumed timelines of attacker and defender deliberation in deterrence games, we may be able to discount one of the utility theories. Perhaps the biases of PT are more salient when terrorists are opportunistic, and perhaps SEU is more salient when they have more time to deliberately plan.

Direct Incorporation of PT Principles into Utility Functions

As mentioned, we did not leverage utility functions that directly incorporated Kahneman and Tversky's principles from the literature. Future work may leverage Verendel's Kahneman and Tversky modified expected utility functions as *input* to a deterrence quantification game, such that the equilibrium results could be compared.

Prospect Theory and Information Availability Circumstances

Kahneman and Tversky claimed that ambiguity might influence decision weights.⁷⁸ They did not explicitly survey respondents with prospects that involved incomplete or ambiguous information.

Also, Kahneman and Lovallo wrote: “[t]he experimental evidence indicates that the certainty effect is not eliminated when probabilities are vague or ambiguous, as they are in most real life situations – and the effect may even be enhanced.”⁷⁹

Therefore, any future elicitations of preferences for prospects in a game theoretical context might include options where information is partially obfuscated. Future work may calibrate the respondent's preferences in absence of complete information, to approximate organizational obfuscation biases.

Also, future work might address the effect of cognitive biases on deterrence in iterative Stackelberg games of multiple rounds.

Attacker Intent Proxy Options

We have not focused on games that could result in a mixed strategy equilibrium in this work. Mixed strategies reflect what one player should do to make their opponent indifferent between choices; whereas intent ratios reflect what one player might prefer based on

comparison of their own choices. If we think terrorists might be performing game-theoretical analyses where they evaluate our strategies, exploring mixed strategy results might be a worthwhile exercise. However, if we think that is unlikely, and that they only compare what they believe their own outcomes will be, then perhaps we do not need to explore this option. It is also possible that using intent ratios instead of pure or mixed NE results to proxy attacker preferences avoids the need to speculate how an attacker might estimate our own expected utility functions.

Also, one way to interpret mixed strategy results is a proportion of times that one player should execute one COA vs other COAs. However, we surmise that if an attacker attacks, the defender will adapt and implement additional security measures, thus changing the expected utility function values from the original game. This might render the original equilibrium mixed strategy solution a *sub-optimal solution* for the attacker. There is work on adaptation and learning in games, especially sequential or perfect information games, which might be explored for applicability to deterrence quantification and impact on risk.

Applicability to Cybersecurity?

This approach may be applicable to the cybersecurity world. Cybersecurity and deterrence have been analyzed in the literature, but a cursory Internet search for “cybersecurity, deterrence and prospect theory” yielded no academic literature on these topics.

Incorporation into Existing Risk Models?

Future work may explore how to incorporate this proposed methodology into existing risk models. For example, the United States Coast Guard’s Maritime Security Risk Analysis Model, or MSRAM, is a risk tool that leverages threat, vulnerability, and consequence judgments to “conduct long-term strategic resource planning, identify capabilities needed to combat future terrorist threats, and identify the highest risk scenarios and targets in the maritime domain.”⁸⁰

Major Takeaways – Future Research

In sum, this approach to quantifying deterrence and exploring the effects of obfuscating or publicizing CIKR deterrence investment information is ripe for further exploration. First, we encourage future research in the area of games of perfect information, wherein a defender would make their CIKR investment decisions known to prospective attackers in an effort to deter attacks. These decisions would entail which CIKR received grant investments. The different effects of complete and incomplete information could then be explored given the existence of perfect information. The results could then be compared to those of the case study in this paper to inform decision making.

Second, we urge modifying the expected utility functions used in deterrence games by incorporating defender budgets and best estimates of attacker budgets. Also, the consequence component of the expected utility functions might be modified, to test results sensitivity.

Third, we encourage exploration of this process when estimates of attacker capabilities differ. Fourth, we suggest consideration of how investments to improve CIKR resilience and/or mitigate attacker capabilities to launch attacks might influence the outcome of deterrence games, thereby informing decisions on whether to obfuscate or publicize deterrence investment information.

Fifth, we advocate exploring this approach under different assumptions about what quantity we are trying to “optimize.” Sixth, we urge exploration of this process under different assumptions derived from principles of Prospect Theory. We have only leveraged a couple of principles in our current approach. Seventh, we suggest exploring games where mixed strategies result, as we have only focused on games that result in pure strategy Nash Equilibria here.

Eighth, and perhaps most importantly to practitioners, we suggest efforts to incorporate theoretical findings into real-world risk models.

Conclusion

Taquechel and Lewis claimed, with respect to the initial results of their deterrence quantification methodology: “In order to generalize these findings, any advantage of a specific information availability circumstance must be robust given utility theory assumptions.”⁸¹

We have shown how this deterrence quantification methodology can be expanded to account for organizational obfuscation biases and prospect bias. This has implications for the deterrence effectiveness of our potential CIKR investments, and for the resulting defender risk. The advantage of obfuscating information was NOT shown to be robust across different utility theories. This was because the advantage of information obfuscation held under SEU assumptions, but there was neither quantifiable advantage nor disadvantage under PT assumptions. Thus, more analysis is needed.

We summarize here our extensions of previous work. First, we have extended Lebow and Stein’s work by claiming that the application of measurable deterrence to CIKR risk analysis and protection must account for different utility calculations. Additionally, we have extended Taquechel and Lewis’ generic approach to quantifying deterrence and shown how cognitive biases may influence “deterrence portfolios” to support decision making. Furthermore, we have extended Berejikian’s concept of applying PT to deterrence by showing how PT may affect the *quantification* of deterrence and resulting change in risk.

Chilton and Weaver⁸² have suggested that deterrence theory is applicable to many of the 21st century threats the US will face, but how the theory is put into practice, or “operationalized”, needs to be advanced. We believe this extension of Taquechel and Lewis’ original deterrence quantification work, as explained herein and in the case study appendix, helps advance “operationalization” of deterrence theory to a very relevant 21st century threat: terrorist attacks against CIKR. And, Berejikian asserts that prospect theory can explain why deterrence succeeds and fails.⁸³ We have taken a slightly different approach: we explored whether prospect theory influences the relative deterrence effectiveness of different investments and resulting risk, not deterrence success or failure per se.

Revisiting Moran vs. Chilton/Weaver, we have shown experimentally that ambiguity can enhance the *end result* of deterrence, or the unconditional risk in deterrence portfolios. When we compared SEU results for complete vs. incomplete information, we found that obfuscating information resulted in lower average defender post-deterrence unconditional risk. However, when we compared PT results for complete vs. incomplete, we found that information obfuscation made no difference in the *end result of deterrence*. Thus, a definitive answer to settle the “Moran-Chilton/Weaver debate” is not known at this time. This is encouraging for the “prospects” of future work in this area!

Glossary

Expected Utility Theory (EUT): a theory that assumes people find the best possible solution from among all known options, choosing a solution that will maximize their expected utility. They act according to their preferences, and preferences are consistent regardless of how options are presented

Prospect Theory (PT): an alternative theory of utility which has shown experimentally that decisions made may be inconsistent with EUT, depending on how options are presented or framed

Critical infrastructure and key resources (CIKR): systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters⁸⁴

Information Obfuscation Biases (IOB): decision-making biases resulting from information imperfection and information incompleteness

Course of Action (COA): in this case, the defender's options for how to invest at CIKR, or the attacker's choice of what CIKR to attack

Strong Stackelberg Equilibrium (SSE): a specific type of equilibrium occurring in modeled attacker-defender security games, which predicts what utility-maximizing COA an attacker who observes the defender's security posture would take

CREATE: Center for Risk and Economic Analysis of Terrorism, at the University of Southern California

PROTECT: Port Resilience Operational/Tactical Enforcement to Counter Terrorism model, developed by CREATE for the U.S. Coast Guard to plan/execute homeland security patrols

Subjective Expected Utility (SEU): the utility from a specific course of action, modified by the subjective probability of attaining that utility

Certainty Equivalent (CE): the maximum amount someone would pay for some expected utility

Quantal Response (QR): a theory that suggests that people will select non-optimal COAs with probability inversely proportional to costs of making an error

Organizational Obfuscation Bias (OOB): an information obfuscation bias that reflects organizational tendencies – reflected in what a terrorist organization might estimate to be the impact of a CIKR defender's investment, when the terrorist does not know the actual investment levels

Nash Equilibrium (NE): Theoretical equilibrium solution to a non-cooperative game. A pure NE means that in theory each player should prefer their equilibrium COA with 100% intent. If all players chose their respective equilibrium COAs during one round of the game, the NE solution means each player gets their best possible expected utility given all other players are simultaneously trying to maximize their own expected utility. For a mixed equilibrium, "mixed strategies" reflect what one player should do to make their opponent indifferent

between choices; in a game with two COAs, the first player's preferences in a mixed strategy reflect a probabilistic distribution where they should execute one COA $x\%$ of the time, and the other $1-x\%$ of the time. Technically a pure strategy NE is one kind of mixed strategy NE.

Maritime Security Risk Analysis Model (MSRAM): a risk tool developed by the U.S. Coast Guard, which leverages CIKR threat, vulnerability, and consequence information to conduct long-term strategic resource planning, identify capabilities needed to combat future terrorist threats, and identify the highest risk scenarios and targets in the maritime domain

Port Security Grant Program (PSGP): a Federal Emergency Management Agency (FEMA) program to prioritize investments to secure CIKR

Return on Investment (ROI): in this case, the CIKR risk reduced per dollar spent at that CIKR

Appendix A: Case Study: Port Security Grants

Here we apply the deterrence quantification methodology and game types to a case study of two notional CIKR. Our two CIKR are target A, a chemical facility in a port, and target B, a ferry terminal nearby. For our example deterrence games, the attacker capability to attack all permutations of these targets is the same, the pre-deterrence vulnerability of the chemical facility is 0.25 and the pre-deterrence vulnerability of the ferry terminal is 0.50, the defender has a vulnerability reduction (deterrence) budget of \$2,000,000 from a possible port security grant, the maximum economic consequence of losing the chemical facility is \$5,000,000, and the maximum economic consequence of losing the ferry terminal is \$10,000,000.

The estimated deaths from an attack on the chemical facility are the same as those from an attack on the ferry terminal, and are monetized.⁸⁵ The budget data could be estimated from previous PSGP applications and the risk data from an existing CIKR terrorism risk model. As a baseline, we first create deterrence portfolios under assumptions of SEU and complete information.

Data – Complete vs Incomplete Information- SEU

We compare deterrence portfolios that result from games of imperfect and complete information, or Types 1 and 2 (SEU), and from games of imperfect but *incomplete* information, or Types 3 and 4 (SEU). Each type has subtype “a”, reflecting our different proxies for attacker intent. We use this form of an imperfect game:

	Optimal Investment	Suboptimal Investment
Attack A	$U_e T _{A, \$_{opt}}^{post}, U_e G _{\$_{opt}, A}^{post}$	$U_e T _{A, \$_{sub}}^{post}, U_e G _{\$_{sub}, A}^{post}$
Attack B	$U_e T _{B, \$_{opt}}^{post}, U_e G _{\$_{opt}, B}^{post}$	$U_e T _{B, \$_{sub}}^{post}, U_e G _{\$_{sub}, B}^{post}$
Attack A+B	$U_e T _{AB, \$_{opt}}^{post}, U_e G _{\$_{opt}, AB}^{post}$	$U_e T _{AB, \$_{sub}}^{post}, U_e G _{\$_{sub}, AB}^{post}$
Restrain	$U_e T _{O, \$_{opt}}^{post}, U_e G _{\$_{opt}, O}^{post}$	$U_e T _{O, \$_{sub}}^{post}, U_e G _{\$_{sub}, O}^{post}$

Figure 4. Generic Simultaneous Game for Case Study

Complete Information

Type 1 (SEU): Intent ratios

For Type 1(SEU) and Type 2(SEU), both simultaneous games, our pure Nash Equilibrium result predicts the attacker will attack the ferry terminal and the defender will invest by distributing optimally the available grant money amongst the two CIKR. The deterrence portfolios for Type 1 (SEU) reflect the assumption that intent ratios are a suitable proxy for attacker intent. The deterrence portfolio for optimal investment in a Type 1(SEU) game is:

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt} |_A = -9.73\% & E_{\$opt} |_B = -9.25\% \\ E_{\$opt} |_{AB} = 56.52\% & E_{\$opt} |_0 = n/a \end{array} \right) \\ \overline{R} |_{\$opt,k}^{post} = \$17,910,708.57 \\ \Delta \left(\overline{R} |_{\$opt} \right) = \$25,107,293.30 \\ ROI_{\$opt}^A = 0.88, ROI_{\$opt}^B = 0.44 \end{array} \right]$$

Figure 5. Deterrence portfolio of optimal investment, Type 1(SEU)⁸⁶

$E_{\$opt} |_A$ represents the quantified deterrence effectiveness of optimal deterrence investments, given the attacker attacks the chemical facility. The other three $E_{\$opt} |$ terms represent quantified deterrence effectiveness of optimal deterrence investment given the other three attacker COAs. Positive results mean the attacker is incentivized to attack that target or combination of targets, whereas negative results mean the attacker is deterred from attacking that target or combination of targets. In this example the attacker is incentivized to attack the chemical facility, and is incentivized to attack the ferry terminal, but is deterred from attacking both simultaneously. One might intuit that the incentive to attack multiple targets simultaneously would exceed that of attacking a single target, but here we show that is not necessarily true.

$R |_{\$opt,k}^{post}$ represents averaged post-deterrence unconditional risk to the defender given optimal investments, averaged across all four possible attacker COAs, and leveraging intent ratios from the deterrence game. $\Delta(R |_{\$opt})$ represents the change from averaged pre-deterrence unconditional risk to averaged post-deterrence unconditional risk, given optimal deterrence investments. In this case averaged pre-deterrence risk was monetized as \$43,018,001.87 so risk has indeed decreased as a result of our optimal deterrence investment.

Finally, $ROI |_{\$opt}^A$ represents the return on the optimal investment at the chemical facility, and $ROI |_{\$opt}^B$ represents the return on the optimal investment at the ferry terminal.

Type 1a (SEU): Pure NE result

What if our deterrence portfolio reflects the assumption that the attacker's intent is proxied by the pure Nash Equilibrium outcome rather than by intent ratios as reflected in Figure 5? Note the absence of deterrence quantification metrics below in Figure 6; we do not feel this metric adds anything meaningful for a decision maker when intent=100% as reflected by a pure NE. This is because the quantification of deterrence will always be negative so the attacker will always be incentivized to attack rather than be deterred, which seems unsatisfying. In this case there is one post-deterrence risk value versus an average, since we use the equilibrium result only, and change in risk will use pre-deterrence risk from the same attacker COA reflected by the deterrence game attacker NE COA:

$$\left[\begin{array}{l} R|_{\$opt,B}^{post} = \$90,944,659.34 \\ \Delta(R|_{\$opt}) = \$40,402,597.52 \\ ROI_{\$opt}^A = 0.46, ROI_{\$opt}^B = 0.23 \end{array} \right]$$

Figure 6. Deterrence portfolio of optimal investment, Type 1a(SEU)

Is it meaningful to compare Figure 6 to Figure 5? Perhaps not; unconditional risk from a NE result would be much higher due to a lack of “dampening” intent ratio <100%. Conceptually, are we more comfortable assuming attacker intent will be 100% if the game yields a pure NE? Or are we more comfortable hedging for the possibility that the attacker may not pick an equilibrium solution, and intent ratios are sufficient proxies for attacker intent? We can at minimum show variations in deterrence portfolio data across these assumptions.⁸⁷

Type 2 (SEU): Prospects – Single Result

What if the attacker evaluates prospects according to the “aggregate prospect” approach? When only one prospect is preferred to all others, in Type 1a, we do not have data to support quantification of deterrence, just as with a pure NE result. In this case, attacking the ferry terminal presents the maximum value attacker prospect regardless of whether the defender invests optimally or suboptimally.⁸⁸ Attacker intent in this case will still be to attack the ferry terminal with 100% intent, but this intent considers both possible defender COAs. Since the NE predicts what we should do, we show one post-deterrence risk value versus an average. The deterrence portfolio is the same as Type 1(SEU) and is shown:

$$\left[\begin{array}{l} R|_{\$opt,B}^{post} = \$90,944,659.34 \\ \Delta(R|_{\$opt}) = \$40,402,597.52 \\ ROI_{\$opt}^A = 0.46, ROI_{\$opt}^B = 0.23 \end{array} \right]$$

Figure 7. Deterrence portfolio of optimal investment, Type 2(SEU)

Type 2a (SEU): “Prospect intent ratios”

An attacker would only execute one COA, but we could hedge for non-maximization and estimate their intent for each individual prospect, creating a probability reflecting the ratio of the attacker expected utility of that prospect to the total attacker expected utility from all prospects. The deterrence portfolio in this case is:

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$} |_A = 52.12\% & E_{\$} |_B = -45.18\% \\ E_{\$} |_{AB} = 77.10\% & E_{\$} |_0 = n/a \end{array} \right) \\ \overline{R|_{\$opt,k}^{post}} = \$20,515,854.65 \\ \Delta(\overline{R|_{\$opt}}) = \$22,502,147.22 \\ ROI_{\$opt}^A = 0.79, ROI_{\$opt}^B = 0.39 \end{array} \right]$$

Figure 8. Deterrence portfolio of optimal investment, Type 2a(SEU)

Notice that the deterrence effectiveness terms are with respect to all possible defender investments, rather than just optimal investment. This is because the attacker considers prospects that account for all possible deterrence investments as with Type 2. However, we still calculate unconditional risk, change in unconditional risk, and ROI *given the equilibrium* COA. This is because we assume that COA will have been implemented when the attacker attacks, and the risk results and ROI should reflect that implementation.

In comparison to the results of the Type 1 game, where intent ratios governed unconditional risk, notice that the attacker is now deterred (rather than incentivized) from attacking the

chemical facility, and is more incentivized to attack the ferry terminal. This may be because across defender COAs, the aggregate intent ratio for attacking the ferry terminal is larger than it is with respect to only one defender COA. The attacker is also more deterred from attacking both targets simultaneously.

Unfortunately, the average unconditional risk in Type 2a is greater, also lowering the change in risk and ROI. The attacker would be more incentivized to attack the ferry terminal, a higher consequence target, if they evaluated aggregate prospects, as opposed to evaluating game outcomes individually. This is important to consider if we are not sure how an attacker evaluates our deterrence investment communications.

Overall Findings – SEU, Complete Information

Overall, we see that unconditional defender risk is highest if we assume the attacker will pick, with 100% certainty, either the pure NE result attacker COA to attack the ferry terminal, or the maximum value prospect across all defender COAs, which coincidentally is to also attack the ferry terminal. This is the most conservative assumption, and is also most consistent with traditional game theoretical and prospect evaluation approaches. Unfortunately, if we are interested in the quantification of deterrence as a stand-alone metric, these assumptions do not facilitate that.

For proxies of attacker intent where we can quantify deterrence, we would caution against claiming deterrence investments are quantifiably more effective (or more ineffective) against different attacker COAs, for different attacker intent proxies. This is because the deterrence effectiveness of a single investment COA (e.g. mathematically optimal distribution between targets) is not the same concept as the aggregate deterrence effectiveness of our “strategy space” or all possible investments. But, we have more confidence that the averaged defender unconditional risk is higher if we assume the attacker compares prospects than if they compare individual outcomes given the defender’s equilibrium COA. Thus, we might take a conservative approach and assume the attacker evaluates “aggregate prospects” rather than “equilibrium prospects.”

Incomplete Information

Type 3 (SEU): Intent ratios

Under assumptions of incomplete information in Type 3(SEU) and Type 4(SEU), as with the previous types, our pure Nash Equilibrium result predicts the attacker will attack the ferry terminal and the defender will invest optimally. This holds across all 3 OOBs.

As with Type 1(SEU), the deterrence portfolios for Type 3 (SEU) reflect the assumption that intent ratios are a suitable proxy for attacker intent. We now show deterrence portfolios for Type 3(SEU) games, one for each OOB:

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -74.24\% & E_{\$opt|B} = 12.88\% \\ E_{\$opt|AB} = 96.52\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R_{\$opt,k}^{post}} = \$16,990,866.67 \\ \Delta \left(\overline{R_{\$opt}} \right) = \$26,027,135.20 \\ ROI_{\$opt}^A = 0.91, ROI_{\$opt}^B = 0.45 \end{array} \right]$$

Figure 9. Deterrence portfolio of optimal investment, Type 3(SEU), PESSIMISTIC ATTACKER

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -70.84\% & E_{\$opt|B} = 14.58\% \\ E_{\$opt|AB} = 82.92\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R_{\$opt,k}^{post}} = \$16,703,383.63 \\ \Delta \left(\overline{R_{\$opt}} \right) = \$26,314,618.24 \\ ROI_{\$opt}^A = 0.92, ROI_{\$opt}^B = 0.46 \end{array} \right]$$

Figure 10. Deterrence portfolio of optimal investment, Type 3(SEU), NEUTRAL ATTACKER

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -66.77\% & E_{\$opt|B} = 16.61\% \\ E_{\$opt|AB} = 66.65\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R_{\$opt,k}^{post}} = \$16,359,430.71 \\ \Delta \left(\overline{R_{\$opt}} \right) = \$26,658,571.16 \\ ROI_{\$opt}^A = 0.93, ROI_{\$opt}^B = 0.46 \end{array} \right]$$

Figure 11. Deterrence portfolio of optimal investment, Type 3(SEU), OPTIMISTIC ATTACKER

When we obfuscate information about our notional port security grant investments, optimal investment deters the attacker from attacking the ferry terminal, as opposed to incentivizing them to attack that target when we publicize information. Even for an optimistic attacker, we

show that deterrence is quantifiably MORE effective against simultaneous attacks on both the chemical facility and ferry terminal, when we obfuscate information. Also, we show that deterrence is quantifiably *less ineffective* against attacks on (i.e. the attacker is *less incentivized* to attack) the chemical facility, the lower consequence target, when we obfuscate information.

Ensuring a game of incomplete information thus means, in this specific case, that we incentivize an attacker to attack a lower value target, and deter them from attacking a higher value target, which may be a decision maker's goal. Examining the rest of each portfolio, average unconditional risk is always lower for incomplete information than for complete information, regardless of OOB. Notice that, perhaps counter to intuition, the average unconditional risk actually *decreases* as attacker confidence increases. It is possible that the averaging of unconditional risk *across attacker COAs* "smoothes out" any increase in unconditional risk we might expect given *one attacker COA*, as attacker confidence increases.

An interesting side note is that incomplete information decreases what we call an "attractiveness differential" between the chemical plant and ferry terminal, increasingly so as attacker confidence increases. This is because the difference between overall intent to attack the chemical plant and overall intent to attack the ferry terminal is smaller in the games of incomplete information than in complete games. This may be another reason to obfuscate information, although again, the entire deterrence portfolio should be considered.⁸⁹

Type 3a (SEU): Pure NE result:

The deterrence portfolio when we use intent=100% is the same for all three OOBs. This is because OOB changes attacker intent ratios, but when intent ratios are not used, the defender's game yields the same results regardless of attacker OOB, so the deterrence portfolio for this game is the same as Figure 6.

$$\left[\begin{array}{l} R|_{\$opt,B}^{post} = \$90,944,659.34 \\ \Delta(R|_{\$opt}) = \$40,402,597.52 \\ ROI_{\$opt}^A = 0.46, ROI_{\$opt}^B = 0.23 \end{array} \right]$$

Figure 12. Deterrence portfolio of optimal investment as a pure NE defender COA, Type 3a(SEU)

Type 4 (SEU): Prospects – single result

If the attacker evaluates prospects according to the "aggregate prospect" approach, even under assumptions of incomplete information, attacking the ferry terminal is the attacker

COA that presents the maximum value prospect. Attacker intent in this case will still be to attack B with 100% intent. The deterrence portfolio is shown:

$$\left[\begin{array}{l} R|_{\$opt,B}^{post} = \$90,944,659.34 \\ \Delta\left(R|_{\$opt}\right) = \$40,402,597.52 \\ ROI_{\$opt}^A = 0.46, ROI_{\$opt}^B = 0.23 \end{array} \right]$$

Figure 13. Deterrence portfolio of optimal investment, Type 4(SEU)

Type 4a (SEU): “Prospect intent ratios”

What if the attacker evaluates prospect intent ratios, but this time with incomplete information? For a pessimistic attacker:

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -17.80\% & E_{\$opt|B} = -9.67\% \\ E_{\$opt|AB} = 74.29\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R|_{\$opt,k}^{post}} = \$18,169,105.15 \\ \Delta\left(\overline{R|_{\$opt}}\right) = \$24,848,896.72 \\ ROI_{\$opt}^A = 0.87, ROI_{\$opt}^B = 0.43 \end{array} \right]$$

Figure 14. Deterrence portfolio of optimal investment, Type 4a(SEU), PESSIMISTIC ATTACKER

For a neutral attacker,

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -20.40\% & E_{\$opt|B} = -3.20\% \\ E_{\$opt|AB} = 53.56\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R|_{\$opt,k}^{post}} = \$17,478,581.69 \\ \Delta \left(\overline{R|_{\$opt}} \right) = \$25,539,420.18 \\ ROI_{\$opt}^A = 0.89, ROI_{\$opt}^B = 0.44 \end{array} \right]$$

Figure 15. Deterrence portfolio of optimal investment, Type 4a(SEU), NEUTRAL ATTACKER

For an optimistic attacker,

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -19.41\% & E_{\$opt|B} = 1.36\% \\ E_{\$opt|AB} = 33.29\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R|_{\$opt,k}^{post}} = \$16,918,133.28 \\ \Delta \left(\overline{R|_{\$opt}} \right) = \$26,099,868.59 \\ ROI_{\$opt}^A = 0.91, ROI_{\$opt}^B = 0.45 \end{array} \right]$$

Figure 16. Deterrence portfolio of optimal investment, Type 4a(SEU), OPTIMISTIC ATTACKER

Here we see that the attacker, if evaluating aggregate prospects under incomplete information circumstances, would be incentivized to attack the higher consequence target instead of deterred as in Type 3(SEU) games (with the exception of the optimistic attacker). And again we see the defender's average post-deterrence unconditional risk is higher, for all three attacker OOBs, when the attacker evaluates prospect intent ratios than when they evaluate intent ratios based on equilibrium solutions.

Also, again we see the counterintuitive result that average unconditional risk DECREASES as attacker confidence increases. And information incompleteness decreases the "attractiveness differential" between the chemical plant and ferry terminal, *when using an aggregate intent ratio*, just as it did when using intent ratios for equilibrium solutions.

Overall Findings – SEU, Incomplete Information

As with complete information, under conditions of incomplete information we see that unconditional defender risk is highest if we assume the attacker will pick, with 100% certainty, either the pure NE attacker COA to attack the ferry terminal, or the maximum value prospect across all defender COAs, which coincidentally is to also attack the ferry terminal.

For attacker intent proxy options where we can quantify deterrence, we see that unconditional risk is greater, across all OOBs, if we assume the attacker evaluates prospect intent ratios as opposed to intent ratios. Thus, as with assumptions of complete information, under incomplete information circumstances we might conservatively assume the attacker evaluates “aggregate prospects” instead of “equilibrium prospects.”

Data – Complete VS Incomplete Information – PT

Complete Information

Type 1(PT): Feasibility of Creating Intent Ratios for Attacker Intent Proxy

Our notional game yields the same pure NE solution as under SEU assumptions. However, intent ratios under PT assumptions should reflect desirability of individual attacker COAs, based on principles of Prospect Theory derived by Kahneman and Tversky (hereafter referred to as “KT principles”).

When Kahneman and Tversky elicited preferences for prospects during their development of Prospect Theory, the results yielded what we will hereafter refer to as “KT preference distributions” for the prospects. For example, they posed the following question⁹⁰:

“Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

Decision (i) Choose between:

- A. a sure gain of \$240 [84%]
- B. 25% chance to gain \$1000 and 75% chance to gain nothing [16%]

Decision (ii) Choose between:

- C. a sure loss of \$750 [13%]
- D. 75% chance to lose \$1000 and 25% chance to lose nothing [87%]”

The bracketed percentages represent examples of preference distributions: here the proportion of the sampled respondents who preferred A or B in Decision (i), and preferred C or D in Decision (ii). One might say these percentages represent intent ratios for the respondents. One could even surmise this could proxy a preference ratio for an individual – e.g. if offered Decision (i) 100 times, 84 times one respondent would prefer a sure gain of \$240; the other 16 times they would prefer the risky prospect.

However, the KT preference distributions shown in the above example and in many of Kahneman and Tversky's other published examples reflect preferences elicited outside the context of a non-cooperative competition. Thus, for our present work, we are not comfortable trying to derive insights from KT preference distributions to help approximate intent ratios as proxies for attacker preferences under Prospect Theory assumptions.

Alternatively, one might create expected utility functions to use in our deterrence game that directly incorporate KT principles, and then evaluate intent ratios as we did in Type 1 (SEU). Verendel⁹¹ leverages into his utility functions (with some slight modifications) the decision weights and values from Cumulative Prospect Theory, which Tversky and Kahneman proposed to advance their original theories.⁹² However, Verendel's equations use KT principles that make assumptions about the relationship between probabilities of all outcomes in a prospect, and assumptions about the relationship between utilities of various outcomes, but these relationships do not hold in our work.⁹³ Therefore, we were not confident we could defensibly apply his utility functions in our deterrence games.

Type 1a (PT): Feasibility of Leveraging Pure NE Result for Attacker Intent Proxy

How could we leverage a pure NE result under KT principles, such that we can compare effects of obfuscating information to the effects of publicizing information?

First, we might create utility functions that directly leverage KT principles and see if a pure NE is the game result. Unfortunately, the same issue we encountered in reviewing Verendel's work applies here.

Second, we might use insights from Metzger and Rieger.⁹⁴ They focus primarily on application of KT principles to games that yield mixed strategies, but mixed strategies are conceptually problematic for our approach. They do show an example where pure strategy NE results would create a reference point against which one player might evaluate possible outcomes. However, we interpreted this was shown to contrast the effects of the pure NE-induced reference frame against the effects of a different reference frame induced by mixed strategy equilibria results. We did not find this helpful for explaining how pure strategy equilibria might help us proxy attacker intent in a way that would show differences in outcomes.

Type 2 (PT): Feasibility of Prospects – Single Result

What if an attacker evaluates prospects, but according to KT principles? This approach most closely resembles the way Kahneman and Tversky elicited preferences – without considering game equilibria, and simply proposing probabilistic outcomes as either losses or gains.

We first consider a notional reference point. If the attacker's reference point

(1) is organizationally driven, rather than situationally driven by an equilibrium game result as in Metzger and Rieger, and

(2) equals, for example, the maximum monetized death/injury and economic consequence from destruction of both targets,

then we can show that the attacker might prefer attacking both the chemical facility AND the ferry terminal simultaneously. To illustrate, we show the game that was used as the basis for Types 1(SEU) and 2 (SEU), with numerical values of the expected utility functions. This game is applicable to PT assumptions insofar as the form of the expected utility functions does not change:

	Optimal Investment		Suboptimal Investment	
Attack A	\$45,472,329.67	\$1,789,527,670.33	\$7,140,625	\$1,827,859,375
Attack B	\$90,944,659.34	\$1,744,055,340.66	\$229,500,000	\$1,605,500,000
Attack A+B	\$9,029,443.77	\$1,825,970,556.23	\$3,578,125	\$1,831,421,875
Refrain		\$1,835,000,000		\$1,835,000,000

Figure 17. Results of game – SEU or PT, complete information

Next, we show the attacker's results as combinations of probability and utility, rather than the final numerical values:

	Optimal Investment		Suboptimal Investment	
Attack A	.049*\$914M	\$1,789,527,670.33	.088*\$914M	\$1,827,859,375
Attack B	.099*\$914M	\$1,744,055,340.66	.25*\$918M	\$1,605,500,000
Attack A+B	.005*\$1,832M	\$1,825,970,556.23	.002*\$1,832M	\$1,831,421,875
Refrain		\$1,835,000,000		\$1,835,000,000

Figure 18. Results of game – SEU or PT, complete information – attacker results expanded

We can now show attacker prospects in a way that allows us to estimate what an attacker might prefer by leveraging KT principles. The attacker's prospect from attacking the chemical facility is shown as a combination of probabilities and utilities (outcomes):

$$U_e T_A^{post} = .049(\$914M) + .008(\$914M) = .057(\$921M)$$

Equation 6. Attacker prospect from attacking the chemical facility, post-deterrence, SEU or PT, complete information

Notice that we change the final utility to a negative number because it reflects a loss relative to the attacker's reference point. For the remaining three prospects, we have:

$$U_e T|_B^{post} = .099(\$918M) + .25(\$918M) = .349(-\$917M)$$

Equation 7. Attacker prospect from attacking the ferry terminal, post-deterrence, SEU or PT, complete information

$$U_e T|_{AB}^{post} = .005(-\$1,832M) + .002(-\$1,832M) = .007(-\$3M)$$

Equation 8. Attacker prospect from attacking both targets simultaneously, post-deterrence, SEU or PT, complete information

$$U_e T|_o^{post} = (-\$1,835M)$$

Equation 9. Attacker prospect from refraining, post-deterrence, SEU or PT, complete information

We now compare these prospects using KT principles. First, because of the certainty effect for losses, or the reflection effect, we can rule out the certain loss of refraining from attack.

Next, we compare the prospect of attacking the ferry terminal to the prospect of attacking the chemical plant. The attacker would probably prefer attacking the ferry terminal to the chemical facility, because the latter has a very small probability of a greater loss, as opposed to the former which has a small probability of a smaller loss. We estimate this preference because KT preference distributions showed that when losses are considered, and both outcomes in a prospect have small probabilities, the smaller loss is often preferred (low probabilities are overweighted which increases the unattractiveness of the larger loss).

Then, we compare the prospect of attacking both targets to the prospect of attacking the ferry terminal. In this case, we predict the attacker would prefer to attack both simultaneously. This is because there is a small probability of a much smaller loss. However, notice that this is different than what the intent ratio, pure NE, and prospect evaluation approaches (assuming SEU) predicted for the attacker COA. These approaches predicted the attacker should attack the ferry terminal. Under SEU, they would be maximizing their expected utility. However, under PT, they are evaluating prospects against a reference point. This example illustrates how different utility theories predict different results.

The resulting deterrence portfolio, which reflects the defender's equilibrium COA of optimal investment, looks like:

$$\begin{aligned}
 R|_{\$opt,AB}^{post} &= \$9,029,442.77 \\
 \Delta(R|_{\$opt}) &= -\$855,971.20 \\
 ROI_{\$opt}^A &= -0.16, ROI_{\$opt}^B = -0.08
 \end{aligned}$$

Figure 19. Deterrence portfolio of optimal investment, Type 2 (PT)

We note a difference between this result and that of Type 2(SEU). In this case, the change in unconditional risk (given defender optimal investment) is *negative*. This means that unconditional risk from an attack on both targets has actually *increased* given our optimal investment. The unconditional post -deterrence risk is much less than it was under SEU assumptions when attacking the ferry terminal was most attractive, but the *change* in unconditional risk from pre-deterrence to post-deterrence is reversed: it increases rather than decreases.

Type 2a (PT): Feasibility of Prospects – Prospect Intent Ratios

If we assume attacker intent for each COA is proxied by the ratio of the numerical value of that COA's prospect, to the aggregate value of all prospects in the game, this is no different from SEU assumptions. Therefore we do not see any value in exploring this option, absent evidence that terrorists make decisions this way under PT assumptions. Alternatively, if one developed defensible utility functions that incorporate KT principles directly, we could compare the new prospects to yield intent ratios. However, the same concerns with defensible utility functions discussed earlier apply here as well.

Overall Findings – PT, Complete Information

We only found one approach that yielded information to create a deterrence portfolio, and one without meaningful deterrence quantification values at that. When an attacker evaluates prospects, in this case they would be more likely to choose attacking both targets simultaneously, as opposed to attacking the ferry terminal under SEU assumptions. This was given a specific reference point, maximum value of both targets. This result yields lower unconditional risk for the defender in deterrence portfolios than under SEU conditions. This comparison between SEU and PT results, assuming complete information, is encouraging because if PT represents reality, we may face less risk than originally thought.

Importantly, we predict that attacking both targets simultaneously is the attacker's most desirable prospect based on similarities between probabilities/outcomes that influenced KT preference distributions, and probabilities/outcomes in our deterrence game prospects.

However, we reiterate that Kahneman and Tversky did not elicit preference distributions under game theoretic circumstances.

Incomplete Information

Since we only established one approach for evaluating games of complete information under PT assumptions, we revisit that approach but instead assume incomplete information.

Type 4 (PT): Prospects – Single Result – Incomplete Information

We revisit the deterrence portfolio for SEU assumptions, incomplete information, for a pessimistic attacker:

$$\left[\begin{array}{l} \left(\begin{array}{ll} E_{\$opt|A} = -74.24\% & E_{\$opt|B} = 12.88\% \\ E_{\$opt|AB} = 96.52\% & E_{\$opt|0} = n/a \end{array} \right) \\ \overline{R|_{\$opt,k}^{post}} = \$16,990,866.67 \\ \Delta \left(\overline{R|_{\$opt}} \right) = \$26,027,135.20 \\ ROI_{\$opt}^A = 0.91, ROI_{\$opt}^B = 0.45 \end{array} \right]$$

If the attacker evaluates aggregate prospects, the “pessimistic attacker’s game” is shown, with numerical values of the expected utility functions, as follows:

	Optimal Investment		Suboptimal Investment	
Attack A	\$4,570,000	\$1,830,430,000	\$28,562,500	\$1,806,437,500
Attack B	\$4,590,000	\$1,830,410,000	\$57,375,000	\$1,777,625,000
Attack A+B	\$45,800	\$1,834,954,200	\$3,578,125	\$1,831,421,875
Refrain		\$1,835,000,000		\$1,835,000,000

Figure 20. Results of “attacker’s game” – SEU or PT, incomplete information, PESSIMISTIC attacker

Next, we show the attacker results:

$$U_e T_A^{post} = .005(\$914M) + .031(\$914M) = .036(-\$921M)$$

Equation 10. Attacker prospect from attacking the chemical facility, post-deterrence, SEU or PT, incomplete information, PESSIMISTIC ATTACKER

$$U_e T|_B^{post} = .005(\$918M) + .063(\$918M) = .068(-\$917M)$$

Equation 11. Attacker prospect from attacking the ferry terminal, post-deterrence, SEU or PT, incomplete information, PESSIMISTIC ATTACKER

$$U_e T|_{AB}^{post} = .000025(\$1,832M) + .002(\$1,832M) = .002(-\$3M)$$

Equation 12. Attacker prospect from attacking both targets simultaneously, post-deterrence, SEU or PT, incomplete information, PESSIMISTIC ATTACKER

$$U_e T|_o^{post} = (-\$1,835M)$$

Equation 13. Attacker prospect from refraining, post-deterrence, SEU or PT, incomplete information, PESSIMISTIC ATTACKER

Just as we did under complete information circumstances, we now compare these prospects using KT principles. We can rule out the certain loss of refraining from attack again.

Next, we compare the prospect of attacking the ferry terminal to the prospect of attacking the chemical facility. Again, because these are losses, and both outcomes have small probabilities, the smaller loss is often preferred, so we predict the attacker would prefer attacking the ferry terminal over attacking the chemical facility.

Then we compare the prospect of attacking both targets to the prospect of attacking the ferry terminal. In this case, once again we predict the attacker would prefer attacking both. This is the same result as predicted when evaluating prospects under complete information circumstances. When assuming PT, the bottom line for comparison of results under different information availability circumstances seems to be that the attacker's predicted preference does NOT change.

Given our prediction that a pessimistic attacker will prefer to attack both targets, but we will invest optimally as the pure NE solution, the deterrence portfolio is shown:

$$\left[\begin{array}{l} R|_{\$opt,AB}^{post} = \$9,029,442.77 \\ \Delta(R|_{\$opt}) = -\$855,971.20 \\ ROI_{\$opt}^A = -0.16, ROI_{\$opt}^B = -0.08 \end{array} \right]$$

Figure 21. Deterrence portfolio of optimal investment, Type 3 (PT), PESSIMISTIC ATTACKER

What happens when we assume a neutral attacker?

	Optimal Investment		Suboptimal Investment	
Attack A	\$22,850,000	\$1,812,150,000	\$57,125,000	\$1,777,875,000
Attack B	\$22,950,000	\$1,812,050,000	\$144,750,000	\$1,720,250,000
Attack A+B	\$1,145,000	\$1,833,855,000	\$14,312,500	\$1,820,687,500
Refrain		\$1,835,000,000		\$1,835,000,000

Figure 22. Results of “attacker’s game” – SEU or PT, incomplete information, NEUTRAL attacker

The attacker prospects are:

$$U_e T|_A^{post} = .025(\$914M) + .063(\$914M) = .088(-\$921M)$$

Equation 14. Attacker prospect from attacking the chemical facility, post-deterrence, SEU or PT, incomplete information, NEUTRAL ATTACKER

$$U_e T|_B^{post} = .025(\$918M) + .125(\$918M) = .15(-\$917M)$$

Equation 15. Attacker prospect from attacking the ferry terminal, post-deterrence, SEU or PT, incomplete information, NEUTRAL ATTACKER

$$U_e T|_{AB}^{post} = .000625(\$1,832M) + .008(\$1,832M) = .009(-\$3M)$$

Equation 16. Attacker prospect from attacking both targets simultaneously, post-deterrence, SEU or PT, incomplete information, NEUTRAL ATTACKER

$$U_e T|_o^{post} = (-\$1,835M)$$

Equation 17. Attacker prospect from refraining, post-deterrence, SEU or PT, incomplete information, NEUTRAL ATTACKER

Comparing the two targets, again we believe the attacker would prefer the ferry terminal. Comparing attacking both simultaneously to the ferry terminal, we again believe the attacker would prefer to attack both. As expected, the aggregate probability of the outcome from attacking both targets is greater for a neutral attacker than for a pessimistic attacker. However, it is less than the aggregate probability of achieving the outcome from attacking both targets under complete information circumstances. The deterrence portfolio of optimal investment in this case is the same as Figure 22.

What if the attacker is optimistic?

	Optimal Investment		Suboptimal Investment	
Attack A	\$45,700,000	\$1,789,300,000	\$85,687,500	\$1,749,312,500
Attack B	\$45,900,000	\$1,789,100,000	\$172,125,000	\$1,662,875,000
Attack A+B	\$4,580,000	\$1,830,420,000	\$32,203,125	\$1,802,796,875
Refrain		\$1,835,000,000		\$1,835,000,000

Figure 23. Results of “attacker’s game” – SEU or PT, incomplete information, OPTIMISTIC attacker

The attacker prospects are:

$$U_e T|_A^{post} = .05(\$914M) + .094(\$914M) = .144(-\$921M)$$

Equation 18. Attacker prospect from attacking the chemical facility, post-deterrence, SEU or PT, incomplete information, OPTIMISTIC ATTACKER

$$U_e T|_B^{post} = .05(\$918M) + .188(\$918M) = .238(-\$917M)$$

Equation 19. Attacker prospect from attacking the ferry terminal, post-deterrence, SEU or PT, incomplete information, OPTIMISTIC ATTACKER

$$U_e T|_{AB}^{post} = 00.3(\$1,832M) + .018(\$1,832M) = .021(-\$3M)$$

Equation 20. Attacker prospect from attacking both targets simultaneously, post-deterrence, SEU or PT, incomplete information, OPTIMISTIC ATTACKER

$$U_e T|_o^{post} = (-\$1,835M)$$

Equation 21. Attacker prospect from refraining, post-deterrence, SEU or PT, incomplete information, OPTIMISTIC ATTACKER

Again we predict the attacker will prefer the ferry terminal to the chemical facility and will ultimately prefer to attack both simultaneously. The aggregate probability of the outcomes in the preferred prospect is predictably higher for an optimistic attacker than for a neutral attacker. And it is now greater than that from attacking both under complete information conditions. The deterrence portfolio is the same as it is for neutral and pessimistic attackers. However, we are not convinced we should ensure complete information if we assume PT conditions apply.

Overall Findings – PT, Incomplete Information

Just as with complete information, we found that an attacker would probably prefer to attack both targets simultaneously over all other COAs, *given our specified reference point*. This was consistent across all OOBs.

About the Authors

Eric F. Taquechel is a U.S. Coast Guard officer with experience in shipboard operations, port operations, critical infrastructure risk analysis, contingency planning/force readiness, operations analysis, and planning, programming, budgeting, and execution process support. He has authored various publications including “Layered Defense: Modeling Terrorist Transfer Threat Networks and Optimizing Network Risk Reduction,” in *IEEE Network Magazine*; “How to Quantify Deterrence and Reduce Critical Infrastructure Risk,” in *Homeland Security Affairs Journal*; “Options and Challenges of a Resilience-Based, Network-Focused Port Security Grant Program,” in the *Journal of Homeland Security and Emergency Management*; and most recently “Measuring the Deterrence Value of Securing Maritime Supply Chains against WMD Transfer and Measuring Subsequent Risk Reduction,” in *Homeland Security Affairs Journal*. LCDR Taquechel earned a master’s degree in Security Studies from the Naval Postgraduate School and prior to that earned his undergraduate degree at the U.S. Coast Guard Academy, and is currently a prospective MPA candidate at Old Dominion University. LCDR Taquechel may be contacted at nixhex3092@yahoo.com.

Ted G. Lewis is a retired professor of computer science and former executive director of the Center for Homeland Defense and Security at the Naval Postgraduate School. He spent forty years in academic, industrial, and advisory capacities, ranging from academic appointments at the University of Missouri-Rolla, University of Louisiana, and Oregon State University, to senior vice president of Eastman Kodak Company, to CEO and president of DaimlerChrysler Research and Technology, North America. Dr. Lewis has published over thirty books and 100 research papers. He is the author of *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation* (2006, second edition 2014), *Network Science: Theory and Applications* (2009), *Bak’s Sand Pile* (2011), and *Book of Extremes* (2014). He received his Ph.D. in computer science from Washington State University. Dr. Lewis may be contacted at tedglewis@redshift.com.

Acknowledgements

The authors wish to thank the anonymous referees whose feedback greatly improved the paper.

Disclaimer

The original opinions and recommendations in this work are those of the authors and are not intended to reflect the positions or policies of any government agency.

Notes

- 1 Herbert Simon, *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* (New York: The Free Press, 1997).
- 2 Eric Taquechel and Ted Lewis, "How to Quantify Deterrence and Reduce Critical Infrastructure Risk," *Homeland Security Affairs* 8, (August 2012). Taquechel and Lewis used the term "information availability circumstances" to mean the extent to which an attacker knows what a defender is doing to defend a CIKR.
- 3 U.S. Department of Homeland Security, *DHS Risk Lexicon* (2010), <https://www.dhs.gov/xlibrary/assets/dhs-risk-lexicon-2010.pdf>, Web accessed July 1, 2015.
- 4 Ibid.
- 5 Richard N. Lebow, "The Cuban Missile Crisis: Reading the Lessons Correctly," *Political Science Quarterly* 98 (1983): 431-458.
- 6 Taquechel and Lewis, "How to Quantify Deterrence and Reduce Critical Infrastructure Risk."
- 7 Nikhil S. Dighe, Jun Zhuang, and Vicki M. Bier, "Secrecy in Defensive Allocations as a Strategy for Achieving more Cost Effective Deterrence," *International Journal of Performability Engineering* 5 (2009): 31- 43.
- 8 Andrew R. Morral and Brian A. Jackson, "Understanding the Role of Deterrence in Counterterrorism Security," RAND Occasional Paper (2009), http://www.rand.org/pubs/occasional_papers/OP281.html, Web accessed July 1, 2015.
- 9 Taquechel and Lewis, "How to Quantify Deterrence and Reduce Critical Infrastructure Risk."
- 10 Ibid.
- 11 Daniel Moran, "Strategic Insight: Deterrence and Preemption," *Strategic Insights* 1 (2002), <https://www.hsdl.org/?view&did=1428> Web accessed July 1, 2015.
- 12 Kevin Chilton and Greg Weaver, "Waging Deterrence in the Twenty-First Century," *Strategic Studies Quarterly* (2009): 31-42. <http://www.au.af.mil/au/sss/2009/Spring/chilton.pdf> Web accessed July 1, 2015.
- 13 Mikhael Shor, "Sequential Game," Dictionary of Game Theory Terms, Game Theory .net, <http://www.gametheory.net/dictionary/SequentialGame.html> Web accessed May 22, 2014.
- 14 Gerald Brown et al., "Defending Critical Infrastructure," *Interfaces* 36 (2006): 530-544.
- 15 James Pita et al., "A Robust Approach to Addressing Human Adversaries in Security Games," Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, Vol 3, 1297-1298, 2012. http://teamcore.usc.edu/papers/2012/MATCH_ECAI_final2.pdf. Web accessed July 1, 2015.
- 16 Mikhael Shor, "Simultaneous Game," Dictionary of Game Theory Terms, Game Theory .net, <http://www.gametheory.net/dictionary/SimultaneousGame.html> Web accessed July 1, 2015.
- 17 Kjell Hausken, Vicki M. Bier, and Jun Zhuang, "Defending Against Terrorism, Natural Disaster, and All Hazards," *Game Theoretic Risk Analysis of Security Threats*, Bier, Vicki and Azaiez, M. Naceur, eds, Vol. 128 of International Series in Operations Research and Management Science, 2009: 65-97.
- 18 Zhengyu Yin et al. "Stackelberg vs. Nash in Security Games: Interchangeability, Equivalence, and Uniqueness," *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems* 1 (2010), <http://teamcore.usc.edu/papers/2010/AAMAS10-OBS.pdf>. Web accessed July 1, 2015.
- 19 Louis Philips, *The Economics of Imperfect Information* (Cambridge, UK: Cambridge University Press, 1988).
- 20 Erik Jenelius, Jonas Westin, and Åke J. Holmgren, "Critical Infrastructure Protection under Imperfect Attacker Perception," *International Journal of Critical Infrastructure Protection* 3 (2010): 16-26.

- 21** Jenelius et al. (op. cit.) do not explicitly define “observation error” in their work, but give examples such as “uncertainty about the level of vulnerability of one of the targets”; inability to perceive “precise gains associated with attacking a target...due to factors such as undisclosed information, surveillance, complex system structure and geographical separation between the antagonist and the target”; and inability to “perfectly observe the utilities associated with attacking elements of an infrastructure system.” Also, they use the terms “imperfect attack perception” and “imperfect observations,” and occasionally use the term “incomplete information.” Given these examples, we infer that Jenelius et al. are referring to the traditional concept of incomplete information when they use the terms “observation error” or “imperfect attack perception,” although introducing the word “imperfect” may blur the lines created by the textbook differentiation of “imperfect information” and “incomplete information.”
- 22** M. Naceur Azaiez, “A Bayesian Model for a Game of Information in Optimal Attack/Defense Strategies,” in *Game Theoretic Risk Analysis of Security Threats*, Vicki M. Bier and M. Naceur Azaiez, eds, (New York: Springer, 2009), 99-123.
- 23** Edi Karni, “Savage’s Subjective Expected Utility Model,” 2005, <http://www.econ2.jhu.edu/people/Karni/savageseu.pdf> Web accessed July 1, 2015.
- 24** Morral and Jackson, “Understanding the Role of Deterrence in Counterterrorism Security.”
- 25** Karni, “Savage’s Subjective Expected Utility Model.”
- 26** Theodore G. Lewis, *Network Science: Theory and Application* (Hoboken: Wiley Interscience, 2009).
- 27** W.I. Al-Mannai and T. Lewis, “A General Defender-Attacker Risk Model for Networks,” *The Journal of Risk Finance* 9, no.3 (2008):244 - 261.
- 28** Vicki M. Bier et al., “Optimal Resource Allocation for Defense of Targets Based on Differing Measures of Attractiveness,” *Risk Analysis* 28 (2008): 763-770.
- 29** See Theodore G. Lewis, *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation* (Hoboken: Wiley Interscience, 2006) and Theodore G. Lewis, *Network Science: Theory and Application* (Hoboken: Wiley Interscience, 2009).
- 30** Lagrange multipliers are a calculus technique to find the “saddle point” or critical point of some performance metric subject to a constraint, by converting a constrained objective function into an unconstrained function and solving for the optimal distribution of some quantity that maximizes that metric. In the case of investment to protect critical infrastructure, if the metric is expected utility, the constraints may include available budget to invest, and the quantities to be optimally distributed are the investments at each infrastructure in the “game” being played. However, saddle points are not guaranteed to be the maxima of the function. Thus, we used a bordered Hessian matrix to prove the optimal investments derived using Lagrange multipliers truly yield maximum defender expected utility, rather than minimum. Alternatively, we might brute force substitute the optimal investment values back into the defender expected utility equation.
- 31** Gregory Levitin, “Optimizing Defense Strategies for Complex Multi-State Systems,” *Game Theoretic Risk Analysis of Security Threats*, Vicki Bier and Azaiez, M. Naceur, eds, Vol. 128 of International Series in Operations Research and Management Science, 2009: 33-64.
- 32** Richard N. Lebow and Janet G. Stein, “Rational Deterrence Theory: I think, Therefore I deter,” *World Politics* 41 (1989): 208-224.
- 33** Karni, “Savage’s Subjective Expected Utility Model.”
- 34** Paul J. H. Schoemaker, “The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations,” *Journal of Economic Literature* 20 (1982): 529-563.
- 35** See Mikhael Shor, “Risk and Certainty Equivalent Applet,” Dictionary of Game Theory Terms, Game Theory .net, <http://www.gametheory.net/mike/applets/Risk/> for a good explanation of expected utility and certainty equivalents. Web accessed July 1, 2015.
- 36** Schoemaker, “The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations.”
- 37** Kahneman and Tversky, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica* 47, no.2 (Mar., 1979): 263-292.

- 38** Tversky and Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, New Series, 211, no.4481 (Jan. 30, 1981): 453-458.
- 39** Ibid.
- 40** Ibid.
- 41** Ibid.
- 42** Ibid.
- 43** We originally considered prospects with more than two outcomes in this approach, but we could not find literature supporting predictions when more than two outcomes in a prospect are considered. Therefore, we limited the number of outcomes in an attacker prospect to two: the only defender COAs are to invest optimally or to invest suboptimally.
- 44** Kahneman and Tversky, "Prospect Theory: An Analysis of Decision under Risk."
- 45** Elinor Ostrom, "A Behavioral Approach to the Rational Choice Theory of Collective Action," *American Political Science Review* 92 (1998): 1-22.
- 46** Gary Schaub, Jr, "Deterrence, Compellence, and Prospect Theory," *Political Psychology* 25 (2004): 389-411.
- 47** Rong Yang, "Human Adversaries in Security Games: Integrating Models of Bounded Rationality and Fast Algorithms," PhD Dissertation, University of Southern California, April 2014. http://teamcore.usc.edu/yangrong/thesis_Yang.pdf, Web accessed July 1, 2015.
- 48** Bo An et al., "A Deployed Quantal Response Based Patrol Planning System for the U.S. Coast Guard," *INFORMS* 43 (2013), <http://pubsonline.informs.org/doi/abs/10.1287/inte.2013.0700>, Web accessed July 1, 2015.
- 49** Ibid.
- 50** Jenelius et al., "Critical Infrastructure Protection under Imperfect Attacker Perception."
- 51** An et al., "A Deployed Quantal Response Based Patrol Planning System for the U.S. Coast Guard."
- 52** Pita et al., "A Robust Approach to Addressing Human Adversaries in Security Games."
- 53** Taquechel and Lewis, "How to Quantify Deterrence and Reduce Critical Infrastructure Risk."
- 54** The authors are grateful to Dr. David Alderson, Naval Postgraduate School for pointing this out.
- 55** See Taquechel and Lewis, "How to Quantify Deterrence and Reduce Critical Infrastructure Risk."
- 56** Details include expected utility based on a specific investment, the amount of the investment, etc.
- 57** M. Naceur Azaiez, "A Bayesian Model for a Game of Information in Optimal Attack/Defense Strategies," in *Game Theoretic Risk Analysis of Security Threats*, Vicki M. Bier and M. Naceur Azaiez, eds, (New York: Springer, 2009), 99-123.
- 58** OOB reflects attacker estimates of the *effects* of defender deterrence investment, such as resulting target vulnerability, rather than estimates of the investments themselves.
- 59** For incomplete information, we use a neutral attacker's estimate of the vulnerability that would result from a defender's optimal investment, as a baseline from which we modify optimistic and pessimistic attacker estimates. For simplicity, our neutral attacker's estimate of the effects of defender optimal investment will be the minimum theoretically possible vulnerability of each CIKR target in the game, assuming no cognitive biases. This minimized vulnerability is the elimination fraction EF_i in Equation 1. In this case, we have set the elimination fraction to 5%. Thus, a pessimistic attacker's estimate of the results of a defender's optimal investment will put the elimination fraction at less than 5% , and an optimistic attacker's estimate of the same will put the elimination fraction at more than 5%. Future work can modify this "baseline" estimate of the effects of optimal defender investment, to gauge sensitivity of results.
- 60** This requires us to assume that the pre-deterrence vulnerability of all targets in the game is greater than 10%.

61 Non-cooperative games have a Nash Equilibrium (NE) solution, which represents the “optimal” solution concept, consisting of one COA for each player, where each player could choose from multiple COAs at the start of a game. A pure NE means that in theory each player should prefer their equilibrium COA with 100% intent. If *all* players chose their respective equilibrium COAs during one round of the game, the NE solution means each player gets their best possible or “optimal” expected utility *given all other players are simultaneously trying to maximize their own expected utility*. Each player’s expected utility from such a solution may be less than what each player could achieve if their opponents *were not also trying to maximize their own utilities*; hence our notation of “optimal” in quotation marks in this game theoretical context. This concept of “optimal” also differs from the mathematically optimal defender investment COA, which is an *input* to the deterrence game; the “optimal” NE solution is an *output* of a game.

62 Mixed strategies reflect what one player should do to make their opponent indifferent between choices; in a game with two COAs, the first player’s preferences in a mixed strategy reflect a probabilistic distribution where they should execute one COA $x\%$ of the time, and the other $1-x\%$ of the time. Technically a pure strategy NE is one kind of mixed strategy NE.

63 Eric Rasmusen, “Mixed and Continuous Strategies,” www.rasmusen.org/GI/chapters/chap03_mixed.pdf.

64 Finding a pure or mixed strategy NE may first require elimination of dominated strategies if any exist. A strategy is dominated if, regardless of what any other players do, the strategy earns a player a smaller payoff than some other strategy. See <http://www.gametheory.net/dictionary/DominatedStrategy.html> Web accessed July 1, 2015.

65 Otherwise, the attacker could work backwards to calculate defender deterrence budget, optimal investments, and resulting attacker expected utility, defeating the purpose of obfuscating those details in the first place.

66 Otherwise, the attacker might be able to “reverse engineer” and calculate the optimal investment amounts.

67 See Eric Taquechel, “Options and Challenges of a Resilience-Based, Network-Focused Port Security Grant Program,” *Journal of Homeland Security and Emergency Management* 10 (2013), 521–554. Taquechel shows how changing investments in port resilience influences the “organic failure susceptibility” of each infrastructure node and thus changes supply chain network resilience, thereby creating a deterrent effect against prospective supply chain attackers. However, Taquechel does not explore how changing network attributes (e.g. node degree) and thus changing “inherited failure susceptibility” influences network resilience and deterrence. Future work may explore how “re-wiring” the network or changing its topology or other attributes influences deterrence and network resilience.

68 This is an amount that would place averaged attacker expected utility (across all four attacker COAs) halfway between what it would be if the defender invested optimally and what it would be if the defender refrained from deterrence investment.

69 Amos Tversky and Daniel Kahneman, “Loss Aversion in Riskless Choice: A Reference-Dependent Model,” *The Quarterly Journal of Economics* 106, no.4 (Nov., 1991): 1039–1061.

70 Abigail Linnington, “Unconventional Warfare as A Strategic Foreign Policy Tool: The Clinton Administration in Iraq and Afghanistan,” Master’s thesis, Tufts University, MA, 2004, <http://dl.tufts.edu/bookreader/tufts:UA015.012.DO.00032#page/5/mode/2up> Web accessed May 22, 2014.

71 Ibid.

72 Amos Tversky and Daniel Kahneman, “Rational Choice and the Framing of Decisions.” *The Journal of Business* 59, No. 4, Part 2: The Behavioral Foundations of Economic Theory (Oct., 1986): S251-S278.

73 Kahneman and Tversky. “Prospect Theory: An Analysis of Decision under Risk.”

74 Ibid.

75 Tversky and Kahneman, “Loss Aversion in Riskless Choice: A Reference-Dependent Model.”

76 Daniel Kahneman and Amos Tversky, “Choices, Values and Frames,” *American Psychologist*, Vol 39, No 4, 341-350, April 1984. There is an example elicitation on preferences for different vaccination programs, after which the authors claim reference points changed between two different formulations of the same elicitation.

- 77** T. K. Das and Bing-Sheng Teng, "Strategic Risk Behavior and its Temporalities :Between Risk Propensity and Decision Context," *Journal of Management Studies* 38 (2001): 515-534.
- 78** Kahneman and Tversky, "Prospect Theory: An Analysis of Decision under Risk."
- 79** Daniel Kahneman and Dan Lovallo, "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking," *Management Science* 39 (1993): 17-31.
- 80** Brady Downs, "The Maritime Security Risk Analysis Model," *Coast Guard Proceedings* (2007): 36-39.
- 81** Taquechel and Lewis, "How to Quantify Deterrence and Reduce Critical Infrastructure Risk."
- 82** Chilton and Weaver, "Waging Deterrence in the Twenty-First Century."
- 83** Berejikian, "A Cognitive Theory of Deterrence, " *Journal of Peace Research* 39, no.2 (March 2002): 165-183.
- 84** National Infrastructure Protection Plan. 2013, <https://www.dhs.gov/sites/default/files/publications/National-Infrastructure-Protection-Plan-2013-508.pdf> Web accessed May 9, 2016.
- 85** This is known as a value of a statistical life. See for example W. Kip Viscusi and Joseph E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," National Bureau of Economic Research Working Paper 9487, 2003, <http://www.nber.org/papers/w9487>, web accessed July 1, 2015.
- 86** Because there is no equilibrium solution to the pre-deterrence "game", we have no baseline to which we can compare the risk determined from the deterrence game equilibrium. Therefore, we use averages of pre-deterrence risk and averages of post-deterrence risk for consistency in our comparisons.
- 87** Importantly, use of intent ratios does not reflect our belief that the attacker will not attack, or will not choose a "pure strategy." Instead, it is our proxy for the possibility that the attacker will choose a non-equilibrium COA.
- 88** It may be a coincidence that attacking the ferry terminal, given defender optimal investment, yields the best result for the attacker when we evaluate the game results based on a pure NE.
- 89** The notion of fostering attacker indifference between their COAs, outside of a mixed strategy result, is explored by Major in his concept of "equilibrium expected loss." This is interesting if we want to make an attacker indifferent between targets of equally low value, but in our approach the noted decrease in attractiveness differential may represent an emergence of an "attractiveness equilibrium" brought on by changes in attacker OOB, rather than a "loss equilibrium" that reflects only target value and does not consider nuances of attacker estimates under conditions of incomplete information. For a discussion of equilibrium expected loss, see: John A. Major, "Advanced Techniques for Modeling Terrorism Risk," *Journal of Risk Finance* 4 (2002): 15-24.
- 90** Tversky and Kahneman, "The Framing of Decisions and the Psychology of Choice."
- 91** Vilhelm Verendel, "A Prospect Theory Approach to Security," Chalmers University of Technology/Goteborg University, Technical Report No. 08-20, 2008. Web accessed July 1, 2015 at <http://www.cse.chalmers.se/~vive/prospectTR.pdf> .
- 92** Tversky and Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty* 5 (1992): 297.
- 93** Kahneman and Tversky proposed a specific equation for the overall value of a prospect wherein the sum of probabilities of all outcomes = 1, or the utility of one outcome is a gain while the utility of the other outcome is a loss. Alternatively, they proposed a different equation for prospects wherein outcomes are all either strictly positive (Gains) or all strictly negative (Losses). See Kahneman and Tversky, "Prospect Theory: An Analysis of Decision under Risk", p.276. However, neither of the aforementioned sets of prerequisites to use these specific equations applies in our work.
- 94** Lars P. Metzger and Marc Oliver Rieger, "Equilibria in Games with Prospect Theory Preferences," National Centre of Competence in Research Financial Valuation and Risk Management Working Paper No. 598, 2009, Web accessed July 1, 2015 at: http://www.nccr-finrisk.uzh.ch/media/pdf/wp/WP598_A1.pdf.

Copyright © 2016 by the author(s). Homeland Security Affairs is an academic journal available free of charge to individuals and institutions. Because the purpose of this publication is the widest possible dissemination of knowledge, copies of this journal and the articles contained herein may be printed or downloaded and redistributed for personal, research or educational purposes free of charge and without permission. Any commercial use of Homeland Security Affairs or the articles published herein is expressly prohibited without the written consent of the copyright holder. The copyright of all articles published in Homeland Security Affairs rests with the author(s) of the article. Homeland Security Affairs is the online journal of the Naval Postgraduate School Center for Homeland Defense and Security (CHDS).