Applied Adversarial Reasoning and Risk Modeling: Papers from the 2011 AAAI Workshop (WS-11-06)

Toward Addressing Human Behavior with Observational Uncertainty in Security Games

James Pita and **Rong Yang** and **Milind Tambe** and **Richard John**
University of Southern California, Los Angeles, CA 90089

Abstract

Stackelberg games have recently gained significant attention for resource allocation decisions in security settings. One critical assumption of traditional Stackelberg models is that all players are perfectly rational and that the followers perfectly observe the leader's strategy. However, in real-world security settings, security agencies must deal with human adversaries who may not always follow the utility maximizing rational strategy. Accounting for these likely deviations is important since they may adversely affect the leader's (security agency's) utility. In fact, a number of behavioral game-theoretic models have begun to emerge for these domains. Two such models in particular are COBRA (Combined Observability and Bounded Rationality Assumption) and BRQR (Best Response to Quantal Response), which have both been shown to outperform game-theoretic optimal models against human adversaries within a security setting based on Los Angeles International Airport (LAX). Under perfect observation conditions, BRQR has been shown to be the leading contender for addressing human adversaries. In this work we explore these models under limited observation conditions. Due to human anchoring biases, BRQR's performance may suffer under limited observation conditions. An anchoring bias is when, given no information about the occurrence of a discrete set of events, humans will tend to assign an equal weight to the occurrence of each event (a uniform distribution). This study makes three main contributions: (i) we incorporate an anchoring bias into BRQR to improve performance under limited observation; (ii) we explore finding appropriate parameter settings for BRQR under limited observation; (iii) we compare BRQR's performance versus COBRA under limited observation conditions.

Introduction

Game theory has become a useful tool for reasoning about security settings and assigning limited security resources (Jain et al. 2010). A benefit of game theory is that it allows security agencies to reason about the interactions between players (i.e., the security agency and their adversaries) and decide their optimal security policy. In fact, a number of real-world systems use game-theoretic approaches at the heart of their decision making processes including ARMOR, IRIS, and GUARDS which provide assistance for resource allocation decisions at Los Angeles International Airport, for the Federal Air Marshals Service, and for the Transportation Security Administration respectively (Jain et al. 2010; Pita et al. 2011). One possible complication of this approach is that these systems use the standard game theoretic assumptions. Specifically, the existing algorithms for these systems find the optimal security policy under the strict assumption that all players are perfectly rational and that the followers (i.e., adversaries) perfectly observe the leader's policy. In the real-world such strong assumptions rarely hold when dealing with human players.

It is well known that standard game-theoretic assumptions of perfect rationality are not ideal for predicting the behavior of humans in strategic decision problems (Camerer 2003). A large variety of alternative models have been proposed in behavioral game theory and cognitive psychology to attempt to capture some of the deviations of human decisions from perfect rationality. Recently, attempts have been made to try and integrate some of these more realistic models of human behavior into the computational analysis of security resource allocation problems (Pita et al. 2010; Yang et al. 2011). The goal of this type of research is to improve the decisions of resource allocation tools such as ARMOR to aid security agencies in dealing with human opponents.

While there are a number of potential challenges to address in dealing with human decision makers, our work will focus on addressing human deviation due to one of these challenges. Specifically, we will focus on human deviations due to limited observation conditions. To investigate these deviations we examine a security setting based on Los Angeles International Airport (LAX) proposed by Pita et al. (Pita et al. 2010) and later extended by Yang et al. (Yang et al. 2011). Given that the objective of this study is to address non-optimal and uncertain human responses, traditional proofs of correctness or optimality are insufficient: it is necessary to experimentally test our models.

We investigate two of the leading behavioral game-theoretic models for security resource allocation problems (Yang et al. 2011) under limited observation conditions. Previous work has shown that under perfect observation conditions a strategy known as BRQR (Best Response to Quantal Response) performs the best experimentally against humans. However, under limited observation conditions, humans may have an anchoring bias which may cause them to

deviate from the utility maximizing strategy in a particular way. We extend the previous work in three ways: (i) we incorporate an anchoring bias into BRQR to improve performance under limited observation; (ii) we explore finding appropriate parameter settings for BRQR under limited observation; (iii) we compare BRQR's performance versus COBRA under limited observation conditions, showing that BRQR is still potentially the superior method for dealing with human adversaries even under limited observation conditions.

Background

Stackelberg Security Game

Stackelberg games have been shown to be a natural model for analyzing real-world security resource allocation problems (Jain et al. 2010). In a Stackelberg game, one player, the leader, commits to a strategy publicly before the remaining players, the followers, make their decision. In a security game there are two agents – the defender (security agency) and an attacker – who act as the leader and the follower in a Stackelberg game. The defender has K resources to protect a set of targets, $t_i \in T$, which have a unique reward and penalty to both the defender and attacker. Thus, some targets may be more valuable to the defender than others. Once a resource is allocated to a target it is marked as covered, otherwise it is marked as uncovered. If the attacker attacks an uncovered target he gets his reward and the defender her corresponding penalty else vice versa. The defender's goal is to maximize her reward given that the attacker will attack with knowledge of the defensive strategy the defender has chosen. For example, in an airport there may be eight terminals serving passengers, but only four bomb sniffing canine units to patrol the terminals. In this scenario, the canine units must decide on a patrolling strategy over these eight terminals first, while their adversaries are able to conduct surveillance and act taking this committed strategy into account. In such security settings the optimal Stackelberg solutions, based on a perfectly rational adversary, are typically mixed strategies (i.e., a randomized patrolling strategy).

There exist a number of algorithms and techniques for optimally solving security games given the standard Stackelberg assumptions (Jain et al. 2010; Pita et al. 2011). In this paper we use DOBSS (Paruchuri et al. 2008) as the baseline algorithm for determining the optimal security policy assuming a utility maximizing perfectly rational adversary.

Anchoring Bias

Support theory is a theory of subjective probability (Tversky and Koehler 1994) and has been used to introduce anchoring biases (See, Fox, and Rottenstreich 2006). An anchoring bias is when, given no information about the occurrence of a discrete set of events, humans will tend to assign an equal weight to the occurrence of each event (a uniform distribution). This is also referred to as giving full support to the ignorance prior (See, Fox, and Rottenstreich 2006). It has been shown through extensive experimentation that humans are particularly susceptible to giving full support to the ignorance prior before they are given any information and that, once given information, they are slow to update away from this assumption (See, Fox, and Rottenstreich 2006).

Models have been proposed to address this bias and predict what probability a human will assign to a particular event x from a set of events X based on the evaluative assessment (i.e., assessment based on events actually viewed) they have made for the occurrence of that event. We utilize a model where it is assumed the estimated probabilities are directly calculated using a simple linear model: $P(x') = \alpha(1/|X|) + (1-\alpha)P(x)$. Here, α represents the bias humans will have toward the ignorance prior. The more information a human is given to assess, the less bias they will have toward the ignorance prior (i.e., the smaller the value of α). Although this is not the only possible model for determining anchoring bias, it is ideal since the odds form model (Fox and Rottenstreich 2003) is not easily representable in an Mixed Integer Linear Program (MILP).

Behavioral Models

In order to examine the benefits of an anchoring bias we will examine two existing behavioral game-theoretic approaches. To the best of our knowledge there have only been three proposed behavioral game-theoretic approaches for dealing with human adversaries in this particular resource allocation problem known as security games (Pita et al. 2010; Yang et al. 2011; Yin et al. 2010). Other work has examined the impact of limited observation in security domains such as patrolling (Agmon et al. 2009). At this time we briefly introduce the two approaches we examine in this study[1].

COBRA

The COBRA (Combined Observability and Bounded Rationality Assumption) algorithm was developed to account for human deviations based on bounded rationality and observational uncertainty (Pita et al. 2010). Here, COBRA assumes a boundedly rational human follower is willing to choose any ϵ-optimal response strategy, i.e., the follower may choose any of the responses within ϵ-utility of the optimal utility strategy. Thus COBRA takes a robust approach and attempts to maximize the minimum utility the leader obtains for any ϵ-optimal response strategy by the follower. To account for observational uncertainty, COBRA uses the linear model explained previously. Specifically, COBRA sets a parameter $\alpha \in [0 \dots 1]$ that determines the human's bias toward the ignorance prior. This value is set based on the number of observations the human follower is expected to take before making his decision. If the human has no information (i.e., no observations) then $\alpha = 1$ and the follower is expected to be guided entirely by their belief in the ignorance prior. The value of α decreases as the number of observations the human is expected to take increases until $\alpha = 0$ when the human perfectly observes the leader strategy. The MILP for COBRA is as follows:

[1]We chose to omit the third approach as it has been shown to be outperformed by BRQR in the perfect observation condition.

$$\max_{x,q,h,a,\gamma} \gamma$$

$$\text{s.t.} \sum_{i \in X} x_i = 1 \quad (1)$$

$$\sum_{j \in Q} q_j \geq 1 \quad (2)$$

$$\sum_{j \in Q} h_j = 1 \quad (3)$$

$$0 \leq (a - \sum_{i \in X} C_{ij} * x'_i) \leq (1 - h_j)M \quad (4)$$

$$\varepsilon(1 - q_j) \leq a - \sum_{i \in X} C_{ij} * x'_i \leq \varepsilon + (1 - q_j)M \quad (5)$$

$$M(1 - q_j) + \sum_{i \in X} R_{ij} x_i \geq \gamma \quad (6)$$

$$h_j \leq q_j \quad (7)$$

$$x_i \in [0 \dots 1] \quad (8)$$

$$q_j, h_j \in \{0, 1\} \quad (9)$$

$$a \in \Re \quad (10)$$

$$x'_i = (1/|X|) * (\alpha) + (1 - \alpha) * x_i \quad (11)$$

Here, the index sets of leader and follower pure strategies are denoted by X and Q respectively. The leader's mixed strategy is denoted by x, a probability distribution over the vector of the leader's pure strategies. The value x_i is the proportion of times in which pure strategy $i \in X$ is used in the strategy. The payoff matrices of the leader and the follower are indexed by the matrices R and C respectively where R_{ij} and C_{ij} represent the reward to the leader and follower if the leader takes action i and the follower action j. The variable h_j is used to identify the optimal strategy for the follower with a value of a in the third and fourth constraints. The variable q represents all ϵ-optimal strategies for the follower; the second constraint allows for one or more strategies for the follower. The fifth constraint ensures that $q_j = 1$ for every action j such that $a - \sum_{i \in X} C_{ij} \leq \epsilon$, since in this case the middle term in the inequality is less than ϵ and the left inequality is then only satisfied if $q_j = 1$. The sixth constraint helps define the objective value against the follower, γ, which must be lower than any leader reward for all actions $q_j = 1$. Since the objective is to maximize γ, forcing γ to the minimum leader reward of all ϵ-optimal actions allows COBRA to robustly guard against the worst case scenario over all ϵ-optimal actions.

BRQR

Yang et al. (Yang et al. 2011) presented an efficient model for computing a strategy based on Quantal Response Equilibrium (QRE) for security games. QRE suggests that instead of strictly maximizing utility, individuals respond stochastically in games: the chance of selecting non-optimal strategies increases as the cost of such an error decreases. In applying the QRE model to security games Yang et al. (Yang et al. 2011) only add noise to the response function for the adversary, so the defender computes an optimal strategy assuming the attacker responds with a noisy best-response. The parameter λ represents the amount of noise in the attacker's response. Given λ and the defender's mixed strategy x, the adversary's quantal response q_i (i.e., probability of taking action i) can be written as:

$$q_i = \frac{e^{\lambda U_i^a(x)}}{\sum_{j=1}^{n} e^{\lambda U_j^a(x)}} \quad (12)$$

where, $U_i^a(x) = x_i P_i^a + (1 - x_i) R_i^a$ is the adversary's expected utility for attacking $t_i \in T$ and x is the defender's strategy.

$$q_i = \frac{e^{\lambda R_i^a} e^{-\lambda(R_i^a - P_i^a)x_i}}{\sum_{j=1}^{n} e^{\lambda R_j^a} e^{-\lambda(R_j^a - P_j^a)x_j}} \quad (13)$$

The goal is to maximize the defender's expected utility given q_i, i.e., $\sum_{i=1}^{n} q_i(x_i R_i^d + (1 - x_i) P_i^d)$. Combined with Equation (13), the problem of finding the optimal mixed strategy for the defender can be formulated as

$$\max_x \; q_i((R_i^d - P_i^d)x_i + P_i^d) \quad (14)$$

$$\text{s.t.} \sum_{i=1}^{n} x_i \leq K \quad (15)$$

$$0 \leq x_i \leq 1, \quad \forall i, j \quad (16)$$

Given that the objective function in Equation 14 is nonlinear and non-convex in its most general form, Yang et al. (Yang et al. 2011) chose to focus on methods to find local optima. To compute an approximately optimal QRE strategy they develop the Best Response to Quantal Response (BRQR) heuristic described in Algorithm 1. They first take the negative of Equation 14, converting the maximization problem to a minimization problem. In each iteration, they find the local minimum[2] using a gradient descent technique from the given starting point. If there are multiple local minima, by randomly setting the starting point in each iteration, the algorithm will reach different local minima with a nonzero probability. By increasing the iteration number, $IterN$, the probability of reaching the global minimum increases.

Algorithm 1 BRQR

```
opt_g ← −∞; {Initialize the global optimum}
for i ← 1, ..., IterN do
    x_0 ← randomly generate a feasible starting point
    (opt_l, x*) ← FindLocalMinimum(x_0)
    if opt_g > opt_l then
        opt_g ← opt_l, x_opt ← x*
    end if
end for
return opt_g, x_opt
```

Parameter Estimation: The λ-parameter in BRQR represents the amount of noise in the best-response function of the attacker. One extreme case is $\lambda = 0$, which represents uniformly random play on behalf of the attacker. The other extreme is $\lambda = \infty$, when the attacker's response is identical

[2]They use *fmincon* in Matlab to find the local minimum.

to the game-theoretic optimal response. The λ-parameter is sensitive to the game payoff structure, so tuning λ is a crucial step in applying the QRE model. Yang et al. (Yang et al. 2011) proposed using Maximum Likelihood Estimation (MLE) to fit λ using previously gathered data. Given the defender's mixed strategy x and N samples of the players' choices, the logarithm likelihood of λ is

$$\log L(\lambda \mid x) = \sum_{j=1}^{N} \log q_{\tau(j)}(\lambda)$$

where $\tau(j)$ denotes the target attacked by the player in sample j. Let N_i be the number of subjects attacking target i. Then, we have $\log L(\lambda \mid x)$=$\sum_{i=1}^{n} N_i \log q_i(\lambda)$. Combining with Equation (12),

$$\log L(\lambda \mid x) = \lambda \sum_{i=1}^{n} N_i U_i^a(x) - N \cdot \log(\sum_{i=1}^{n} e^{\lambda U_i^a(x)})$$

It has been shown that $\log L(\lambda \mid x)$ only has one local maximum (Yang et al. 2011). One potential difficulty with this approach is that data may be difficult to gather in real-world security settings. However, it may be a reasonable estimate to collect sample data from subjects in experimental settings based on these real-world domains.

Anchoring Bias for BRQR

While COBRA already accounts for an anchoring bias, we will need to modify the formulation of BRQR in order to reason about an anchoring bias. We will refer to this new formulation as BRQRA. In order to extend BRQRA to handle an anchoring bias we will need to alter the way the adversary perceives his reward. Specifically, instead of basing his decisions on the strategy x he will now base his decisions on the strategy x'. As in COBRA, x' is determined using the linear model presented previously for anchoring bias. We point out that if $\alpha = 0$ BRQRA becomes identical to BRQR and so BRQRA is applicable for all observation conditions. The adversary's quantal response can now be written as:

$$q_i^* = \frac{e^{\lambda U_i^a(x')}}{\sum_{j=1}^{n} e^{\lambda U_j^a(x')}} \tag{17}$$

where, $U_i^a(x') = (\alpha/|X| + (1-\alpha) * x_i) P_i^a + (1 - (\alpha/|X| + (1-\alpha)x_i)) R_i^a$ is the adversary's expected utility according to his anchoring bias for attacking $t_i \in T$ and x is the defender's strategy.

$$q_i^* = \frac{e^{\lambda R_i^a} e^{\frac{\lambda\alpha}{|X|}(P_i^a - R_i^a)} e^{\lambda(P_i^a - R_i^a)x_i} e^{\lambda\alpha(R_i^a - P_i^a)x_i}}{\sum_{j=1}^{n} e^{\lambda R_i^a} e^{\frac{\lambda\alpha}{|X|}(P_i^a - R_i^a)} e^{\lambda(P_i^a - R_i^a)x_j} e^{\lambda\alpha(R_i^a - P_i^a)x_j}} \tag{18}$$

The goal again is to maximize the defender's expected utility given q_i^*, i.e., $\sum_{i=1}^{n} q_i^*(x_i R_i^d + (1 - x_i) P_i^d)$. As in BRQR, the problem of finding the optimal mixed strategy for the defender can be formulated as in Equations 14$\sim$16, however, we replace q_i with q_i^*[3].

[3]As before we use the same heuristic described in Algorithm 1.

Experiments

In order to evaluate the benefit of including an anchoring bias we conduct an empirical evaluation of the presented algorithms with human subjects playing an online game. There are two goals we seek to address with our analysis. Our first goal is to see if including an anchoring bias will potentially increase the performance (expected utility) of BRQRA over BRQR. If we are able to show an increase in performance it will strengthen the case that humans have an anchoring bias under low observation and that accounting for such biases is important when addressing humans in game-theoretic settings. Second, we want to compare BRQR/BRQRA against COBRA to see if BRQR remains a superior model for addressing human behavior in such security settings, even under low observation conditions. To that end, we use the same model presented by Yang et al. (Yang et al. 2011) when they initially tested BRQR against other behavioral models in security settings. Specifically, this model is based on Los Angeles International Airport (LAX), which has eight terminals that can be targeted in an attack (Pita et al. 2008). Subjects play the role of an attacker and are able to observe the defender's mixed strategy (i.e., randomized allocation of security resources) before making a decision about which gate to attack.

Experimental Setup

Given the eight terminal scenario determined by Yang et al. (Yang et al. 2011), our experimental domain has three guards – jointly acting as the leader – guarding eight gates, and each individual human subject acts as a single adversary. Each of the eight gates has a unique reward and penalty associated with it for both the subjects as well as the guards – a non zero-sum game. In each game instance, the subject's goal is to choose the gate that will maximize his expected utility given the defender's strategy. If the subject chose a door that was guarded he would receive his penalty for that gate and the guard her reward for that gate, else vice-versa. A key difference between our experiments and those run by Yang et al. (Yang et al. 2011) is that subjects are no longer provided with the defender strategy in advance. Instead, subjects are given only five observations of the defender strategy to try and infer what that strategy is. This is because under perfect observation conditions humans should not be influenced by their anchoring bias. The game interface subjects were presented with can be seen in Figure 1.

For these experiments we use the four new reward structures presented by Yang et al. (Yang et al. 2011), which were chosen to be representative of the payoff structure space. We omit the specific details, however, these four payoff structures were chosen from a sample of 1000 randomly generated payoff structures. These 1000 payoff structures were classified based on eight features into four clusters and each of the four payoff structures represents one of the clusters.

As seen in Figure 1, the five observations the subjects received were presented to them as a set of 5 triplets. That is, a single observations is seen as [x,y,z] where the letters (i.e., x, y, and z) correspond to the gates that were guarded in that particular observation. Subjects are given an unlimited amount of time to study both the reward structure and

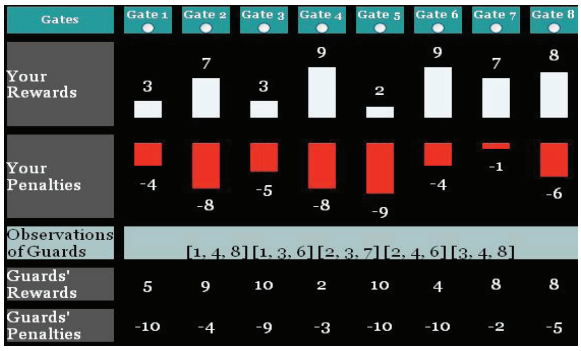


Figure 1: Game Interface

the observations they have been presented before making a decision on which gate to attack.

In order to motivate the subjects, they would earn or lose money based on whether or not they were successful in attacking a gate. Subjects started with an endowment of 8 dollars and for each reward point they earned they received an additional 10 cents. For example, examining Figure 1 if the subject chose gate 2 and was successful (i.e., there was no guard) he would receive 70 cents. Similarly, for each penalty point they lost they would lose an additional 10 cents. Regardless of how many points the subject lost they were guaranteed to be paid at least 5 dollars. There was no limit placed on the maximum amount of money a subject could earn.

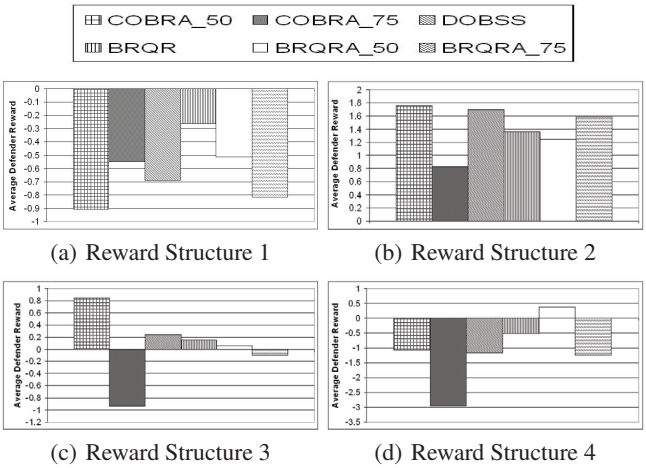
Given this experimental setup we ran two separate experiments. The first set of experiments was performed to collect initial data for the four reward structures under low observation conditions and also to compare the performance of BRQR/BRQRA and COBRA against DOBSS. We then used this initial set of data to better estimate both the α and λ parameters to see if BRQR/BRQRA would indeed outperform COBRA under low observation.

First Experiment

In the first experiment, for each payoff structure we tested the mixed strategies generated by four algorithms: DOBSS, COBRA, BRQR, and BRQRA. For both the λ-parameter and the ϵ-parameter we used the parameter settings given by Yang et al. (Yang et al. 2011). The α-parameter should be set according to the number of observations the adversary is expected to take. For the α-parameter we explored two settings with $\alpha = .50$ and $\alpha = .75$. Since we are examining a low observability condition we would expect there to be a strong anchoring bias. However, since it is costly to run too many experiments we chose to explore a half anchoring bias and three quarters anchoring bias. Of course $\alpha = 1$ is too extreme in this case since it often leads to a deterministic defender strategy (i.e., the defender guards the top 3 doors since the attacker believes the defender is choosing doors uniformly at random).

There were a total of 24 payoff structure/strategy combinations and each subject played all 24 combinations. To mitigate the order effect on subject responses, a total of 24 different orderings of the 24 combinations were generated using Latin Square design. Every ordering contained each of the 24 combinations exactly once, and each combination appeared exactly once in each of the 24 positions across all 24 orderings. The order played by each subject was drawn uniformly at random from the 24 possible orderings. In an attempt to keep each subject's strategy consistent, no feedback was given for all 24 games until the end of the experiment. A total of 33 human subjects played in this experiment.

Results: In order to evaluate the performance of each algorithm and parameter setting we computed the expected leader reward for each follower, i.e., for each choice of gate by subject. We then found the average expected reward for a given algorithm using the actual gate selections from the 33 subject trials. Figure 2 (a) shows the average expected leader reward for our first reward structure, with each data-point averaged over 33 human responses. Figures 2 (b-d) show the same for the second, third, and fourth reward structures. In all figures, the y-axis shows the average expected reward each strategy obtained and the number next to any strategy represents the α-parameter setting. For example, examining Figure 2(b) BRQRA_75 ($\alpha = .75$) obtains an average expected reward of 1.59 and DOBSS obtains an average expected reward of 1.70.



(a) Reward Structure 1 (b) Reward Structure 2

(c) Reward Structure 3 (d) Reward Structure 4

Figure 2: Average Expected Utility of Defender

Analysis: The key result of this first experiment is that DOBSS is outperformed by at least one setting of COBRA in each reward structure. While the results were not statistically significant, these trends continue to suggest the importance of addressing human adversaries in game-theoretic settings. In fact, it has been shown in 4 different reward structures under the same experimental setting (i.e., 5 observations) that COBRA statistically significantly outperforms DOBSS (Pita et al. 2010).

Given these results, we need to determine appropriate λ and α settings for both BRQR/BRQRA and COBRA. In order to make a fair comparison of the two algorithms we will need to explore a wider range of α-settings to find the best performance for each algorithm in each reward structure. In our second experiment we will also explore one possible method for appropriately determining α for BRQR.

Second Experiment

For this experiment, for each payoff structure we only tested mixed strategies generated by COBRA and BRQR/BRQRA. Once again for the ϵ-parameter we used the parameter settings given by Yang et al. (Yang et al. 2011). The ϵ-setting should not be affected by the observation condition. We used the procedure given by Yang et al. (Yang et al. 2011) based on Maximum Likelihood Estimation (MLE), which we described previously, to estimate the λ-setting based on the data from our first experiment and the current choice of α. Specifically, the λ-setting is determined as the average Maximum Likelihood Estimation (MLE) over all four reward structures and all strategies tested. However, because we are assuming the subject bases his utility on his anchoring bias (i.e., $U_i^a(x') = (\alpha/|X| + (1-\alpha) * x_i)P_i^a + (1 - (\alpha/|X|+(1-\alpha)x_i))R_i^a$) it is necessary to compute the MLE for each α-setting. This is a modification from the standard procedure which depends on $U_i^a(x)$. We present the MLE of λ for each setting of α used in this experiment in Table 1

	$\alpha = 0$	$\alpha = .10$	$\alpha = .15$	$\alpha = .20$	$\alpha = .25$
λ	.12	.16	.18	.20	.22
	$\alpha = .30$	$\alpha = .40$	$\alpha = .50$	$\alpha = .55$	$\alpha = .60$
λ	.24	.25	.25	.24	.23
	$\alpha = .65$	$\alpha = .75$	$\alpha = .85$		
λ	.21	.18	.15		

Table 1: MLE Estimates of λ

For the α-settings we tried to explore a wide range for both COBRA and BRQRA. In total for each reward structure we chose 5 new settings of α for COBRA and BRQRA along with the 3 original settings for BRQRA given the new λ-settings (i.e., BRQR or $\alpha = 0$, $\alpha = .50$, and $\alpha = .75$). This lead to a total of 13 strategies for each reward structure. For both algorithms, we chose α-settings that lead to the largest range of resulting mixed strategies to get the best idea of overall performance. We also attempted to find an optimal α-setting using an MLE method similar to that previously described for the λ-parameter. Given the defender's mixed strategy x and N samples of the subjects' choices, the logarithm likelihood of α is $\log L(\alpha \mid x) = \sum_{i=1}^{n} N_i \log q_i(\alpha)$. Here, we set $q_i(\alpha)$ as follows:

$$q_i = \begin{cases} 1 & \text{for } U_i^a(x') \geq \max_{j \in Q}(U_j^a(x')) - \epsilon \\ 1e-20 & \text{for } U_i^a(x') < \max_{j \in Q}(U_j^a(x')) - \epsilon \end{cases}$$

For the ϵ-parameter we use the setting used for COBRA in each reward structure. As with the λ-parameter we found the average MLE α-setting based on the results over all four reward structures and all strategies. The MLE of α is 0.55 for the data used from the first experiment.

Given that each reward structure had 13 new strategies there were a total of 52 payoff structure/strategy combinations. To alleviate the time it would take a subject to finish the experiment we decided to separate the reward structures into two groups. The first group was reward structures 1 and 2 while the second group was reward structures 3 and 4. Thus, each subject played 26 combinations and were assigned to either group 1 or group 2. As before, the 26 orderings were generated using Latin Square design. A total of 19 human subjects played the game for the first group and 18 human subjects played the game for the second group. No subject was allowed to play both groups.

Results: As before we computed the expected leader reward for each follower and then averaged the expected reward for a given algorithm from the 19/18 subject trials. For each reward structure we have results for 13 strategies in total (5 settings of COBRA, 7 settings of BRQRA, and BRQR). In Figure 3 we will only present results comparing the top three performing settings for both COBRA and BRQRA as well as BRQR[4]. Figure 3(a) shows the average expected leader reward for our first reward structure, with each data-point averaged over 19 human responses. Figures 3(b-d) show the same for the second, third, and fourth reward structures. In all figures, the y-axis shows the average expected reward each strategy obtained and the number next to any strategy represents the α-parameter setting. For BRQRA the λ-parameter setting is determined based on Table 1 and the α-parameter setting. For example, in reward structure 1 for BRQRA_40 we set $\lambda = .25$.

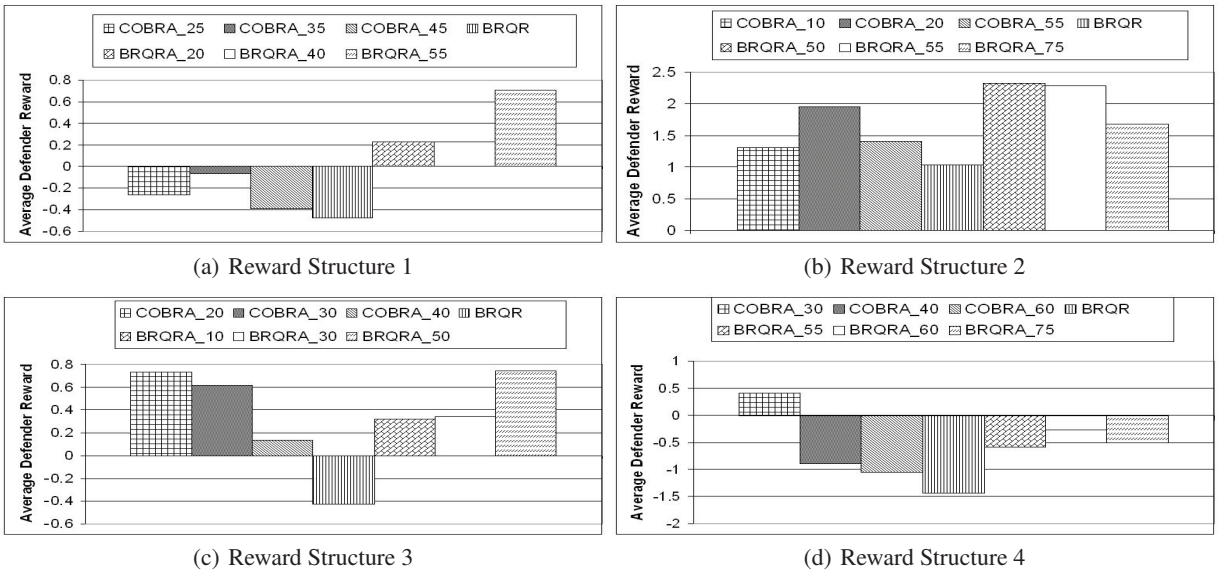
Analysis: Based on the results of this experiment there are four main conclusions: (i) incorporating an anchoring bias can help improve performance; (ii) our heuristic method for estimating α using an MLE method was a good estimation for BRQRA; (iii) appropriately estimating both the λ and α parameters enhances the performance of BRQR/BRQRA; (iv) BRQR/BRQRA stands as the leading contender for use in security domains to schedule limited security resources against human adversaries. While these results are not statistically significant, they are a first step toward appropriately addressing human adversaries in limited observation conditions. We will now more closely examine each of these conclusions.

First, in all four reward structures BRQRA obtained a higher expected utility than BRQR. This is an indication that accounting for human anchoring biases can be valuable and that extending game-theoretic models to address specific human factors, such as anchoring biases, can lead to an improvement in performance.

Second, in all four reward structures BRQRA's performance was maximized approximately around the MLE of α, which was determined from the data in the first experiment. Since BRQRA is a new model there did not exist any method for determining an optimal α-setting under limited observation conditions. Given that no method existed, this is indeed a promising result. We will need to further test this method to see if it continues to be good for determining α in different security settings under different conditions.

Third, in reward structures 1-3 at least 1 setting of BRQRA obtained a higher expected utility than all settings of BRQR/BRQRA in the first experiment. This demonstrates the benefit of appropriately adjusting both λ and α in combination to ensure the best results. In fact, in reward structure 2 BRQRA with $\alpha = .50$ performed the best in this experiment while in the first experiment BRQR outperformed BRQRA with $\alpha = .50$. Furthermore, BRQR itself

[4]To see the full list of results please refer to http://teamcore.usc.edu/pita/results.html



(a) Reward Structure 1

(b) Reward Structure 2

(c) Reward Structure 3

(d) Reward Structure 4

Figure 3: Average Expected Utility of Defender

did not perform better in the second experiment than in the first experiment for some of the reward structures. These two examples show how appropriately adjusting both parameters in combination is important. We also point out that these results show that a poor choice of α can lead to decreased performance against BRQR.

Finally, in reward structures 1-3 BRQRA is now seen outperforming COBRA, even when compared against the results from the first experiment. This is an important result as it follows on the results found by Yang et al. (Yang et al. 2011) showing that a strategy based on QRE can lead to improved results against human adversaries. Given these three results, BRQRA currently stands as the leading contender for use in security domains to schedule limited security resources against human adversaries. The only reward structure where the last two results fail to hold is reward structure 4, which we will more closely examine in order to explain this degradation in performance.

Analysis of Reward Structure 4: To guide this analysis we present reward structure 4 and the strategies generated by BRQR in both experiment 1 and experiment 2 in Table 2. Here, the letters D and A represent the defender and attacker respectively (i.e., D. Reward is the defender's reward). For the mixed strategies generated by BRQR, we present the probability that a guard will protect each of the gates. For example, the strategy generated by BRQR_76 (i.e., BRQR with $\lambda = .76$) protects gate 2 with a guard 33% of the time.

Given the data presented in Table 2 we can begin to see why BRQR performed so poorly in reward structure 4. Specifically, one of the major problems with the strategy generated by BRQR_12 is its protection of gate 4. Notice that on gate 4 the adversary obtains a reward of 9 points. While his penalty is also high, the expectation on this door is 6.42 for the adversary against BRQR_12. For BRQR_76 on the other hand, the expectation for the adversary is only

Gate:	1	2	3	4	5	6	7	8
D. Reward	5	9	10	2	10	4	8	8
D. Penalty	-10	-4	-9	-3	-10	-10	-2	-5
A. Reward	3	7	3	9	2	9	7	8
A. Penalty	-4	-8	-5	-8	-9	-4	-1	-6
BRQR($\lambda = .76$)	.35	.33	.30	.44	.20	.62	.36	.42
BRQR($\lambda = .12$)	.47	.29	.48	.15	.39	.62	.23	.37

Table 2: BRQR in reward structure 4

1.54. More importantly, the defender's expectation is -2.24 in BRQR_12 versus -.807 in BRQR_76. Out of the 18 people who played against BRQR_12, exactly half (9 people) chose gate 4. It is clear that with such a low λ-setting, BRQR's resulting strategy leaves a significant weakness to be exploited. In fact, given the mixed strategy produced by BRQR_12, gate 4 has the highest expectation for the adversary and thus is the rational choice. This shows how a poor choice of λ can lead to significant consequences. In addition, increasing the value of α in the case of BRQRA only enhances this weakness further.

The analysis of BRQR_76 versus BRQR_12 demonstrates why BRQR performed poorly in experiment 2, however, we still need to examine BRQR versus COBRA. Between the 2 experiments, the highest average expected reward COBRA obtained was .42 and the highest BRQR/BRQRA obtained was .38. In both cases, the reason the algorithms performed well is they exploited the value of gate 7. Here the reward for both the attacker and defender was relatively high and the penalty was relatively low. In the case of BRQRA_50 (BRQRA with $\alpha = .50$ in experiment 1) this gate had the highest expected utility at 3.2208 for the attacker, but still gave the defender an expected utility of 2.724. Similarly for

COBRA_30 (COBRA with $\alpha = .30$) gate 7 had the highest expected utility at 3.728, but still gave the defender an expected utility of 2.09. While there is no definite winner between these strategies, this shows how these strategies can also exploit strengths in the payoff structures as opposed to the vulnerabilities left by BRQR in the second experiment.

Summary

While game-theoretic approaches have seen use in real-world security settings, there still remains an open issue of how to address human adversaries in these settings. Humans may not always respond with the utility maximizing rational strategy and so it is crucial to accommodate for these likely deviations. Particularly since deviating from the game-theoretic optimal strategy can lead to negative results for the defender. To that end, a number of models have been proposed to attempt to account for these deviations.

In this work we examine two models in particular, COBRA and BRQR, that were proposed specifically for security domains that can be modeled using a Stackelberg Security Game framework. BRQR has been shown to be the leading contender for addressing human adversaries under perfect observation conditions. However, we were interested in examining how these models would perform under limited observation conditions. Due to human anchoring biases under limited observation conditions, our expectation was for BRQR's performance to degrade.

Our work makes three important contributions: (i) we incorporate an anchoring bias into BRQR to improve its performance under limited observation; (ii) we explore finding appropriate parameter settings for BRQR under limited observation; (iii) we compare BRQR's performance versus COBRA under limited observation conditions. Given our results we arrived at three key conclusions. First, accounting for human adversaries and the likely deviations they will make from the optimal strategy can lead to performance increases for resource allocation in security domains. In our first experiment we show that DOBSS, which assumes a perfectly rational adversary, is always outperformed by an algorithm that accounts for human adversaries. Second, we have shown that extending BRQR to account for an anchoring bias did lead to improved performance. This shows the potential benefits of analyzing and addressing specific types of deviations that may occur. However, there is still significant work to be done since we have also shown that making a poor choice in either the model or parameter settings can lead to significant weaknesses that can easily be exploited. Finally, BRQR remains one of the leading contenders for addressing resource allocation against human adversaries, however, creating a robust model for guarding against human adversaries remains an open challenge. Indeed, we have examined only one type of potential deviation using only one method based on support theory. There are possibly better alternatives for addressing human biases due to limited observation conditions and there are a number of other human factors that could be addressed in a robust algorithm.

Acknowledgments

This research was supported by the United States Department of Homeland Security through the National Center for Risk and Economic Analysis of Terrorism Events (CREATE) under award number 2010-ST-061-RE0001. However, any opinions, findings, and conclusions or recommendations in this document are those of the authors and do not necessarily reflect views of the United States Department of Homeland Security, the University of Southern California, or CREATE.

References

Agmon, N.; Kraus, S.; Kaminka, G.; and Sadov, V. 2009. Adversarial uncertainty in multi-robot patrol. In *IJCAI*.

Camerer, C. 2003. In *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Fox, C. R., and Rottenstreich, Y. 2003. Partition priming in judgement under uncertainty. *Psychological Science* 14:195–200.

Jain, M.; Tsai, J.; Pita, J.; Kiekintveld, C.; Rathi, S.; Ordóñez, F.; and Tambe, M. 2010. Software assistants for randomized patrol planning for the LAX airport police and the Federal Air Marshals Service. *Interfaces* 40(4):267–290.

Paruchuri, P.; Marecki, J.; Pearce, J.; Tambe, M.; Ordóñez, F.; and Kraus, S. 2008. Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games. In *AAMAS*.

Pita, J.; Jain, M.; Marecki, J.; Ordóñez, F.; Portway, C.; Tambe, M.; Western, C.; Paruchuri, P.; and Kraus, S. 2008. Deployed ARMOR protection: The application of a game theoretic model for security at the Los Angeles International Airport. In *AAMAS*.

Pita, J.; Jain, M.; Tambe, M.; Ordóñez, F.; and Kraus, S. 2010. Robust solution to stackelberg games: Addressing bounded rationality and limited observations in human cognition. *AIJ* 174(15):1142–1171.

Pita, J.; Tambe, M.; Kiekintveld, C.; Cullen, S.; and Steigerwald, E. 2011. GUARDS - game theoretic security allocation on a national scale. In *AAMAS*.

See, K. E.; Fox, C. R.; and Rottenstreich, Y. S. 2006. Between ignorance and truth: Partition dependence and learning in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:1385–1402.

Tversky, A., and Koehler, D. J. 1994. Support thoery: A nonextensional representation of subjective probability. *Psychological Review* 101:547–567.

Yang, R.; Kiekintveld, C.; Ordóñez, F.; Tambe, M.; and John, R. 2011. Improving resource allocation strategy against human adversaries in security games. In *IJCAI*.

Yin, Z.; Korzhyk, D.; Kiekintveld, C.; Conitzer, V.; and Tambe, M. 2010. Stackelberg vs. Nash in security games: Interchangeability, equivalence, and uniqueness. In *AAMAS*.