Learning Bounded Rationality Models of the Adversary in Repeated Stackelberg Security Games*

Debarun Kar, Fei Fang, Francesco Delle Fave*, Nicole Sintov, Arunesh Sinha, Aram Galstyan, Bo An+, Milind Tambe
University of Southern California, Los Angeles, CA, 90089
*Disney Research, Boston, MA, 02142
+Nanyang Technological University, Singapore, 639798
{dkar,feifang,sintov,aruneshs,galstyan,tambe}@usc.edu,
*francesco.dellefave@disneyresearch.com, +boan@ntu.edu.sg

ABSTRACT

Several competing human behavior models have been proposed to model and protect against boundedly rational adversaries in repeated Stackelberg security games (SSGs). However, these existing models fail to address three main issues which are extremely detrimental to defender performance. First, while they attempt to learn adversary behavior models from adversaries' past actions ("attacks on targets"), they fail to take into account adversaries' future adaptation based on successes or failures of these past actions. Second, they assume that sufficient data in the initial rounds will lead to a reliable model of the adversary. However, our analysis reveals that the issue is not the amount of data, but that there just is not enough of the attack surface exposed to the adversary to learn a reliable model. Third, current leading approaches have failed to include probability weighting functions, even though it is well known that human beings' weighting of probability is typically nonlinear. Moreover, the performances of these models may be critically dependent on the learning algorithm used to learn the parameters of these models. The first contribution of this paper is a new human behavior model, SHARP, which mitigates these three limitations as follows: (i) SHARP reasons based on success or failure of the adversary's past actions on exposed portions of the attack surface to model adversary adaptiveness; (ii) SHARP reasons about similarity between exposed and unexposed areas of the attack surface, and also incorporates a discounting parameter to mitigate adversary's lack of exposure to enough of the attack surface; and (iii) SHARP integrates a non-linear probability weighting function to capture the adversary's true weighting of probability. Our second contribution is a comparison of two different approaches for learning the parameters of the bounded rationality models.

Our third contribution is a first "longitudinal study" – at least in the context of SSGs – of competing models in settings involving repeated interaction between the attacker and the defender. This study, where each experiment lasted a period of multiple weeks with individual sets of human subjects, illustrates the strengths and weaknesses of different models and shows the advantages of SHARP.

1. INTRODUCTION

Whereas previous real-world deployments of Stackelberg Security Games (SSGs) to protect airports, ports or flights have been one-shot game models [29], recent work has focused on domains involving repeated interactions between defenders and adversaries. These domains include security of wildlife (repeated interactions between rangers and poachers) [31], security of fisheries (repeated interactions between coast guard and illegal fishermen) [10], forest protection or drug interdiction, and are modeled via repeated SSGs. In a repeated SSG model, the defender periodically deploys new patrol strategies (in "rounds" of the game) and the adversary observes these strategies and acts accordingly.

Research in repeated SSGs has produced different approaches to address uncertainty in key dimensions of the game such as payoff uncertainty (but assuming a perfectly rational adversary) [3, 18, 21] or uncertainty in adversary behavior model (but assuming known payoffs) [10, 31]. Our work follows the second approach, learning a model of boundedly rational adversaries with known adversary payoffs, as (arguably) it provides a better fit for domains of interest in this work. This is because (i) in real-world settings such as wildlife or fishery protection, it is feasible to model adversary payoffs via animal or fish density in different locations; and (ii) there is significant support in the literature for bounded rationality of human adversaries [32, 26].

Unfortunately, despite the promise of Bounded Rationality models in Repeated Stackelberg Games (henceforth referred to as BR-RSG models), existing work in this area [10, 31] suffers from three key limitations which are extremely detrimental to defender performance. First, existing models reason about the adversary's future actions based on past actions taken but *not* the associated successes and failures. Our analysis reveals that the adversary adapts in future rounds based on past success and failure. Hence, failing to consider an adaptive adversary leads to erroneous predictions about his future behavior, and thus significantly flawed defender strategies.

Second, existing approaches for learning BR-RSG models assume that enough data will be collected in the initial rounds to learn a reliable adversary model. Our analysis reveals that the issue is not the amount of data, but insufficient exposure of *attack surface* [12, 19] in initial rounds to gather sufficient information about adversary responses to various strategies and learn a reliable model. Learning is biased towards the limited available information and hence significant losses are incurred by the defender until enough of the *right kind of data* becomes available. This degraded performance in initial rounds may have severe consequences for three reasons: (i) In domains like wildlife crime or fisheries protection, each round lasts for weeks or potentially months and so initial round losses (if massive) could imply irrecoverable losses of resources (e.g., animal populations). (ii) Following heavy losses, human defenders may lose confidence in recommendations provided by the game-theoretic approach. (iii) Given the nature of these domains, re-initialization of the game may periodically be necessary

*This paper is based on the following AAMAS'15 full paper- "A Game of Thrones": When Human Behavior Models Compete in Repeated Stackelberg Security Games. The second contribution mentioned in the abstract and discussed later in the paper along with detailed experimental results is a new contribution as compared to the AAMAS'15 paper.

and thus initial rounds may be repeated; in domains such as wildlife crime, re-initialization can stem from man-made or natural changes in parks, e.g., changes in vegetation, water bodies, or possible developmental activities. The construction of an oil-refinery [1] and the simultaneous re-introduction of rhinos in the Murchison Falls National Park in Uganda is an example. In addition, re-initializing the game after a year or so would mean repeating the initial rounds after four to five rounds, adding to the importance of addressing initial rounds.

Third, BR-RSG models have failed to include probability weighting functions (how humans "perceive" probabilities), even though it is well known that probability weighting curves for humans – e.g., in prospect theory [30] – are typically nonlinear. In light of this, we show that direct application of existing models such as SUQR [26] which assume a linear probability model, provide results that would be extremely detrimental to defender performance.

The primary contribution of this paper is a new model called SHARP (**S**tochastic **H**uman behavior model with **A**tt**R**activeness and **P**robability weighting) that mitigates these three limitations: (i) Modeling the adversary's adaptive decision making process, SHARP reasons based on success or failure of the adversary's past actions on exposed portions of the attack surface. (ii) Addressing limited exposure to significant portions of the attack surface in initial rounds, SHARP reasons about similarity between exposed and unexposed areas of the attack surface, and also incorporates a discounting parameter to mitigate adversary's lack of exposure to enough of the attack surface. (iii) Addressing shortcomings of probability weighting functions, we incorporate a two parameter probability weighting function in existing human behavior models.

One additional aspect that can influence the performance of these behavioral models is the learning algorithm used to learn the parameters of these models. Past research has only focused on using Maximum Likelihood Estimation (MLE) to learn the model parameters [26, 10, 31]. Therefore, the second contribution in this paper is to explore the effects of using Bayesian Updating to learn our model parameters by comparing its performance to that of MLE.

Our third contribution is to provide evidence from the first "longitudinal study" of competing models in repeated SSGs. In our study, a suite of well-established models and SHARP take the battlefield in an attempt to prove themselves best in repeated SSGs. Our results show: (i) SHARP outperforms existing approaches consistently over all rounds, most notably in initial rounds. (ii) As discussed earlier, existing approaches perform poorly in initial rounds with some performing poorly throughout all rounds. (iii) Surprisingly, simpler models which were originally proposed for single-shot games performed better than recent advances which are geared specifically towards addressing repeated SSGs. Taken together, given the critical importance of the repeated 'initial rounds' as discussed above, these results indicate that SHARP should be the model of choice in repeated SSGs.

2. BACKGROUND

2.1 Background on SSGs

In an SSG, the defender plays the role of a leader who protects a set of targets from the adversary, who acts as the follower [4, 27, 16]. The defender's pure strategy is an assignment of a limited number of security resources M to the set of targets T. An assignment of a resource to a target is also referred to as covering a target. A defender's mixed-strategy x ($0 \leq x_i \leq 1; i \in T$) is then defined as a probability distribution over the set of all possible pure strategies. A pure strategy of an adversary is defined as attacking a single target. The adversary receives a reward R_i^a for selecting i if it is not covered and a penalty P_i^a for selecting i if it is covered. Similarly, the defender receives a reward R_i^d for covering i if it is selected by the adversary and penalty P_i^d for not covering i if it is selected. Then, utility for the defender for protecting target i while playing mixed strategy x is:

$$U_i^d(x) = x_i R_i^d + (1 - x_i) P_i^d \tag{1}$$

Similarly, the utility for the adversary for attacking target i is:

$$U_i^a(x) = (1 - x_i) R_i^a + x_i P_i^a \tag{2}$$

Recent work has focused on modeling boundedly rational adversaries in SSGs, developing models discussed below.

Subjective Utility Quantal Response (SUQR): SUQR [26] builds upon prior work on quantal response (QR) [23] models. In the QR models, rather than strictly maximizing utility, an adversary stochastically chooses to attack targets, i.e., the adversary attacks a target with higher expected utility with a higher probability. SUQR proposes a new utility function called Subjective Utility, which is a linear combination of key features that are considered to be the most important in each adversary decision-making step. Nguyen et al. [26] experimented with three features: defender's coverage probability, adversary's reward and penalty at each target. According to this model, the probability that the adversary will attack target i is given by:

$$q_i(\omega|x) = \frac{e^{SU_i^a(x)}}{\sum_{j \in \mathbb{T}} e^{SU_j^a(x)}} \tag{3}$$

where $SU_i^a(x)$ is the Subjective Utility of an adversary for attacking target i when defender employs strategy x and is given by:

$$SU_i^a(x) = \omega_1 x_i + \omega_2 R_i^a + \omega_3 P_i^a \tag{4}$$

The vector $\omega = (\omega_1, \omega_2, \omega_3)$ encodes information about the adversary's behavior and each component of ω indicates the relative importance the adversary gives to each attribute in the decision making process. The weights are computed by performing MLE on available attack data.

Bayesian SUQR: SUQR assumes that there is a homogeneous population of adversaries, i.e., a single ω is used to represent an adversary in [26]. However, in the real-world we face heterogeneous populations. Therefore Bayesian SUQR is proposed to learn a particular value of ω for each attacker [31]. Protection Assistant for Wildlife Security (PAWS) is an application which was originally created using Bayesian SUQR.

Robust SUQR: Robust SUQR [10] combines data-driven learning and robust optimization to address settings where not enough data is available to provide a reasonable hypothesis about the distribution of ω. It computes the worst-case expected utility over all previously seen SUQR models of the adversary and deploys the optimal strategy against the adversary type that reduces the defender's utility the most. Robust SUQR is reported to be applied to the fisheries protection domain[10].

2.2 Probability Weighting Functions

Probability weighting functions model human perceptions of probability. Perhaps the most notable one is the weighting function in nobel-prize winning work on Prospect Theory [15, 30], which suggests that people weigh probability non-uniformly, as shown in Fig. 1. It indicates that people tend to overweigh low probabilities and underweigh high probabilities. Other works in this domain propose and experiment with parametric models which capture both inverse S-shaped as well as S-shaped probability curves [8] (Fig. 2).

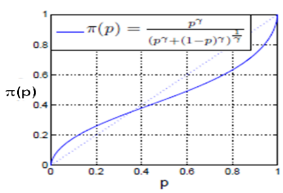


Figure 1: Probability Weighting Function (Prospect Theory)

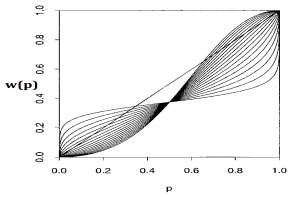


Figure 2: Probability Weighting Function (Gonzalez & Wu, 99)

3. RELATED WORK

We have already discussed related work in SSGs in the previous section, including key behavioral models. Here we discuss additional related work:

Learning in repeated Stackelberg games: The problem of learning the adversary's payoffs in an SSG by launching a minimum number of games against a perfectly rational adversary is studied in [18, 3]. Additionally, Marecki et al. [21] focused on optimizing the defender's overall utility during the learning process when faced with a perfectly rational adversary with unknown payoffs.

Robust strategies in repeated games: In cases where the opponent cannot be successfully modeled, McCracken et al. [22] proposed techniques to generate ϵ-safe strategies which bound the loss from a safe value by ϵ. Johanson et al. [14, 13] studied the problem of generating robust strategies in a repeated zero-sum game while exploiting the tendency in the adversary's decision making and evaluated their technique in a game of two-player, Limit Texas Hold'em. Recently, Ponsen et al. [28] proposed techniques to compute robust best responses in partially observable stochastic games using sampling methods.

All of the above work differs from ours in three ways: (i) They do not model bounded rationality in human behavior; (ii) They do not consider how humans weigh probabilities; and (iii) None of these existing work address the important problem of significant losses in the initial rounds. This is a critical problem in domains such as wildlife security as explained above; requiring a fundamental shift at least in the learning paradigm for SSGs. In addition, work on learning in SSGs differs because in our game, the payoffs are known but we are faced with boundedly rational adversaries whose parameters in their behavioral model are to be learned.

4. WILDLIFE POACHING GAME

We conducted *longitudinal experiments*[1] [20] with human subjects to test the effectiveness of existing behavioral models and algorithms against our proposed approach on repeated SSGs.

4.1 Game Overview

In our game, human subjects play the role of poachers looking to place a snare to hunt a hippopotamus. The game interface is shown in Fig. 3. In the game, the portion of the park shown in the map is divided into a 5*5 grid, i.e. 25 distinct cells. Overlaid on the Google Maps view of the park is a heat-map, which represents the rangers' mixed strategy x — a cell i with higher coverage probability x_i is shown more in red, while a cell with lower coverage probability is shown more in green. As the subjects play the game, they are given the following detailed information: R_i^a, P_i^a and x_i for each target i. However, they do not know the pure strategy that will be played by the rangers, which is drawn randomly from

[1]Whereas "longitudinal study" is often used to describe research that spans years – in which measurement occasions are conducted every X years – we use the term longitudinal study because our study included 5 measurement points with a single population.

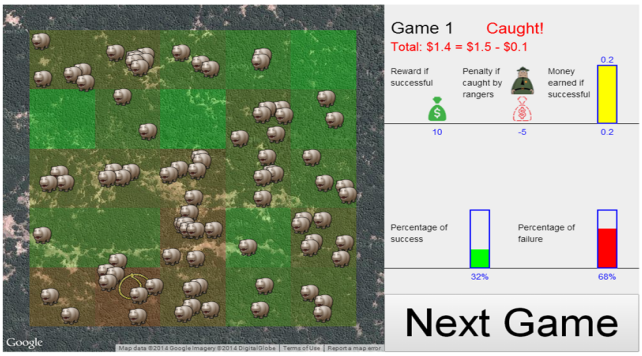


Figure 3: Game Interface for our simulated online repeated SSG (Reward, penalty and coverage probability for a selected cell are shown)

mixed strategy x shown on the game interface. Thus, we model the real-world situation that poachers have knowledge of past pattern of ranger deployment but not the exact location of ranger patrols when they set out to lay snares. In our game, there were $M = 9$ rangers protecting this park, with each ranger protecting one grid cell. Therefore, at any point in time, only 9 out of the 25 distinct regions in the park are protected.

In addition to animal density, which is strongly correlated with high-risk areas of poaching [25, 24, 9], distance is another important factor in poaching, e.g., recent snare-density studies have found that significant poaching happens within 5 kilometers of South Africa's Kruger National Park border [17]. Therefore, the reward obtained by a poacher in successfully placing a snare at target i is calculated by discounting the animal density by a factor of the distance traveled and is calculated as follows:

$$R_i^a = int(\phi_i - \zeta * \frac{D_i}{\max_j(D_j)}) \quad (5)$$

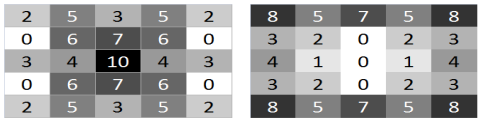
Here, ϕ_i and D_i refer to the animal density at target i and the distance to target i from the poacher's starting location respectively. $int(y)$ rounds the value y to the closest integer value. The parameter ζ is the importance given to the distance factor in the reward computation and may vary based on the domain. When the poacher successfully poaches, he may thus obtain a reward that is less than the animal density (Eqn. 5), but the defender loses a value equal to that of the animal density, i.e., the game is non-zero-sum. For our experiments we set P_i^a and R_i^d to constant values over all targets.

4.2 Experimental Procedures

We recruited human subjects on Amazon Mechanical Turk (AMT). We first primed participants with a background story about the hardships of a poacher's life. To enhance understanding of the game, participants played two trial games, one validation game, and finally the actual game. Data from subjects who played the validation game incorrectly were discarded.

We tested a set of behavioral models introduced in Section 2.1 by deploying the mixed strategy generated based on each of these models repeatedly over a set of five rounds. For each model, we recruited a new set of participants to eliminate any learning bias. Due to unavailability of data, the strategy shown for each first round was Maximin. We then learned the model parameters based on previous rounds' data, recomputed and redeployed strategies, and asked the *same* players to play again in the subsequent rounds. For each model, all five rounds were deployed over a span of weeks.

4.3 Payoff Structures



(a) ADS_1 (b) ADS_2

Figure 4: Animal density structures

The payoff structures used in our human subject experiments vary in terms of the animal densities and hence the adversary rewards. We henceforth refer to payoff structures and animal density structures interchangeably in this paper. The total number of animals in all the payoffs we generate is the same (= 96). However, the variation in these payoffs is in the way that the animals are spread out in the park. In payoff structure 1, the animal density is concentrated towards the center of the park, whereas the animal density is higher towards the edges of the park in payoff structure 2. These represent scenarios that might happen in the real world. The animal density for both payoffs is symmetric, thus eliminating any bias due to the participant's starting point in the game. Figs. 4(a)–4(b) show heatmaps of two animal density structures, denoted as ADS_1 and ADS_2 respectively. More details can be found here[2].

5. SHARP: PROBABILITY WEIGHTING

This paper contributes a novel human behavior model called SHARP for BR-RSG settings. SHARP has three key novelties, of which probability weighting is the first one. The need for probability weighting became apparent after our initial experiments. In particular, initially following up on the approach used in previous work [26, 33, 31, 10], we applied MLE to learn the weights of the SUQR model based on data collected from our human subject experiments and found that the weights on coverage probability were positive for all the experiments. That is, counter-intuitively humans were modeled as being attracted to cells with high coverage probability, even though they were *not* attacking targets with very high coverage but they were going after targets with moderate to very low coverage probability. Examples of the learned weights for SUQR from data collected from the first round deployment of the game for 48 human subjects on ADS_1 and ADS_2 are: (ω_1, ω_2, ω_3)=(2.876, -0.186, 0.3) and (ω_1, ω_2, ω_3)=(1.435, -0.21, 0.3). We prove a theorem (Theorem 5.1) to show that, when the weight on the coverage probability in the SUQR model (ω_1) is found to be positive, the optimal defender strategy is a pure strategy. The proof of the theorem can be found here[2].

THEOREM 5.1. *When $\omega_1 > 0$, the optimal defender strategy is a pure strategy.*

Employing a pure strategy means that there will be no uncertainty about the defender's presence. Several cells will always be left unprotected and in those cells, the attackers will always succeed. In our example domains, even if the top-valued cells are covered by a pure strategy, we can show that such a strategy would lead to significantly worse defender expected utility than what results from the simplest of our defender mixed strategies deployed. For example, if cells of value 4 are left unprotected, the defender expected value will be -4, which is much lower than what we achieve even with Maximin. In repeated SSG domains like wildlife crime, this would mean that the poachers successfully kill animals in each round without any uncertainty of capture by rangers.

We hypothesize that this counter-intuitive result of a model with $\omega_1 > 0$ may be because the SUQR model may not be considering people's *actual* weighting of probability. SUQR assumes that people weigh probabilities of events in a linear fashion, while existing work on probability weighting (Sec. 2.2) suggests otherwise. To address this issue, we augment the Subjective Utility function with a two-parameter probability weighting function (Eqn. 6) proposed by Gonzalez and Wu [8], that can be either inverse S-shaped (concave near probability zero and convex near probability one) or S-shaped.

$$f(p) = \frac{\delta p^\gamma}{\delta p^\gamma + (1-p)^\gamma} \tag{6}$$

The SU of an adversary denoted by 'a' can then be computed as:

$$SU_i^a(x) = \omega_1 f(x_i) + \omega_2 R_i^a + \omega_3 P_i^a \tag{7}$$

where $f(x_i)$ for coverage probability x_i is computed as per Eqn. 6. The two parameters δ and γ control the elevation and curvature of the function (Fig. 2) respectively. $\gamma < 1$ results in an inverse S-shaped curve while $\gamma > 1$ results in an S-shaped curve.

We will refer to this as the PSU (Probability weighted Subjective Utility) function and the models (SUQR, Bayesian SUQR and Robust SUQR) augmented with PSU will be referred to as P-SUQR, P-BSUQR and P-RSUQR respectively. *Our SHARP model will use PSU*. We will use these PSU-based models in our experiments.

One of our key findings, based on experiments with the PSU function is that the curve representing human weights for probability is *S-shaped in nature, and not inverse S-shaped* as prospect theory suggests. The S-shaped curve indicates that people would overweigh high probabilities and underweigh low to medium probabilities. Examples of learned curves are shown in Sec. 9.2. Recent studies [2, 11, 7] have also found S-shaped probability curves which contradict the inverse S-shaped observation of prospect theory. Given S-shaped probability weighting functions, the learned ω_1 was negative as it accurately captured the trend that significantly higher number of people were attacking targets with low to medium coverage probabilities and *not* attacking high coverage targets.

Feature Selection and Weight Learning: In Sec. 4.1, we introduced a new feature – distance – that affected the reward and hence the obvious question for us was to investigate the effect of this new feature in predicting adversary behavior. We considered several variations of PSU with different combinations of features and found that it gives better prediction accuracy when the following four features are used while computing the Subjective Utility of the adversary: coverage probability, animal density, adversary penalty and distance from starting location, as shown in Eqn. 8.

$$SU_i^a(x) = \omega_1 f(x_i) + \omega_2 \phi_i + \omega_3 P_i^a + \omega_4 D_i \tag{8}$$

We learn a 6-tuple $b =< \delta, \gamma, \omega_1, \omega_2, \omega_3, \omega_4 >$ (δ and γ due to inclusion of Eqn. 6) from available data. To learn the behavioral parameters b from available data, we propose an algorithm based on Repeated Random Sub-sampling Validation (see online appendix[2]). For P-SUQR , we learn a single b, while for P-BSUQR and P-RSUQR we learn a set of $b \in \mathbb{B}$ for each attacker.

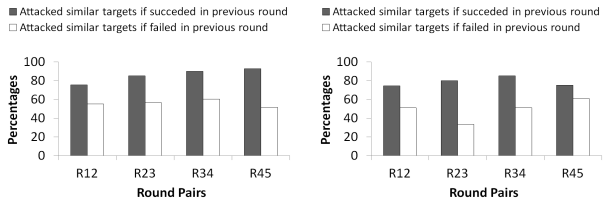
Based on our experiments, in addition to $\omega_1 < 0$, we also found $\omega_2 > 0$, $\omega_3 < 0$ and $\omega_4 < 0$. The rest of the formulations in this paper will be based on these observations about the feature weights.

6. SHARP: ADAPTIVE UTILITY MODEL

A second major innovation in SHARP is the adaptive nature of the adversary and addressing the issue of attack surface exposure. First, we define key concepts, present evidence from our experiments, and then present SHARP's innovations.

Definition The ***attack surface*** α is defined as the n-dimensional

[2] http://onlineappendixalaworkshop2015.weebly.com/



(a) Maximin ADS_2 (b) P-RSUQR ADS_2

Figure 5: Evidence for adaptivity of attackers

space of the features used to model adversary behavior. Formally, $\alpha =< F^1, F^2, ..., F^n >$ for features $F^j (\forall j; 1 \leq j \leq n)$.

For example, as per the PSU model in Eqn. 8, this would mean the space represented by the following four features: coverage probability, animal density, adversary penalty and distance from the starting location.

Definition A ***target profile*** $\beta_k \in \alpha$ is defined as a point on the attack surface α and can be associated with a target. Formally, $\beta_k =< F_k^1, F_k^2, ..., F_k^n >$ denotes the kth target profile on the attack surface.

In our example domain, the kth target profile can be represented as $\beta_k =< x_{\beta_k}, \phi_{\beta_k}, P^a_{\beta_k}, D_{\beta_k} >$, where $x_{\beta_k}, \phi_{\beta_k}, P^a_{\beta_k}$ and D_{β_k} denote values for coverage probability, animal density, attacker penalty and distance from starting location respectively[3]. For example, <0.25, 2, -1, 4> is the target profile associated with the top-leftmost cell in ADS_1 for the first round. Exposing the adversary to a lot of different target profiles would therefore mean exposing the adversary to more of the attack surface and gathering valuable information about their behavior. While a particular target location, defined as a distinct cell in 2-d space, can only be associated with one target profile in a particular round, more than one target may be associated with the same target profile in the same round. β_k^i denotes that target profile β_k is associated with target i in a particular round.

6.1 Observations and Evidence

Below are two observations from our human subjects data, based on the above concepts, that reveal interesting trends in attacker behavior in repeated SSGs.

OBSERVATION 1. *Adversaries who have succeeded in attacking a target associated with a particular target profile in one round, tend to attack a target with 'similar' target profiles in next round.*

OBSERVATION 2. *Adversaries who have failed in attacking a target associated with a particular target profile in one round, tend* ***not*** *to attack a target with 'similar' target profiles in the next round.*

In order to provide evidence in support of Observations 1 and 2, we show results from our data highlighting these trends on ADS_2 in Figs. 5(a) - 5(b). Results of other models on ADS_1 and ADS_2 can be found in the online appendix[2]. In each plot, the y-axis denotes the percentage of (i) attacks on similar targets out of the total successful attacks in the previous round (ζ_{ss}) and (ii) attacks on similar targets out of the total failed attacks in the previous round (ζ_{fs}). The x-axis denotes pairs of rounds for which we are computing the percentages, for example, in R12, 1 corresponds to round $(r-1)$ and 2 means round r in our claim. Thus, ζ_{ss} corresponding to R23 in ADS_2 is 80%, meaning that out of all the people who succeeded in round 2, 80% attacked similar target profiles in round 3. Similarly, ζ_{fs} corresponding to R23 in ADS_2 is 33.43%, meaning that out of all people who failed in round 2, 33.43% attacked similar target profiles in round 3. All statistical significance results reported below are on two-tailed t-tests at confidence=0.05. The average (over all four models on two payoffs and for all round pairs) of ζ_{ss} is 75.2% and the average of ζ_{fs} which is 52.45%. This difference is statistically significant, thus supporting Observation 1 and Observation 2.

These observations are also consistent with the "spillover effect" in psychology [6], which in our case suggests that an adversary will tend to associate properties of unexposed target profiles with knowledge about similar target profiles to which he has been exposed, where similarity is expressed in terms of the Euclidean distance between two points on the attack surface. Smaller distance indicates higher similarity. The above aspects of adversary behavior currently remain unaccounted for, in BR-RSG models. Based on observations above, we define two key properties below to capture the consequences of past successes and failures on the adversary's behavior and reason based on them.

Definition The ***vulnerability*** associated with a target profile β_i which was shown to the adversary in round r, denoted $V^r_{\beta_i}$, is defined as a function of the total number of successes and failures on the concerned target profile in that round (denoted by $success^r_{\beta_i}$ and $failure^r_{\beta_i}$ respectively). This is shown in Eqn. 9:

$$V^r_{\beta_i} = \frac{success^r_{\beta_i} - failure^r_{\beta_i}}{success^r_{\beta_i} + failure^r_{\beta_i}} \quad (9)$$

Therefore, more successful attacks and few failures on a target profile indicate that it was highly vulnerable in that round. Because multiple targets can be associated with the same target profile and the pure strategy generated based on the mixed strategy x in a particular round may result in a defender being present at some of these targets while not at others, there may be both successes and failures associated with the same target profile in that round.

Definition The ***attractiveness*** of a target profile β_i at the end of round R, denoted $A^R_{\beta_i}$, is defined as a function of the vulnerabilities for β_i from round 1 to round R. It is computed using Eq. 10.

$$A^R_{\beta_i} = \frac{\sum_{r=1}^{R} V^r_{\beta_i}}{R} \quad (10)$$

Therefore, we model the attractiveness of a target profile as the average of the Vulnerabilities for that target profile over all the rounds till round R. This is consistent with the notion that a target profile which has led to more successful attacks over several rounds will be perceived as more attractive by the adversary.

6.2 SHARP's Utility Computation

Existing models (such as SUQR, which is based on the subjective utility function (Eqn. 4)) only consider the adversary's actions from round $(r-1)$ to predict their actions in round r. However, based on our observations (Observations 1 & 2), it is clear that the adversary's actions in a particular round are dependent on his past successes and failures. The *adaptive* probability weighted subjective utility function proposed in Eq. 11 captures this adaptive nature of the adversary's behavior by capturing the shifting trends in attractiveness of various target profiles over rounds.

$$ASU^R_{\beta_i} = (1 - d * A^R_{\beta_i})\omega_1 f(x_{\beta_i}) + (1 + d * A^R_{\beta_i})\omega_2 \phi_{\beta_i} + (1 + d * A^R_{\beta_i})\omega_3 P^a_{\beta_i} + (1 - d * A^R_{\beta_i})\omega_4 D_{\beta_i} \quad (11)$$

[3]In our experiments, $\phi_{\beta_i} > 0$, $P^a_{\beta_i} < 0$ and $D_{\beta_i} > 0$

There are three main parts to SHARP's computation: (i) Adapting the subjective utility based on past successes and failures on exposed parts of the attack surface; (ii) Discounting to handle situations where not enough attack surface has been exposed; and (iii) Reasoning about similarity of unexposed portions of the attack surface based on other exposed parts of the attack surface.

The intuition behind the adaptive portion of this model is that, the subjective utility of target profiles which are highly attractive to the adversary should be increased, and that of less attractive target profiles should be decreased, to model the adversary's future decision making. Hence, for a highly attractive target profile β_i, the attacker would view the coverage x_{β_i} and distance from starting location D_{β_i} to be of much lower value, but the animal density ϕ_{β_i} to be of higher value, as compared to the actual values. The contribution of the penalty term would also increase the utility (recall that $P^a_{\beta_i} < 0$ and $\omega_3 < 0$). Taking an example from our game, for a target profile $\beta_i =< 0.25, 2, -1, 4 >$ which had $A^1_{\beta_i} = 1$ after round 1, and the weights learned were $b =< \delta, \gamma, \omega_1, \omega_2, \omega_3, \omega_4 > =< 2.2, 2.4, -3, 0.9, -0.3, -0.5 >$, P-SUQR would compute the subjective utility as -0.29, while (assuming d (explained later) $= 0.25$, for example) SHARP's adaptive utility function would compute the subjective utility as 0.855. In comparison to the original subjective utility function, this function is adaptive due to the positive or negative boosting of model weights based on the defender's knowledge about the adversary's past experiences. Here, learning the model parameters b has been decoupled from boosting the model parameters for future prediction to ensure simplicity in terms of both the model formulation as well the weight learning process.

Now we turn to the next aspect of SHARP's utility computation. Recall the problem that the defender does not necessarily have information about the attacker's preferences for enough of the attack surface in the initial rounds. This is because, the attacker is exposed to only a limited set of target profiles in each round and the defender progressively gathers knowledge about the attacker's preferences for only those target profiles. We provide evidence in support of this observation in Sec. 9.3.

The parameter d ($0 \leq d \leq 1$) in Eqn. 11 mitigates this attack surface exposure problem. It is a discounting parameter which is based on a measure of the amount of attack surface exposed. d is low in the initial rounds when the defender does not have enough of the right kind of data, but would gradually increase as more information about the attacker's preferences about various regions of the attack surface become available. For our experiments, we varied d based on Eqn. 12:

$$d = \frac{1}{N_r - r} \tag{12}$$

where N_r is the total number of rounds and r is the round whose data is under consideration. For example, $N_r = 5$ and $r = 1$ for data collected in round 1 of an experiment conducted over five rounds. The intuition behind this formulation is that, as more rounds are played, more information about the adversary' preferences for a lot of the attack surface will be available and hence d will increase from a very small value gradually as rounds progress.

Finally, we look at how we reason about unexposed portions of the attack surface based on the exposed areas. If a target profile β_u was not exposed to attacker response in round r, the defender will not be able to compute the vulnerability $V^r_{\beta_u}$. Therefore, we will also not be able to estimate the attractiveness for β_u and hence the optimal defender strategy. So, in keeping with our analysis on available data and based on the spillover effect introduced earlier, we use the distance-weighted k-nearest neighbors algorithm [5] to obtain the Vulnerability $V^r_{\beta_u}$ of an unexposed target profile β_u in round r, based on the k most similar target profiles which were exposed to the attacker in round r (Eqns. 13 and 14).

$$V^r_{\beta_u} = \frac{\sum_{i=1}^{k} \theta_i * V^r_{\beta_i}}{\sum_{i=1}^{k} \theta_i} \tag{13}$$

$$\theta_i \equiv \frac{1}{d(\beta_u, \beta_i)^2} \tag{14}$$

where, $d(\beta_u, \beta_i)$ denotes the Euclidean distance between β_u and β_i in the feature space. We use $k = 5$ for our experiments.

7. GENERATING DEFENDER STRATEGIES AGAINST SHARP

While SHARP provides an adversary model, we must now generate defender strategies against such a model. To that end, we first learn the parameters of SHARP from available data (See Sec. 5). We then generate future round strategies against the boundedly rational adversary characterized by the learned model parameters by solving the following optimization problem:

$$\max_{x \in \mathbb{X}} \left[\sum_{i \in \mathbb{T}} U^d_i(x)\, q^R_i(\omega \mid x) \right] \tag{15}$$

$q^R_i(\omega|x)$ is the probability that the adversary will attack target i in round R and is calculated based on the following equation:

$$q^R_i(\omega|x) = \frac{e^{ASU^R_{\beta^i_k}(x)}}{\sum_{i \in \mathbb{T}} e^{ASU^R_{\beta^i_k}(x)}} \tag{16}$$

β^i_k denotes that target profile β_k is associated with target i. $ASU^R_{\beta^i_k}$ and $U^d_i(x)$ are calculated according to Eqns. 11 and 1 respectively.

To solve the non-linear and non-convex optimization problem in Eqn. 15 and generate the optimal defender strategy, we use PASAQ [34] as it provides an efficient approximate computation of the defender strategy with near-optimal solution quality.

8. LEARNING SUQR PARAMETERS: ALTERNATIVES TO MLE

Instead of using MLE to learn a particular $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$ from data collected till a particular round of the game, we can alternatively perform Bayesian updating on available data. Let us consider a discrete set of samples Ω of weight vectors, i.e., $\Omega =< \omega^1, \omega^2, ..., \omega^{|\Omega|} >$ where each element in the set is a 4-tuple, i.e. $\omega^s = (\omega^s_1, \omega^s_2, \omega^s_3, \omega^s_4)\,; \forall s = 1 to |\Omega|$. Let us define a probability distribution $p =< p^1, p^2, ..., p^{|\Omega|} >$ over the elements in Ω. Then, as we collect data in a particular round r, we can update our prior probability distribution before round r, denoted as p_{r-1}, to the posterior distribution p_r. Now, assuming a homogeneous population of adversaries as in SUQR, we can use Algorithm 1 to update p_{r-1} to p_r as all the attacks are considered to be performed by a single adversary. The distribution learned from data collected over multiple rounds tells us which is the most probable adversary type causing the attacks. Algorithm 1 helps us get closer to that true ω. In Algorithm 1, χ_{r-1} is the attack data collected in round $r-1$, χ^i_{r-1} is the target number for the ith attack in round $r-1$ and x_{r-1} is the mixed strategy deployed in round $r-1$.

However, considering a heterogeneous population of adversaries as in Bayesian SUQR, we can use Algorithm 2 to update p_{r-1} to p_r. In Algorithm 2, $|\chi^i_{r-1}|$ denotes the number of attacks performed on target i in round $r-1$. In this case, the probability

Algorithm 1 BU (x_{r-1}, χ_{r-1}, Ω, p_{r-1})

Output: estimated distribution p_r.

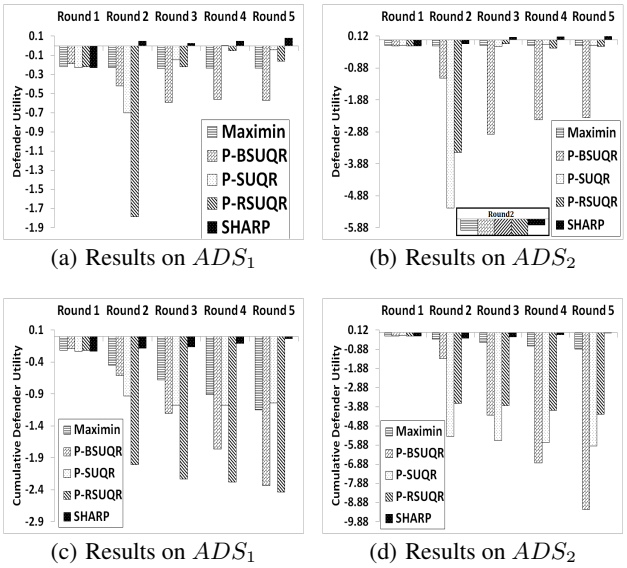
1: $b = p_{r-1}$
2: **for** i=1 to $|\chi_{r-1}|$ **do**
3: **for** s=1 to $|\Omega|$ **do**
4: $b^s = \frac{b^s * q_{\chi^i_{r-1}}(\omega^s, x_{r-1})}{\sum_m b^m * q_{\chi^i_{r-1}}(\omega^m, x_{r-1})}$
5: **end for**
6: **end for**
7: $p_r = b$;

distribution p gives us the probability of occurrence of each attacker type in a heterogeneous population of adversaries. Line 3 of Algorithm 2 updates the prior based on an attack on a particular target. Then line 7 computes the weighted average of these separate probability distributions based on the number of attacks on the corresponding targets (i.e., by the corresponding attacker type). In Section 9.4, we will explore the effects of using Algorithms 1 and 2, as opposed to using MLE.

Algorithm 2 BU-Learn (x_{r-1}, χ_{r-1}, Ω, p_{r-1})

Output: estimated distribution p_r.

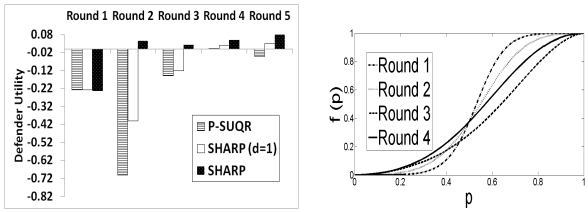
1: **for** i=1 to $|T|$ **do**
2: **for** s=1 to $|\Omega|$ **do**
3: ${}^i p^s_r = \frac{p^s_{r-1} q_i(\omega^s, x_{r-1})}{\sum_m p^m_{r-1} q_i(\omega^m, x_{r-1})}$
4: **end for**
5: **end for**
6: **for** s=1 to $|\Omega|$ **do**
7: $p^s_r = \frac{\sum_i |\chi^i_{r-1}| p^s_r}{\sum_i |\chi^i_{r-1}|}$
8: **end for**



(a) Results on ADS_1 (b) Results on ADS_2

(c) Results on ADS_1 (d) Results on ADS_2

Figure 6: (a), (b): Defender utilities for various models on ADS_1 and ADS_2 respectively; (c), (d): Cumulative defender utilities for various models on ADS_1 and ADS_2 respectively.



(a) Results on ADS_1 (b) Probability curves

Figure 7: (a) Comparison of defender utilities between P-SUQR, SHARP and SHARP(d=1) on ADS_1; (b) Learned probability curves for P-SUQR on ADS_1 from rounds 1 to 4.

9. RESULTS WITH HUMAN SUBJECTS

9.1 Defender Utilities

In Figs. 6(a)- 6(b) we show actual defender utilities obtained over 5 rounds for P-SUQR, P-BSUQR, P-RSUQR, SHARP and Maximin on ADS_1 and ADS_2 respectively, with an average of 37 human subjects playing per round. In the plots, y-axis corresponds to defender utility and the models tested for each round is shown on the x-axis. For example, in Fig. 6(b), P-SUQR performs worst in round 2 with utility of -5.26. In Fig. 6(b), we also show (inset) zoomed in results of the second round to highlight the difference in performance between Maximin (= -0.18) and SHARP (= -0.1). First round utilities for all models are same as Maximin strategy was played due to absence of data. All significance results reported below are computed with bootstrap t-test. Following are key observations from our experiments.

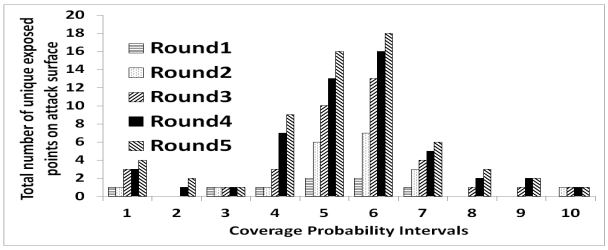
Heavy initial round losses: For all models except SHARP, there is statistically significant (p=0.05) loss in defender utility as compared to Maximin in round 2. P-SUQR recovers from initial round losses and outperforms Maximin in rounds 3, 4 and 5 for ADS_1 (statistically significant at p=0.05), and in round 4 (statistically significant at p=0.15) and round 5 for ADS_2. P-RSUQR, which is a robust model, also outperforms Maximin in rounds 4 and 5 (statistically significant at p=0.05) for ADS_1 after initial round losses. Surprisingly, P-BSUQR, which is the basis for wildlife security application PAWS, performs worst on both payoffs over all rounds. Figs. 6(c)- 6(d) show cumulative defender utility over five rounds on ADS_1 and ADS_2 respectively. Observe that it takes five rounds for P-SUQR to recover from initial round losses and outperform Maximin in terms of cumulative defender utility for ADS_1 (Fig. 6(c)). None of the other models recover from the initial round losses in five rounds, thus highlighting the impact of initial round losses on model performance for a long period of time.

Performance of SHARP against other models: SHARP consistently outperforms (statistically significant at p=0.05) all the models over all rounds (Figs. 6(a)- 6(b)), most notably in initial rounds (round 2) and ends up with significantly high cumulative utility at the end of all rounds (Figs. 6(c)- 6(d)).

Performance of SHARP (with and without discounting): To test the effectiveness of the design decisions in SHARP, we considered SHARP both with and without discounting. SHARP with $d = 1$ is compared against SHARP and P-SUQR on ADS_1 in Fig. 7(a). SHARP($d = 1$) outperforms P-SUQR (statistically significant at p=0.05) because it captures the adaptive nature of the adversary. However, it performs worse than SHARP (statistically significant at p=0.01) as SHARP also trusts the data less when we don't have enough information about the adversary's responses to most of the attack surface; in this case the initial rounds.

Therefore, our results on extensive human subjects experiments on repeated SSGs show SHARP's ability to perform well throughout, including the important initial rounds.

9.2 Learned Probability Curves



(a) For ADS_1

Figure 8: Total number of unique exposed target profiles till the end of each round for each coverage probability interval for ADS_1.

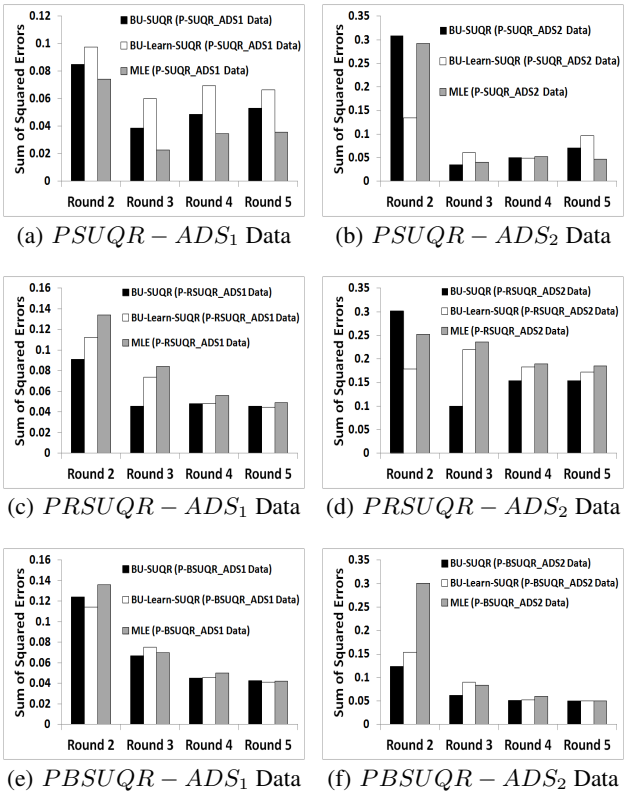
Fig. 7(b) shows human perceptions of probability in rounds 1 to 4 when the participants were exposed to P-SUQR based strategies on ADS_1. Learned curves from P-SUQR and SHARP on all payoffs have this S-shaped nature (See online appendix[2]), showing that even though there is little change in the curvature between rounds, it retains the same S-shape throughout all rounds. The curves indicate that people weigh high probabilities to be higher and low to medium probabilities to be lower than the actual values.

9.3 Attack surface exposure

In our repeated SSG, the only variation in terms of feature values for our model (Eqn. 11) from round to round is the mixed strategy x and hence the coverage x_i at each target. Hence, exposure to various regions of the attack surface means exposure to various values of x_i for fixed values of the other model parameters. Fig. 8 shows how the adversary was exposed to more unique values of the coverage probability, and hence attack surface, over the five rounds for ADS_1. We discretized the range of x_i, i.e. [0,1] into 10 intervals (x-axis) and show the total number of unique coverages exposed till a particular round (y-axis) for each interval. Observe that more interval ranges and more unique coverage probabilities get exposed in rounds 3 to 5. As we showed in Fig. 6(a), the defender performance for P-SUQR improves significantly in rounds 4 and 5. Similar plot for ADS_2 is shown in the online appendix[2].

9.4 Bayesian Updating vs MLE

In this section, we present results of using Algorithms 1 and 2 and evaluate their performance in comparison to MLE on available human subjects data collected in our experiments. We computed the probability of attack on each target in round r (for the actual deployed strategy in round r) after learning from attack data collected in the previous rounds. We then computed the sum of squared errors over all targets with respect to the actual attack probability for each round. The results are shown in Figs. 9(a) - 9(f). We observe that Algorithm 2 (denoted as BU-Learn-SUQR) performs significantly better than MLE in round two on four out of six datasets (statistically significant via t-test with p=0.05). The performances are similar on other rounds. Algorithm 1 (denoted as BU-SUQR) performs significantly better than MLE in round two on two out of six datasets (statistically significant via t-test with p=0.05), on four datasets in round three (statistically significant via t-test with p=0.05) and has similar or better performances in other rounds. This indicates that Bayesian updating can be a competitive approach to learning weights for the behavioral models and highlights the need to conduct human subjects experiments with these algorithms to test their effectiveness.



(a) $PSUQR - ADS_1$ Data (b) $PSUQR - ADS_2$ Data

(c) $PRSUQR - ADS_1$ Data (d) $PRSUQR - ADS_2$ Data

(e) $PBSUQR - ADS_1$ Data (f) $PBSUQR - ADS_2$ Data

Figure 9: Sum of Squared Errors (SSE) over all targets of per round predictions for MLE, BU-Learn-SUQR and BU-SUQR on various datasets.

10. CONCLUSION

This paper provides three major contributions that are critical for important domains such as protection of wildlife, fish and forests. First, it introduces a novel human behavior model called SHARP for repeated SSG settings. SHARP has three major novelties: (i) It models the adversary's adaptive decision making process by reasoning based on success or failure of the adversary's past actions on exposed portions of the attack surface. (ii) It accounts for lack of information about the adversary's preferences due to insufficient exposure to attack surface by reasoning about similarity between exposed and unexposed areas of the attack surface, and also incorporating a confidence based discounting parameter to model the learner's trust in the available data. (iii) It integrates a non-linear probability weighting function. Second, we explored the importance of different learning approaches, for example, Bayesian updating, and found the need to conduct further human subjects experiments to test its performance and compare it to MLE based learning approaches. Third, we conducted the first "longitudinal study" of competing models in repeated SSGs to test the performance of SHARP along with existing approaches. Our results show that: (i) Human perceptions of probability are S-shaped, contradicting the inverse S-shaped observation from prospect theory. (ii) Existing human behavior models and algorithms perform poorly in initial rounds of repeated SSGs. (iii) SHARP performs significantly better than existing approaches consistently over all the rounds, most notably in the initial rounds.

11. ACKNOWLEDGMENTS

This research was supported by MURI Grant W911NF-11-1-03.

12. REFERENCES

[1] Ugandans fear curse of oil wealth as it threatens to blight 'pearl of africa'. http://www.theguardian.com/world/2013/dec/29/ugandans-oil-blight-pearl-africa. Accessed: November 8, 2014.

[2] Y. Alarie and G. Dionne. Lottery decisions and probability weighting function. *Journal of Risk and Uncertainty*, 22(1):21–33, 2001.

[3] A. Blum, N. Haghtalab, and A. Procaccia. Learning optimal commitment to overcome insecurity. In *Proceedings of the 28th Annual Conference on Neural Information Processing Systems, series = NIPS, year = 2014, location = Quebecl, Canada*.

[4] V. Conitzer and T. Sandholm. Computing the optimal strategy to commit to. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, EC '06, pages 82–90, 2006.

[5] S. A. Dudani. The distance-weighted k-nearest-neighbor rule. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-6(4):325–327, April 1976.

[6] J. Elster. A plea for mechanisms. *Social mechanisms: an analytical approach to social theory*.

[7] N. Etchart-Vincent. Probability weighting and the level and spacing of outcomes: An experimental study over losses. *Journal of Risk and Uncertainty*, 39(1):45–63, 2009.

[8] R. Gonzalez and G. Wu. On the shape of the probability weighting function. *Cognitive psychology - Vol 38*, pages 129–166, 1999.

[9] M. Hamisi. *Identification and mapping risk areas for zebra poaching: A case of Tarangire National Park, Tanzania*. Thesis, ITC, 2008.

[10] W. Haskell, D. Kar, F. Fang, M. Tambe, S. Cheung, and E. Denicola. Robust protection of fisheries with compass. In *Innovative Applications of Artificial Intelligence (IAAI)*, 2014.

[11] S. J. Humphrey and A. Verschoor. The probability weighting function: experimental evidence from Uganda, India and Ethiopia. *Economics Letters*, 84(3):419–425, September 2004.

[12] S. Jajodia, A. K. Ghosh, V. Swarup, C. Wang, and X. S. Wang. *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats*. Springer Publishing Company, Incorporated, 1st edition, 2011.

[13] M. Johanson and M. Bowling. Data biased robust counter strategies. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS-09*, 2009.

[14] M. Johanson, M. Zinkevich, and M. Bowling. Computing robust counter-strategies. In *In Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS*, 2007.

[15] D. Kahneman and A. Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–91, 1979.

[16] D. Korzhyk, V. Conitzer, and R. Parr. Complexity of computing optimal stackelberg strategies in security resource allocation games. In *In Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 805–810, 2010.

[17] A. M. Lemieux. *Situational Crime Prevention of Poaching (Crime Science Series)*. Routledge, 2014.

[18] J. Letchford, V. Conitzer, and K. Munagala. Learning and approximating the optimal strategy to commit to. In *Proceedings of the 2Nd International Symposium on Algorithmic Game Theory*, SAGT '09, pages 250–262, Berlin, Heidelberg, 2009. Springer-Verlag.

[19] P. K. Manadhata and J. M. Wing. An attack surface metric. *Software Engineering, IEEE Transactions on*, 37(3):371–386, 2011.

[20] A. Mao, D. Parkes, Y. Chen, A. D. Procaccia, K. Z. Gajos, and H. Zhang. Turkserver: Enabling synchronous and longitudinal online experiments. *In AAAI HCOMP Workshop*, 2012.

[21] J. Marecki, G. Tesauro, and R. Segal. Playing repeated stackelberg games with unknown opponents. In *AAMAS*, pages 821–828, 2012.

[22] P. McCracken and M. Bowling. Safe strategies for agent modelling in games. In *In Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2004.

[23] D. McFadden. Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, 5(4):363–390, 1976.

[24] M. Montesh. Rhino poaching: A new form of organised crime. Technical report, College of Law Research and Innovation Committee of the University of South Africa, 2013.

[25] W. Moreto. *To Conserve and Protect: Examining Law Enforcement Ranger Culture and Operations in Queen Elizabeth National Park, Uganda*. Thesis, Rutgers, 2013.

[26] T. H. Nguyen, R. Yang, A. Azaria, S. Kraus, and M. Tambe. Analyzing the effectiveness of adversary modeling in security games. *In AAAI*, 2013.

[27] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordonez, and S. Kraus. Playing games for security: An efficient exact algorithm for solving bayesian stackelberg games. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS, pages 895–902, 2008.

[28] M. Ponsen, S. D. Jong, and M. Lanctot. Computing approximate nash equilibria and robust best-responses using sampling. *J. Artif. Intell. Res. (JAIR)*, 2011.

[29] M. Tambe. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned*. Cambridge University Press, New York, NY, 2011.

[30] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[31] R. Yang, B. Ford, M. Tambe, and A. Lemieux. Adaptive resource allocation for wildlife protection against illegal poachers. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2014.

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

[33] R. Yang, C. Kiekintveld, F. Ordonez, M. Tambe, and R. John. Improving resource allocation strategies against human adversaries in security games: An extended study. *Artif. Intell.*, 195:440–469, 2013.

[34] R. Yang, F. Ordonez, and M. Tambe. Computing optimal strategy against quantal response in security games. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2*, AAMAS '12, pages 847–854, 2012.