Effectiveness of Probability Perception Modeling and Defender Strategy Generation Algorithms in Repeated Stackelberg Games: An Initial Report

Debarun Kar[1], Fei Fang[1], Francesco Maria Delle Fave[1], Nicole Sintov[1], Milind Tambe[1], Arlette van Wissen[2]

[1]University of Southern California, Los Angeles, CA 90007, USA
{dkar, feifang, dellefav, sintov, tambe}@usc.edu

[2]VU University Amsterdam, 1081 HV Amsterdam, The Netherlands
a.van.wissen@vu.nl

Abstract

While human behavior models based on repeated Stackelberg games have been proposed for domains such as "wildlife crime" where there is repeated interaction between the defender and the adversary, there has been no empirical study with human subjects to show the effectiveness of such models. This paper presents an initial study based on extensive human subject experiments with participants on Amazon Mechanical Turk (AMT). Our findings include: (i) attackers may view the defender's coverage probability in a non-linear fashion; specifically it follows an S-shaped curve, and (ii) there are significant losses in defender utility when strategies generated by existing models are deployed in repeated Stackelberg game settings against human subjects.

Introduction

Algorithms based on Stackelberg security games (SSGs) have been deployed across multiple domains over the past decade and are currently in use by major security agencies such as the US Coast Guard (USCG), the Federal Air Marshal Service (FAMS) and the Los Angeles Airport (LAX) police (Tambe 2011). In most of these applications, the interaction between the defender (the security agency) and the attacker (terrorist) has been represented as a one-shot game where the attacker was defined as a perfectly rational player. The idea was to represent a generic terrorist attack, whereby a single target is attacked, thus generating terrible social and economic consequences.

Recent research in SSGs have started to tackle new domains such as (i) "wildlife security" where rangers and poachers are engaged in a repeated tussle and (ii) "fisheries protection" where defending agencies like the USCG are constantly trying to protect fish stocks from illegal fishermen. Here the defender and the adversary are engaged in repeated interactions where the defender deploys new patrolling strategies periodically and the adversary observes these strategies and acts accordingly. Models and algorithms have been proposed to address repeated SSGs against boundedly rational adversaries (Yang et al. 2014; Haskell et al. 2014). The key idea in this literature is to use behavioral models such as quantal response (QR) (Yang et al. 2011) and subjective utility quantal response (SUQR) (Nguyen et al. 2013) to model human adversaries. Unfortunately, these algorithms have not yet been evaluated against human subjects in repeated games. Therefore, it is unclear as to how these models will perform in repeated games where a defender repeatedly encounters a group of human adversaries. This is also a key requirement since human subjects experiments (HSE) on AMT have become a standard test-bed to determine the quality of behavioral models and algorithms (Pita et al. 2010; 2012; Nguyen et al. 2013; Yang et al. 2013).

To address this challenge, our work presents an empirical study whereby we compare the state-of-the-art behavioral models for SSGs in a number of human subjects experiments on repeated SSGs. To run our experiments, we designed and deployed a "Wildlife Poaching Game" in which human participants played the role of poachers to collect data and use it to estimate their behavior. Whereas this study is still ongoing, we identified two interesting findings.

The first observation is that attackers do not necessarily view probabilities in a linear fashion. In light of this, we show that the direct application of existing models such as QR (Yang et al. 2011) and SUQR (Nguyen et al. 2013) which assume a linear probability model, provide results that would be extremely detrimental to defender performance. To address this shortcoming, we incorporate a 2-parameter probability weighting function in existing human behavior models. This function accounts for the attacker's *actual* perception of coverage probabilities. We present an algorithm to learn the attacker's true perception of probability along with the weights of the behavior models from empirical data. Surprisingly, our results show that people behave following a probability weighting function which is the inverse of the well-known prospect theoretic function (Tversky and Kahneman 1992), i.e. it is an S-shaped function. The second finding is that, results from our human subject experiments show that existing models perform poorly in terms of the defender utility in the initial rounds. Here, one round indicates the deployment of a particular defender strategy in a repeated SSG. In the domain of wildlife poaching, such initial round losses would mean a lot of animals killed, thus adding to the importance of addressing these initial rounds in repeated SSGs.

Background and Related Work

Models for predicting the behavior of human adversaries were developed in order to address bounded rationality in the human decision making process. Below we introduce a few solution concepts and some of these human behavior models along with the key parameters of these models.

Maximin MAXIMIN is a robust game-theoretic solution concept that would generate a defender strategy against an attacker who will attack the target which will minimize the defender utility the most.

Quantal Response (McFadden 1976; McKelvey and Palfrey 1995) It is a stochastic choice model which attributes a probability distribution of attacks over each target, thus introducing randomness in the adversary's decision making process. This model is based on the notion of Quantal Response Equilibrium (QRE) . It is assumed that, instead of strictly maximizing their expected utility, individuals respond stochastically in games, i.e. the adversary will attack a target having higher expected utility with higher probability.

Subjective Utility Quantal Response One variant of the QR model is the Subjective Utility Quantal Response (SUQR) model (Nguyen et al. 2013), which proposes a new utility function called the Subjective Utility. This function is a linear combination of key features that are considered to be the most important factors for the adversary at each decision-making process. Nguyen et al. (2013) experimented with 3 features: the defender's coverage probability, the adversary's reward and penalty at each target. According to this model, the probability that the adversary will attack target t is given by:

$$q_t(\omega|x) = \frac{e^{\lambda * SU_t^a(x)}}{\sum_{t \in \mathbb{T}} e^{\lambda * SU_t^a(x)}} \quad (1)$$

where $SU_t^a(x)$ is the Subjective Utility (SU) of an attacker for attacking target t when the defender employs strategy x and is given by:

$$SU_t^a(x) = \omega_1 x_t + \omega_2 R_t^a + \omega_3 P_t^a \quad (2)$$

The reward and penalty for the attacker for attacking target t are R_t^a and P_t^a respectively. x_t is the coverage probability for target t. The vector $\omega = (\omega_1, \omega_2, \omega_3)$ encodes information about the behavior of the adversary and each component of ω indicates how much importance is given by the attacker to each attribute in his decision making process. The weights are computed by performing Maximum Likelihood Estimation (MLE) on available attacker behavior data.

Bayesian SUQR SUQR assumes that there is a homogeneous population of adversaries and hence there is a single ω used to represent an attacker in (Nguyen et al. 2013). However, in the real-world we face an entire population of heterogeneous adversaries. Therefore Bayesian SUQR is proposed to learn a particular value of ω for each attack (Yang et al. 2014). The most recent deployed application which uses the Bayesian SUQR framework is Protection Assistant for Wildlife Security (PAWS), which is an application deployed at the Queen Elizabeth National Park in Uganda to generate optimal anti-poaching patrols for park rangers.

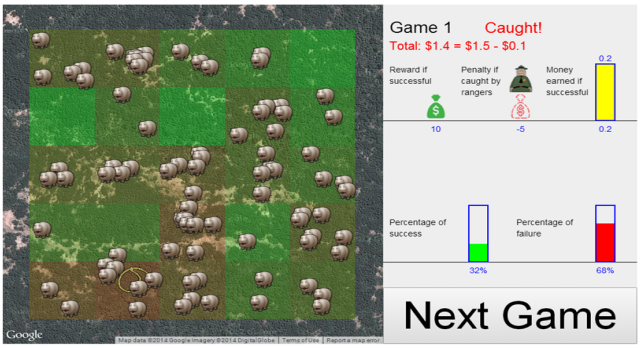

Figure 1: Game Interface for our simulated online SSG

Prospect Theory Prospect Theory is a framework for decision-making under uncertainty that captures (i) variations in how people perceive probabilities through a probability weighting function, and (ii) people's risk preferences. The probability weighting function captures the tendency for individuals to overestimate low probabilities and underestimate high probabilities (Tversky and Kahneman 1992). Yang et al.(2011) proposed two models of human behavior based on Prospect Theory.

Robust SUQR Making a reasonable hypothesis about the distribution of ω may not always be possible due to lack of data in domains like "wildlife crime" and hence, performing updates to improve our estimate of the true probability distribution as proposed by (Yang et al. 2014) will not always be accurate. Robust SUQR (Haskell et al. 2014) combines both data-driven learning and robust optimization into a single framework by computing the worst-case expected utility over all previously seen SUQR models of the adversary and hedging against the adversary type that reduces the defender's utility the most.

Our Contributions

We developed a "Wildlife Poaching Game" to simulate the 'wildlife poaching' scenario and to conduct human subject experiments on repeated games. In this game, the strategies generated by various algorithms are essentially strategies for rangers to protect animals from poachers in a protected wildlife park. The game interface is shown in Fig. 1. Human subjects play the role of a poacher who is looking to place a snare to hunt a hippopotamus. Overlaid on the Google Maps view of the park is a heat-map, which represents the rangers' mixed strategy, i.e., a randomized allocation of security resources at each target i, which is represented by x_i. As the subjects play the game, they are given detailed information about the reward, penalty and coverage probability of a particular region of the park. However, they do not know the exact location of the rangers. The targets to be protected by the rangers in a particular game are drawn randomly from the coverage probability distribution shown on the game interface. A player is said to succeed if he places a snare in a region which is not protected by a ranger, else he is said to be unsuccessful. The reward for successfully placing a snare

in a region is calculated by taking into account: 1) number of animals in the region and 2) the distance the poacher has to cover from his starting location to go to that region.

Probability Weighted SUQR (PW-SUQR)

While performing human subject experiments with the SUQR model, we found that the weights generated for the SUQR model were sometimes unintuitive, given the data from our experiments. More specifically, the learned weight for coverage probability was sometimes found to be *positive*, even when it was evident by observing the data that a significantly higher number of people have attacked targets with low coverage probability as compared to the number of attacks on targets with high coverage. We also propose a theorem (Theorem 1) to show that, when the weight on the coverage probability in the SUQR model (ω_1 in Eqn. 2) is greater than zero, the optimal defender strategy is a pure strategy. Employing a pure strategy is not in the defender's best interest as it would lead to a lower expected utility as compared to the case where the defender employs a mixed strategy. The proof of the theorem can be found here[1].

Theorem 1. *When $\omega_1 > 0$, the optimal defender strategy is a pure strategy.*

This led us to hypothesize that the SUQR model may not be considering people's *actual* perception of probability. SUQR assumes that people view probabilities of events in a linear fashion, while prior work on Prospect Theory suggests that people have a non-uniform perception of probability. The empirical form of the probability weighting function $\pi(p_i)$, where p_i is the actual probability, from their paper (Kahneman and Tversky 1979; Tversky and Kahneman 1992) is shown in Fig. 2. Prospect theory indicates that people tend to be risk averse when making decisions about low probability events but risk seeking with respect to events occurring with high probabilities. In our work, it would mean that targets with low to medium probabilities will not be attacked by a significantly large number of people as they would view low coverage probabilities to be higher than they actual are. Similar arguments can be used to infer about attacks on targets with higher coverage probabilities. However, we observed the opposite trend in our data: a significantly higher number of individuals were attacking targets with low to medium coverage probabilities, indicating that they are risk seeking when making decisions about low probability events but risk averse with respect to events occurring with high probabilities. Therefore, our observations about the attack data indicate that the probability weighting function may be S-shaped in nature, and not inverse S-shaped as prospect theory suggests. Recent studies have also found S-shaped probability curves (Etchart-Vincent 2009) which contradict the inverse S-shaped observation of prospect theory.

In order to address this issue, we augment the Subjective Utility function with a two-parameter probability weighting function (Equation 3) (Gonzalez and Wu 1999), that can be either inverse S-shaped (concave near probability zero and convex near probability one) or S-shaped. The two parameters δ and γ control the elevation and curvature of the function respectively. Intuitively, δ and γ can be interpreted as the attacker's sensitivity to dispersion and skewness of the outcome of the attack. $\gamma < 1$ results in a prospect theoretic curve while $\gamma > 1$ results in an S-shaped curve.

[1] http://onlineappendixrsg.weebly.com/

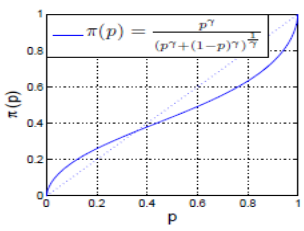

Figure 2: Probability Weighting Function (Prospect Theory)

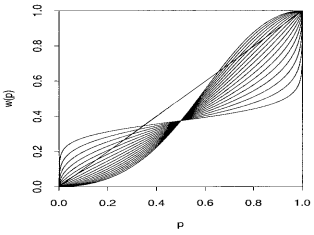

Figure 3: Probability Weighting Function (Gonzalez & Wu, 99)

$$f(p) = \frac{\delta p^{\gamma}}{\delta p^{\gamma} + (1-p)^{\gamma}} \tag{3}$$

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

$$SU_t^a(x) = \omega_1 f(x_t) + \omega_2 R_t^a + \omega_3 P_t^a \tag{4}$$

where $f(x_t)$ for a particular coverage probability x_t is computed as per Eqn. 3. We will henceforth refer to this as the PW-SU (Probability Weighted Subjective Utility) function and the models (SUQR, Bayesian SUQR and Robust SUQR) augmented with PW-SU will be referred to as PW-SUQR, PW-BSUQR and PW-RSUQR respectively. We will use these models in our experiments.

Now, unlike previous work in (Nguyen et al. 2013) where the targets were laid out in a single row, our game is based on a 2-D setting where distance from the starting point may have an impact on the poacher's behavior. Therefore, we considered several variations of PW-SU with different combinations of features and found that the performance of Eqn. 5 is statistically significant as compared to the other models. AD_t refers to the animal density at target t.

$$SU_t^a(x) = \omega_1 f(x_t) + \omega_2 AD_t + \omega_3 P_t^a + \omega_4 D_t \tag{5}$$

So we now have to learn the values of 6 behavioral parameters ($b =< \delta, \gamma, \omega_1, \omega_2, \omega_3, \omega_4 >$) from available data. Learning these values is non-trivial due to the non-convexity of Eqns. 1 and 3. Although a non-linear solver may be applied directly to learn the 6 parameter tuple altogether, it would be inefficient and may lead to high degradation in solution quality due to the high dimensions of the search space and the need for large amounts of data. Therefore, we propose an algorithm (Algorithm 1) based on Repeated Random Sub-sampling Validation to learn b in an effective way. For SUQR, we learn a single b, while for PW-BSUQR and PW-RSUQR we learn a set of $b \in \mathbb{B}$ for each attack.

Empirical Validation

We conducted experiments with human workers on Amazon Mechanical Turk to evaluate the effectiveness of current

Algorithm 1 Algorithm to learn the weights of the PW-SUQR and its variations

OUTPUT: Learned weights (δ_f, γ_f, ω_1, ω_2, ω_3, ω_4).

1: Randomly divide the collected data D into 1 training (Tr) and 1 test (Te) set.
2: Take the training samples (Tr) and randomly divide it into K training (Trv) and validation (Val) splits.
3: Consider a range of values for both δ and γ in Eqn. 3.
4: Discretize each range and consider all possible $\{\delta\ ,\ \gamma\}$ pairs in that range. Let there be M such pairs.
5: **for** i=1 to M **do**
6: **for** j=1 to K **do**
7: Given training split Trv_j, learn the weights ω^j=(ω_1^j, ω_2^j, ω_3^j, ω_4^j) of Eqn. 5.
8: Predict using the learned weights ω^j on the corresponding validation split Val_j.
9: Calculate the prediction error Err_j on the validation set Val_j.
10: **end for**
11: Calculate the average of all K prediction errors Err_j (j=1 to K) and let that be denoted by $AvgErr_i$.
12: **end for**
13: Let p be the index of the $\{\delta\ ,\ \gamma\}$ pair with the minimum $AvgErr_i$ (i=1 to M). Then choose $\{\delta_p\ ,\ \gamma_p\}$ as the parameter values of the probability weighting function that best describes the probability perception of the adversary population.
14: Given training set Tr and $\{\delta_p\ ,\ \gamma_p\}$, learn the weights ω=(ω_1, ω_2, ω_3, ω_4) of Eqn. 5 by performing MLE. The final learned weight set is then (δ_f, γ_f, ω_1, ω_2, ω_3, ω_4).

behavioral models in repeated Stackelberg game settings using our "Wildlife Poaching Game". We recruited a set of participants who played each of 5 rounds of the game, with the initial round being MAXIMIN and the subsequent round strategies were generated based on human behavior models learned from the aggregated data till the previous round. However, as mentioned before, applying SUQR directly results in inappropriate learned weights for the model. An example of the learned weights for SUQR from data collected from the first round deployment of the game for 48 human subjects is (ω_1, ω_2, ω_3)=(2.876, -0.186, 0.3). The most important aspect to note here is that the weight on coverage probability is positive which according to Theorem 1 will generate a pure strategy and is hence harmful to the defenders. Therefore, we learned the weights of PW-SUQR on the same dataset and obtained the following behavioral parameters $b = (\delta, \gamma, \omega_1, \omega_2, \omega_3) = (0.6, 3, -2.72, 0.27, 0.3)$, which is also consistent with our observation of the attack data. Fig. 5 shows people's perception of probability in rounds 1 to 4 when they were exposed to PW-SUQR based strategies. Note that each of these curves is S-shaped. We also conducted experiments with human subjects data from (Nguyen et al. 2013) on one-shot games. The probability perception curves learned from the data is shown in Fig. 7.

In Fig. 6 we show actual defender utilities obtained over 5 rounds for PW-SUQR, PW-BSUQR, PW-RSUQR and

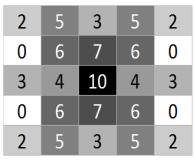

Figure 4: Animal Density Structure

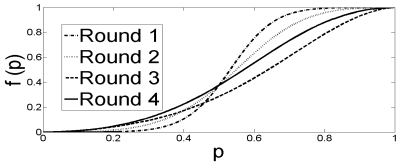

Figure 5: Probability Curves for our wildlife game from round 1 to round 4

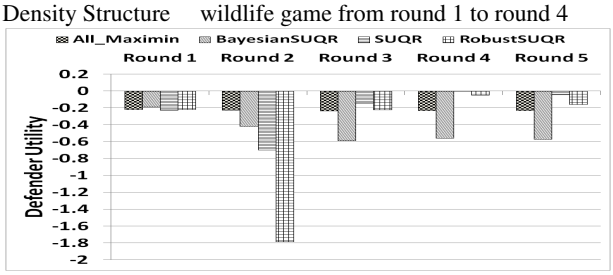

Figure 6: Defender Utilities

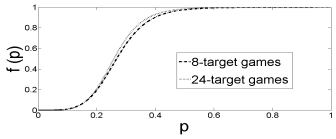

Figure 7: Probability Curves for 8-target and 24-target games on dataset in (Nguyen et al. 2013)

MAXIMIN on the animal density structure shown in Fig. 4. The middle most cell has the highest animal density of 10. The most important observation based on the performance of the models is that there is significant loss in defender utility in the initial rounds for all the behavioral models. While PW-SUQR and PW-RSUQR recover from the initial round losses and perform better than Maximin as rounds progress, it takes 4 rounds of play to recover from the initial round losses and beat Maximin in terms of the cumulative defender utility. PW-BSUQR performs the worst over all the 5 rounds.

Conclusion

We observe that the attacker in an SSG may view probabilities in a non-linear way and it should be incorporated in human behavior models to predict the attacker's behavior more accurately. We show through extensive human subject experiments that modeling it using a 2-parameter probability weighting function (i) results in an S-shaped probability curve which contradicts the inverse S-shaped observation of prospect theory, and (ii) the weights learned for the SUQR model are consistent with the observed attack data, specifically it does not generate positive weights for coverage probability which is otherwise detrimental to defender performance. We also present experimental results with the probability weighted human behavior models on a repeated Stackelberg game setting and show that most of these models perform significantly worse in the initial rounds, thus adding to the importance of addressing these initial rounds in such repeated game settings.

Acknowledgments

We would like to thank Major Adam Ackerman, Elizabeth Carpenter, Robert Guiterrez and Sierra Kelly from the United States Air Force Academy for helping us with the experiments on the dataset from (Nguyen et al. 2013).

References

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

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

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

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

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

McKelvey, R. D., and Palfrey, T. R. 1995. Quantal response equilibria for normal form games. *Games and Economic Behavior* 2:6–38.

Neyman, J., and Pearson, E. S. 1933. On the problem of the most efficient tests of statistical hypotheses. *Royal Society of London Philosophical Transactions Series A* 231:289–337.

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

Pita, J.; Jain, M.; Ordonez, F.; Tambe, M.; and Kraus, S. 2010. Solving stackelberg games in the real-world: Addressing bounded rationality and limited observations in human preference models. *Artificial Intelligence Journal* 174(15):1142–1171.

Pita, J.; John, R.; Maheswaran, R.; Tambe, M.; and Kraus, S. 2012. A robust approach to addressing human adversaries in security games. *In ECAI*.

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

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

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

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

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