Advances in Cognitive Systems X (20XX) 1-6 Submitted X/20XX; published X/20XX

Believe It or Not: Modeling Adversary Belief Formation in Stackelberg Security Games with Varying Information

Debarun Kar[1] DKAR@USC.EDU
Subhasree Sengupta[1] SUBHASRS@USC.EDU
Ece Kamar[2] ECKAMAR@MICROSOFT.COM
Eric Horvitz[2] HORVITZ@MICROSOFT.COM
Milind Tambe[1] TAMBE@USC.EDU
[1] University of Southern California, Los Angeles, CA 90089 USA
[2] Microsoft Research, One Microsoft Way, Redmond WA 98052 USA

Abstract

There has been significant amount of research in Stackelberg Security Games (SSG), and a common assumption in that literature is that the adversary perfectly observes the defender's mixed strategy. However, in real-world settings the adversary can only observe a sequence of defender pure strategies sampled from the actual mixed strategy. Therefore, a key challenge is the modeling of adversary's belief formation based on such limited observations. The SSG literature lacks a comparative analysis of these models and a principled study of their strengths and weaknesses. In this paper, we study the following shortcomings of previous work and introduce new models that address these shortcomings. First, we address the lack of empirical evaluation or head-to-head comparison of existing models by conducting the first-of-its-kind systematic comparison of existing and new proposed models on belief data collected from human subjects on Amazon Mechanical Turk. Second, we show that assuming a homogeneous population of adversaries, a common assumption in the literature, is unrealistic based on our experiments, which highlight four heterogeneous groups of adversaries with distinct belief update mechanisms. We present new models that address this shortcoming by clustering and learning these disparate behaviors from data when available. Third, we quantify the value of having historical data on the accuracy of belief prediction.

1. Introduction

A Stackelberg Security Game (SSG) (Kiekintveld et al. (2009)) is a game between a defender, who plays the role of a leader by deploying her limited security resources to protect a set of targets, and an adversary, who acts as the follower by taking an action after observing the defender's strategy. Single-shot SSGs have been successfully used in the past by security agencies for the protection of airports, ports and flights (Tambe (2011)). Recent research in SSGs has focused on domains involving repeated interactions between the defenders and adversaries, such as the "Green Security Game" domains, i.e. security of wildlife (Fang et al. (2016)) and fisheries (Haskell et al. (2014)).

In an SSG, the defender's pure strategy is an assignment of a limited number of security resources to the set of targets. The defender's mixed-strategy is then defined as a probability distribu-

tion over the set of all possible pure strategies. An equivalent description (Korzhyk et al. (2010)) of these mixed strategies is a probability distribution over the set of targets. In both single-shot and repeated game domains the defender first computes an optimal mixed strategy and then deploys pure strategies (protection at some targets) which are sampled from the optimal mixed strategy. Most existing work in SSGs on modeling adversary behavior assumes that adversaries have access to the actual mixed strategy of the defender while optimizing their own attack strategies (Tambe (2011)).

However, the above assumption does not always hold in real-world setings. Therefore, a key challenge in these settings is the modeling of adversary's belief formation about the defender's strategy based on limited observations. Several models have been proposed, both in the SSG literature (An et al. (2012); Pita et al. (2010)) as well as in psychology (See et al. (2006)) that address this problem in different ways. While An et al. (2012) proposed a Bayesian belief update model assuming perfectly rational adversaries, Pita et al. (2010) proposed a linear mixture model assuming boundedly rational adversaries. The goal of this paper is to present a comprehensive study of belief formation models applicable to SSGs, highlighting their strengths and shortcomings and introducing new computational models to address these shortcomings. The key contributions are as follows.

First, the literature lacks empirical evaluation or a head-to-head comparison of existing belief formation models. Indeed, in the absence of a comprehensive analysis it is unclear as to which model(s) are better suited for estimating adversary beliefs in SSGs. To address this shortcoming, we conducted the first-of-its-kind systematic comparison of existing and new proposed models of adversary belief update. Our extensive analysis with 24 different models (we present 16 models in this paper due to lack of space) on human subjects data collected from Amazon Mechanical Turk (AMT) through a simulated online SSG game highlights key insights about the human belief update process and demonstrates the strengths and weaknesses of these models. Second, existing belief update models assume the presence of a homogeneous population of adversaries with the same belief update mechanism (An et al. (2012); Pita et al. (2010)). However, our analysis shows the presence of four heterogeneous groups of adversaries with distinct belief update processes. We present a new model called $B\text{-}REACT$ (Belief model for heteRogenEous Adversaries using ClusTering) that addresses this shortcoming by learning about the adversary's beliefs based on historical data combined with a clustering based approach. We demonstrate that this new model completely outperforms existing and other proposed models, thus emphasizing the importance of modeling heterogeneity in human belief formation. Third, existing work simply assumes that no historical data about adversary beliefs will be available. Therefore, the literature lacks models that can take advantage of historical data (when available) by learning about the adversary's belief update process and then making more accurate belief predictions. Therefore, we propose models for settings where data is available and quantify the value of having population-wide or historical data on belief prediction accuracy.

2. Background

As briefly mentioned earlier, 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 (Kiekintveld et al. (2009)). 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 in an SSG can be compactly represented as a probability distribution over the set of targets: x $(0 \leq x_i \leq 1; \forall x_i, i \in T; \sum_{i=1}^{|T|} x_i = M)$.

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. The expected utility for the adversary for attacking target i is $U_i^a(x) = (1 - x_i)R_i^a + x_i P_i^a$. For simplification purposes, we assume a zero-sum game and therefore the defender's expected utility is $U_i^d(x) = -U_i^a(x)$. Although a perfectly rational adversary would choose to attack the target with the highest expected utility, more recent work has focused on modeling boundedly rational adversaries in SSGs . Below we introduce two models for generating the optimal defender strategy, one considering a perfectly rational adversary while the other assumes a boundedly rational adversary. Pure strategies sampled from the optimal mixed strategies computed based on these models were used in our game to collect data about the adversary's belief estimation process.

Maximin: A game-theoretic concept that generates an optimal defender strategy assuming a perfectly rational adversary who attacks the target that minimizes the defender's utility the most.

Subjective Utility Quantal Response (SUQR): SUQR (Nguyen et al. (2013)) is a popular human behavior model used in SSGs that builds upon prior work on quantal response (QR) (McFadden (1976)). It proposes a utility function called Subjective Utility ($SU_i^a(x)$; $i \in T$), which is a weighted linear combination of key features that are considered to be the most important in each adversary decision-making step: $SU_i^a(x) = \omega_1 x_i + \omega_2 R_i^a + \omega_3 P_i^a$. The probability that an adversary 'a' will attack target i is given in Eqn. 1. The optimal strategy to deploy against the adversary is then obtained by maximizing the defender's expected utility (Eqn. 2) based on a learned ω.

$$q_i(\omega|x) = \frac{e^{SU_i^a(x)}}{\sum_{j \in T} e^{SU_j^a(x)}} \quad (1) \qquad \max_{x \in \mathbb{X}} \left[\sum_{i \in \mathbb{T}} U_i^d(x)\, q_i(\omega \mid x) \right] \quad (2)$$

3. Belief Modeling Game

We conducted human subjects experiments on AMT to collect data about how humans update their beliefs about the defender's mixed strategy while acting as adversaries based on their observations about the defender. Below is an overview of our experimental game, the payoff structures and defender strategies used and the model categories tested.

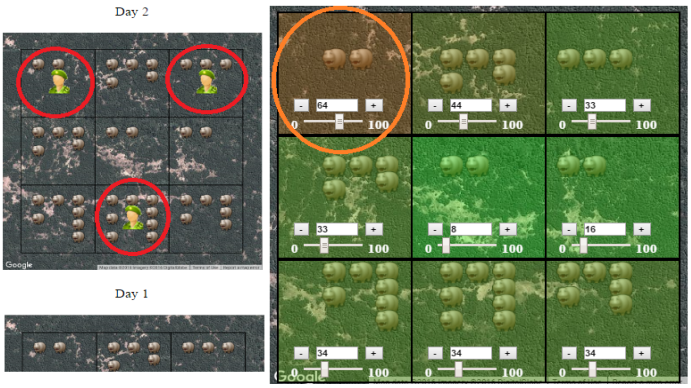


Figure 1. Game Interface for simulated online belief modeling game

3.1 Game Overview

In our game, human subjects play the role of poachers (a type of adversary) who are trying to estimate the defender's mixed strategy by observing 10 consecutive pure strategies sampled independently from the corresponding defender

mixed strategy. Each pure strategy corresponds to the strategy used by the defender on one particular day for patrolling the protected park area. At the end of each day, the participants were required to enter their beliefs about the defender's mixed strategy based on their pure strategy observations till the current day. The game interface is shown in Fig. 1.

In our game, the Google maps view of the portion of the park shown in the interface is divided into a 3*3 grid, i.e. 9 distinct target cells. Overlaid on this map to the right of the interface is a heat-map which represents the participants' current belief about the rangers' mixed strategy x — a cell i where the participant believes that a defender has higher coverage probability x_i is shown more in red, while a cell with lower coverage probability is shown more in green. The participants can use the sliders, text boxes and +/- buttons to enter their beliefs about the percentage likelihood of a ranger being present in each cell and this change will be reflected by the color of that cell. As the subjects play the game, they are given information about the presence/absence of a ranger for each target i for each day as shown by the map in the left of the game interface. Fig. 1 shows the defender's pure strategy for Day 2 in the map on the left (three rangers are circled) and in the right map the participant is currently entering his/her beliefs (64% coverage on top leftmost target) about the defender's strategy after having observed two days of defender patrols. The participant can check all the previous days' patrols (pure strategies) by scrolling down in the left side of the interface before entering their beliefs. In our game, $M = 3$ rangers were protecting 3 out of 9 grid cells in the park. So, for any day, only 3 out of the 9 targets are shown to be protected in the per day maps shown in the left of the interface.

As mentioned earlier, the pure strategies shown to the left were drawn independently from a defender mixed strategy x. This is the mixed strategy that the participants were asked to estimate based on the pure strategy observations. This setting simulates a real-world situation where poachers have knowledge of previous ranger deployments in terms of their exact locations per day and they are tasked to form beliefs about the actual mixed strategy based on these observations. In this paper we are only interested in modeling the belief formation and update procedures in such scenarios and hence only collect data about their beliefs and do *not* ask them to choose a target to attack after any day of play in the game.

3.2 Experimental Procedure

After an introduction to the game setting, the participants had to answer two validation questions which tested their understanding of the game, and were allowed to proceed to a trial and then the actual game if they answered them correctly. In the actual game, one of four mixed strategies was randomly selected for each participant to eliminate any bias and he/she was shown the 10 pure strategies sampled from the chosen mixed strategy.

Payment Scheme: We set up the payment scheme to not only reward participation but also to incentivize truthful reporting of the participants' beliefs. Specifically, each participant was paid a 'base compensation' for participation. To motivate the participants to enter their beliefs accurately after each day, we gave them an incentive called 'performance bonus', based on the difference between the entered beliefs after each day and the actual mixed strategy from which the pure strategies were sampled. The total reward was the sum of their performance bonuses and base compensation.

Payoff Structures: We randomly generated two game boards showing how animals are spread out across the 9 targets, which determines the payoff structure for the game. We henceforth refer to payoff structures and animal density structures interchangeably in this paper. The total number of animals on the board is constant across games (= 40). Figs. 2(a)–2(b) show animal densities used; they are referred to as ADS_1 and ADS_2 respectively in the paper.

Defender Strategies: We experimented with four different defender strategies to test how humans form and update their beliefs when faced with different strategies. These are: (i) Maximin, (ii) Proportional, (iii) SUQR, and (iv) Uniform. We show Maximin and SUQR strategies for ADS_1 in Figs. 2(c) – 2(d). Proportional strategy puts coverage probabilities on targets in proportion to the number of animals in that target. In a Uniform strategy, each target is covered with equal probability. Since three defenders were protecting 9 targets, sum of the coverages (in terms of percentages) is ≤ 300. We ensured that participants do not enter beliefs for each target outside the range of [0,100] and/or enter beliefs such that the sum is more than 300 by showing a pop-up message if they attempted to submit beliefs outside the allowable range or if the sum is ≥ 300. Coverages and adversary's beliefs about the coverages can be computed in terms of either probabilities or percentages.

2	5	3
5	2	2
7	8	6

(a) ADS_1

8	3	5
10	4	2
4	2	2

(b) ADS_2

0	49	15
49	0	0
64	68	58

(c) Maximin

0	41	0
41	0	0
66	100	52

(d) SUQR

Figure 2. (a,b): Animal Densities; (c) Maximin; (d) SUQR

We deployed our game on AMT and collected data for 191 and 160 participants for ADS_1 and ADS_2 respectively. Since each participant was randomly allocated to a condition corresponding to one of the four mixed strategies, the number of participants for each condition in the resulting data set varies. In our experiments with ADS_1, Maximin, Proportional, SUQR and Uniform strategies were played by 35, 55, 44 and 57 participants respectively. We divided each of these four groups of participants randomly into 10 train-test splits with 70% of the participants in the training data and remaining 30% from the same split in the test data. Training data (whenever used) is for learning our models. We will make belief predictions for participants in the test sets. Non-learning models were evaluated on the same test sets as the learning models to enable fair comparison.

Models Tested: The literature on belief modeling can be broadly categorized as: (a) Bayesian updating models; (b) Heuristic belief updating; (c) Bayesian Theory of Mind (BTOM); and (d) Level-k models. In this paper, we will provide a description of models that fall in categories (a) and (b) only, as these were earlier shown to be the best performing models in the SSG literature and other related fields (e.g., psychology). We have extensively experimented with such models and we present those results in Sec. 7. We will not be presenting models that belong to categories (c) and (d) because: BTOM models (Baker et al. (2011)) use POMDPs to model beliefs, and are therefore not easily applicable in our setting due to infinite state space (all possible mixed strategies); and Level-k models (Wright & Leyton-Brown (2014)) have only been used to predict actions in simultaneous-move games and it is non-trivial to adapt to our belief updating setting in repeated SSGs.

Earlier work on belief modeling in categories (a) and (b) can be broadly classified into two types based on the assumption about the amount of information available. First is the case when no prior data is available to learn about the belief formation and update process of human agents in a given

Data avaiable for inference	Rationality	Proposed(P)/ Existing(E)	Model Name	Section #
No training data	Perfect (Bayesian)	E	B_u	4.1
		P	B_i^s (Uninformed adversary)	4.2.1
			${}^{I}B_u$ (Informed adversary)	4.2.2
	Bounded (Heuristic)	E	${}_{0.6}M_u{}^{A}$	4.1
		P	${}_{exp}M_u{}^{E}$, ${}_{hyp}M_u{}^{E}$, ${}_{lin}M_u{}^{E}$, ${}_{lin}M_{\{u,p\}}{}^{E}$	4.2
Training data	Perfect (Bayesian)	E	---	---
		P	$B_{learn}{}^{s}$	5
	Bounded (Heuristic)	E	${}_{learn}log_u$	5.1
		P	${}_{learn}M_u{}^{E}$, ${}_{learn}M_{\{u,p\}}{}^{E}$, B-REACT$_{c=4}{}^{wt}$	5.2
Training and Test data	Perfect (Bayesian)	E	---	---
		P	---	---
	Bounded (Heuristic)	E	---	---
		P	IBL$_k$, B-REACT$_c{}^{k=1}$, ^{best}B-REACT$_c{}^{k=1}$	6.1

Figure 3. Model names and assumptions

situation. This is what has been used in SSGs. Second is the scenario when historical belief update data for a group of human agents is available (training set). This facilitates learning a generalized model of human belief formation and update, and apply the learned model to predict belief updates for a previously unknown set of human agents (testing set). The assumption about having access to training data is common in the psychology literature and we adapt one popular model from that literature to SSGs. In this paper, we will also discuss another setting where, in addition to the training data about a group of participants, we will use information about the previously unseen (test set) participants' past beliefs (when available) to predict their future beliefs. Fig. 3 provides a summary of the models presented in this paper along with the corresponding assumptions.

4. Setting without training data

Here we discuss models for the situation where we do not have any training data to learn the adversary's belief update procedure.

4.1 Existing work

In the absence of any training data, previous work on modeling adversary beliefs in SSGs has focused on two aspects depending on assumptions about the adversary's rationality: (i) Bayesian update models typically associated with a perfectly rational adversary (An et al. (2012)); and (ii) heuristic belief update associated with "boundedly rational adversaries" (Pita et al. (2010)).

Perfectly Rational Adversary: An et al. (2012) proposed a Stackelberg Game with Limited Observation (SGLS) model where a perfectly rational adversary updates his beliefs about the defender's actual mixed strategy x given his prior beliefs and τ observations, where each observation is one of the defender's pure strategies $j \in \mathcal{P}$. They represent the sequence of observations compactly in terms of an observation vector $\mathcal{O}^r =< o_j^r >$ in which o_j^r is the number of times pure strategy j is observed until day r. They represented the adversary's belief distribution over the set of all pure strategies as Dirichlet distributions characterized by a parameter vector $\alpha =< \alpha_1, ..., \alpha_{|\mathcal{P}|} >$. They

assumed uniform Dirichlet distribution as prior. Then they use Bayesian updates to compute the posterior belief distribution over pure strategies based on the observation vector $\mathcal{O}^r$. For example, assuming $\alpha_k = 0; \forall k = 1to|\mathcal{P}|$ before day 1, and then after 5 days (one observation per day) we have observed pure strategy $j \in \mathcal{P}$ three times, then $o_j^5 = 3$ and the posterior $\alpha_j + o_j^5 = 0 + 3 = 3$ at the end of day 5. The adversary's belief b_i^r about the marginal coverage of target i after the r^{th} observation could then be computed from the posterior belief distribution over pure strategies as in Eqn. 3. In our experimental setting, with 9 targets and 3 defenders, $|\mathcal{P}| = \binom{9}{3} = 84$, and τ=10 (total number of days of observations). $j_i = 1$ (or 0) depending on whether target i is protected in pure strategy j (or not). This model will be referred to as B_u, where B represents Bayesian models and u stands for uniform prior.

$$b_i^r = \frac{\sum\limits_{j \in \mathcal{P}} j_i(\alpha_j + o_j^r + 1)}{\sum\limits_{j \in \mathcal{P}} \alpha_j + |\mathcal{P}| + \tau} \tag{3}$$

Boundedly Rational Adversary: Pita et al. (2010) proposed a linear mixture model to account for the belief update of boundedly rational adversaries. They model the adversary's beliefs b based on a weighted linear combination of two components: a prior belief ρ and the actual mixed strategy x. For any target i, this is shown in Eqn. 4. They assume the prior to be an uniform distribution of the number of defenders over the given set of targets. They further assume a fixed weight ($\mu \in [0, 1]$) on the prior for their experimental setting and do not provide any justification for their choice of the fixed weight. So, given 9 targets and 3 defenders, ρ_i at any target i is $\frac{300}{9} \approx 33$. If x_i at some target i is 50 and μ is 0.60, then the adversary's belief b_i (in percentage) about the defender's coverage at target i is $0.60 * 33 + (1 - 0.60) * 50 \approx 40$. We will refer to this model as ${}_{0.6}M_u^A$, where M denotes mixture models, A represents actual mixed strategy and 0.6 is the fixed weight on the prior.

$$b_i = \mu * \rho_i + (1 - \mu) * x_i \tag{4}$$

4.2 Proposed Models

For this setting where we have no training data, we first applied the above models and observed that these models perform poorly in terms of predicting beliefs of the adversaries. Therefore, we developed new models assuming both perfectly rational and boundedly rational adversaries that improve the state-of-the-art by providing new methods for (a) prior initialization and (b) the updating scheme. Performance results for all models proposed in this section are reported in Sec. 7.1.

4.2.1 Perfectly Rational Adversary

In this section, we consider two scenarios for modeling perfectly rational adversaries that make different assumptions about the amount of information the adversary may have about the strategies the defender is employing. In the first setting, we assume that the adversary knows nothing about the possible set of defender strategies. In the second setting, the adversary knows a set of candidate strategies of size $|\Theta|$ (=4 in our case) the defender may employ but does not know which strategy among this set the defender chooses to implement. In our experiments, this candidate set is composed of Maximin, SUQR, Uniform and Proportional strategies. The motivation for the second

scenario is that there may be an inside informant on the defender side who has secretly revealed this information to the adversary, and therefore we were interested in investigating the performance of a belief prediction model that accounts for this.

Uninformed Adversary: In the existing belief update model for a perfectly rational adversary in SSGs discussed in Section 4.1 , the adversary has no information about the types of strategies the defender may deploy. In their model, An et al. (2012) further assume that the adversary (a) starts from a uniform Dirichlet prior and (b) only updates the prior corresponding to the observed pure strategies. We relax these assumptions and improve the existing approach by proposing an informative Dirichlet prior based on domain features and a similarity based updating mechanism.

We hypothesize that instead of starting from a uniform Dirichlet prior the adversary may start with an informative Dirichlet prior based on the features of the domain. Intuitively, in our game, since animal density is the most important factor in determining defender allocations in the wildlife crime domain, we compute an informative Dirichlet prior which puts prior values on each pure strategy in proportion to the sum of the animal densities at the targets protected by that pure strategy. The intuition behind our novel updating method is generalizing our observations about pure strategies employed by the defender to other, similar pure strategies so that a more informed updated belief can be generated even after making limited pure strategy observations. In this work, we say that two pure strategies are similar if they differ in terms of defender allocation in only one of the three protected targets. For example, in Eqn. 3, if a pure strategy $j \in \mathcal{P}$ is observed three times in 5 days, then not only is $\alpha_j + o_j^5 = 3$, but also $\alpha_k + o_k^5 = 3$ for pure strategy k which was never observed during the 5-day timeframe but is similar to pure strategy j. This model will be referred to as B_i^s, where s denotes the similarity based updating procedure and i denotes informative prior.

Informed Adversary: Let us denote the set of mixed strategies that the defender chooses from as $\Theta =< \theta_1, \theta_2, ..., \theta_{|\Theta|} >$. For the case where the adversary has complete knowledge that the defender is deploying one of these $|\Theta|$ different mixed strategies, a perfectly rational adversary will perform Bayesian updates on their belief distribution over these strategies (represented as $\xi =< \xi_1, \xi_2, ..., \xi_{|\Theta|} >$) based on the sequence of pure strategy observations. The updated probability for the k^{th} mixed strategy θ_k (a vector denoting the coverage probabilities over all the targets) after observing the pure strategy on day r, denoted as ξ_k^r is computed using Eqn. 5 , where S_r denotes the set of all targets protected in pure strategy observation on day r, and x_i^k denotes the coverage probability at target i for the kth mixed strategy. His belief of the defender's mixed strategy after observing pure strategy on day r (denoted as b^r, which is a vector denoting the beliefs over all the targets) can then be computed as a weighted average of all the mixed strategies, where the weights are the updated probabilities (Eqn. 6). We denote this model as ${}^I B_u$, where I denotes informed adversary and u indicates that we start with a uniform prior over the set of mixed strategies.

$$\xi_k^r = \frac{\xi_k^{r-1} * \prod_{i \in S_r} x_i^k}{\sum_k (\xi_k^{r-1} * \prod_{i \in S_r} x_i^k)} \qquad (5) \qquad\qquad b^r = \frac{\sum_k (\xi_k^r * \theta_k)}{\sum_k (\xi_k^r)} \qquad (6)$$

4.2.2 Boundedly Rational Adversary

The previously proposed belief update model for boundedly rational adversaries in an SSG setting with no training data combines the actual mixed strategy of the defender with a uniform prior belief about the coverage at each target by keeping a fixed weight on the prior for all days. The existing

model does not offer a way to weight the prior beliefs over days of the game. The construction of the model has a number of shortcomings that may explain why the existing model performs poorly in experiments (see Sec. 7.1).

First, the adversary would only observe the defender's pure strategies and *not* know the exact mixed strategy. Therefore, he can only reason based on the empirical probability distribution of protection at each target. Second, he can have non-uniform prior beliefs about the coverage probabilities. Finally, an exploration of different weighting methods is necessary as the adversary can have any arbitrary weighting function for the prior weights over days of the game. None of these has ever been taken into consideration in existing work in SSGs. Our contributions here are to address these shortcomings and improve the state-of-the-art belief model for boundedly rational adversaries.

First, we incorporate in the existing model (Eqn. 4) the empirical mixed strategy (instead of actual mixed strategy) of the defender computed using all the pure strategy observations till the current day under consideration. So, when reasoning about the adversary's beliefs for day i, our model (Eqn. 7) would compute the empirical strategy (x^E) based on all pure strategy observations until day i. Second, since we assume that the defender has no prior training data about belief updates, it is not possible to learn about the belief update patterns of humans in this scenario. Therefore, instead of learning a function of how the adversary's reliance on his prior beliefs changes over days, we experiment with three different types of discounting functions and compare their performances: (a) linear, (b) hyperbolic, and (c) exponential. We chose hyperbolic and exponential since these are the most popular discounting methods in the literature (Samuelson (1937); Farmer & Geanakoplos (2009)). These models are denoted as ${}_{lin}M_u^E$, ${}_{hyp}M_u^E$ and ${}_{exp}M_u^E$, where u denotes uniform prior, E denotes that empirical strategy, and lin, hyp and exp denote linear, hyperbolic and exponential discounting functions respectively. Linear discounting based mixture model is shown in Eqn. 7. Similarly for hyperbolic (Eqn. 8) and exponential discounting (Eqn. 9).

$$b = \mu^{lin} * \rho_u + (1 - \mu^{lin}) * x^E \tag{7}$$

$$b = \mu^{hyp} * \rho_u + (1 - \mu^{hyp}) * x^E \tag{8}$$

$$b = \mu^{exp} * \rho_u + (1 - \mu^{exp}) * x^E \tag{9}$$

$\mu =< \mu_1, \mu_2, ..., \mu_\tau >$ denotes the weight on the prior for each of the τ days of observations. In terms of the ith day of the game, μ^{hyp} and μ^{exp} are computed as in Eqns. 10 and 11 respectively.

$$\mu_i^{hyp} = \frac{1}{i} \tag{10}$$

$$\mu_i^{exp} = \frac{1}{\exp(i-1)} \tag{11}$$

Although the above models assume a uniform prior, we observed during our analysis that not all participants start with a prior close to the uniform prior; in fact some participants start with a prior similar to the proportional prior and then update their beliefs. Since it is unknown which category of prior belief a previously unseen adversary would belong to, we apply a model ${}_{linear}M_{\{u,p\}}^E$ shown in Eqn. 12 that uses a weighted (weight=β) combination of uniform and proportional strategies as the prior (Eqn. 13). Due to absence of data to learn from, we assume β=0.5 in our experiments.

$$b = \mu^{lin} * \rho_{comb} + (1 - \mu^{lin}) * x^E \tag{12}$$

$$\rho_{comb} = \beta * \rho_u + (1 - \beta) * \rho_p \tag{13}$$

5. Setting with training data

This section discusses models for the situation where we have training data to learn the adversary's belief update procedure. Although we have experimented with a learning version (denoted as B^s_{learn}, where we learn the adversary's prior belief and do similar pure strategy based updating) of the non-learning Bayesian update model B^i_s, we show in Sec. 7.2 that this model did not yield promising results. Therefore, we only discuss heuristic belief update models for the learning setting.

5.1 Existing work

One popular learning belief model in psychology is a non-linear mixture model called the log-odds model (See et al. (2006)) shown in Eqn. 14. This model computes the log of odds metric between an event F (in our setting it is the adversary's belief b_i that a target i is covered by the defender) and the alternate event A (the adversary's belief that a target is *not* covered by the defender, i.e., (1-b_i)).

$$ln\frac{b_i}{1-b_i} = a_1 + a_2 * ln\frac{n^i_F}{n^i_A} + a_3 * ln\frac{f^i(F)}{f^i(A)} \tag{14}$$

Here, n^i_F (n^i_A) represents the adversary's prior belief about the number of ways target i is protected (or not). Similarly, $f^i(F)$ and $f^i(A)$ represent the influence of the actual observations on the beliefs formed by the adversaries. $ln\frac{n^i_F}{n^i_A}$ and $ln\frac{f^i(F)}{f^i(A)}$ denote the influence of the adversary's prior beliefs and actual observations respectively on his future beliefs. Parameters a_1, a_2 and a_3 are learned by performing linear regression on the training data. We call this model ${}_{learn}log_u$.

5.2 Proposed Models

In order to explore the benefits of having prior training data on model performances, here we propose two types of learning models assuming boundedly rational adversaries: (i) linear mixture models; and (ii) clustering based models that exploit the heterogeneity in adversary behavior.

Linear Mixture Models: Given training data about belief formation and update from a set of participants, the defender can learn the weighting function for the prior that best fits the training data. Here, by best fit we mean that we compute the weight vector $\mu =< \mu_1, \mu_2, ..., \mu_\tau >$ (where τ is the number of days) that minimizes the average root mean squared error (rmse) between the model's predicted beliefs and those of the training set participants. Therefore, instead of using a fixed weighting function as in our proposed models in Section 4.2, we use model ${}_{learn}M^E_u$ shown in Eqn. 15. Consistent with our non-learning model with combined prior in Sec. 4.2, we propose and experiment with a learning variant ${}_{learn}M^E_{\{u,p\}}$ (Eqn. 16) where we learn β in Eqn. 13 along with μ.

$$b = \mu^{learn} * \rho_u + (1-\mu^{learn}) * x^E \tag{15}$$

$$b = \mu^{learn} * \rho^{learn}_{comb} + (1-\mu^{learn}) * x^E \tag{16}$$

Clustering based Models: During our analysis of the performances of different models on the belief data collected on AMT, we observed the following heterogeneous behavior among adversaries in terms of their belief formation and update process. Adversaries can be clustered into four distinct groups based on their belief updates: (a) participants who start from a uniform prior and then update

their beliefs by taking into account the empirical distribution, (b) participants who start from a proportional prior and then update their beliefs by taking into account the empirical distribution, (c) participants who only update based on the empirical distribution and start with no prior, and (d) participants whose updates have no clear pattern and could be termed as random players.

This observation inspired us to apply clustering techniques on the belief data of the training set participants, learn a separate model for each cluster and use the learned models to predict the beliefs of test set participants. We propose a weighted clustering based approach to model and predict beliefs of a heterogeneous population of adversaries. First, we perform c-means clustering on the 10 day belief data of the training set participants to determine the clusters. Once the clusters are generated, we learn our model ${}_{learn}M^{E}_{\{u,p\}}$ (Eqn. 16) for each of the c clusters. ${}_{learn}M^{E}_{\{u,p\}}$ was chosen as it performed best (see Section 7.2) among all previously discussed models, and it is also the most generalized mixture model presented. Next, we compute the model's predicted beliefs for any participant after observing pure strategy for day r as a weighted average of the predictions of each of the models: $b^r = \frac{\sum_{i=1}^{c} \gamma_i * {}_i b^r}{\sum_{i=1}^{c} \gamma_i}$. Here, $\gamma_i = N_c$ is the weight given to cluster i and is the number of training set participants that belong to that cluster. ${}_i b^r$ denotes the belief predicted for day r by the learned model ${}_{learn}M^{E}_{\{u,p\}}$ for cluster i. The intuition behind weighting each cluster's model with the number of participants in that cluster is that we assume that the test set participant distribution will be similar to the training set. So, we give higher importance to clusters containing higher number of participants, and vice versa. We will refer to this model as $B\text{-}REACT^{wt}_{c}$.

6. Setting with training and test data

In the setting studied in Section 5, training data collected from a group of participants are used to predict belief updates of a completely new set of participants in test set. In this section, we assume that in addition to the training data we also have some data collected from the participants in the test set (earlier days of belief updates), which are used to predict belief updates for the following days.

Instance based Learning Models: Instance-Based Learning Theory (IBLT) (Gonzalez et al. (2003)) is a popular model used in Cognitive Science that attempts to explain human decision making in dynamic tasks. Based on past data about various situations and actions of different agents in such situations, IBLT attempts to predict the behavior of an agent in some situation by reasoning about known actions of other agents in similar situations. We propose an IBL model for belief prediction of an unknown adversary T_m after observing pure strategy j^r on day r.

We assume that in addition to knowing beliefs over all days of a set of adversaries (training set), we also gain information about the beliefs of the previously unseen test set adversaries at the end of each day. This could be achieved by placing an informant or spy among the poachers who would provide the defender information about the poacher's day-to-day beliefs. This allows the defender to make future belief predictions about the test set adversaries using their leaked beliefs till the current day by reasoning about beliefs of similar adversaries that are in the training data. In order to achieve this task, our model first computes similarity between beliefs (until day $r-1$) of a test set adversary and beliefs (until day $r-1$) of all training set adversaries. We then choose the k most similar training set adversaries and compute the belief of test set adversary T_m upon observing the r^{th} pure strategy based on day r beliefs of the k most similar training set participants based on

Eqn. 17, where $\theta_i \equiv \frac{1}{d(i,T_m)^2}$ and $d(i, T_m)$ denotes the dissimilarity between the test set adversary T_m and its ith most similar training set adversary.

$$ {}^{T_m}b^r = \frac{\sum_{i=1}^{k} \theta_i * {}^{k}b^r}{\sum_{i=1}^{k} \theta_i} \tag{17} $$

We will refer to this as the IBL_k model, e.g., a model based on four nearest neighbors will be referred to as $IBL_{k=4}$. Comparison results for various values of k are shown in Section 7.3.

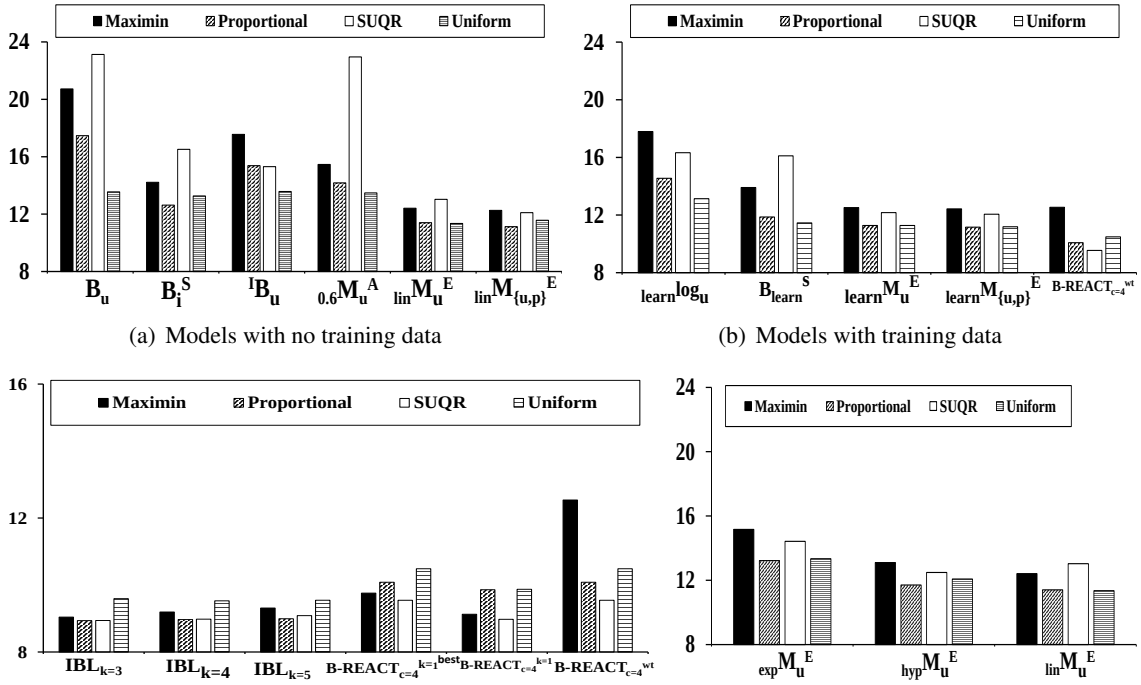
Clustering based Models: We customize our previously proposed clustering based model to take advantage of additional information about the test set participants (when available). We consider two scenarios of information availability: (a) before each day the defender has complete information about a test set adversary's beliefs till the previous day– this is same as the assumption for IBL models; and, (b) the defender knows the exact cluster a test set adversary belongs to.

For case (a), we compute for each test set adversary, the nearest (k=1) cluster he belongs to based on the known beliefs of that participant until day $i-1$ and apply the model for that cluster to predict his/her day i beliefs. This model is represented as $B\text{-}REACT_c^{k=1}$.

For case (b), since we assume that the exact cluster for each test set adversary is known to the defender, we apply the corresponding cluster's learned model to predict their beliefs for any day i. This is a somewhat unrealistic best-case scenario which gives us an important lower bound and therefore forms a baseline for comparing other models. In order to implement this, an important question is: how do we determine the exact cluster for a test set participant? In our game, since we have each participant's belief information for each of the 10 days, we assume the ideal scenario where we know the beliefs of all the 10 days for any test set participant ahead of time. This allows us to perform an exact nearest neighbor computation w.r.t. the c cluster centroids and determine the cluster for any test set adversary. The model is henceforth referred to as ${}^{best}B - REACT_c^{k=1}$ and its performance is shown in Sec. 7.3.

7. Experimental Results

In this section, we present results for existing and our proposed models (see Fig. 3 for all model names and their assumptions). We report the performance of all the models in estimating the beliefs of the test data set participants in terms of the average root mean squared errors (rmse) between the human entered beliefs and the models' predicted beliefs. The averaging is done over all targets for all days over the total number of participants in the respective test sets and over the total number of train-test splits. We show results on ADS_1 data in the paper. Results on ADS_2 have the same trends for all the models that we tested, thus confirming the value of our modeling and analysis. *In the figures, model names are on the x-axis and average rmse (lower is better) is on the y-axis. We start y-axis from 8 instead of 0 to show differences between the model performances more prominently.* The four defender strategies for which we conducted our experiments (Maximin, Proportional, SUQR and Uniform) are shown by the colored/patterned bars for each model in each of the figures. Any mention of statistical significance indicates that the discussed model performances are statistically significant based on two-tailed t-tests at confidence=0.05.



(a) Models with no training data

(b) Models with training data

(c) Models with training and testing data (max. Y-axis value is set to 16 so that differences are clearly visisble)

(d) Discounting Functions (Models with no training data)

Figure 4. Belief Estimation Errors (average RMSE)

7.1 Setting without training data

In Fig. 4(a) and 4(d) we first show performances for previously existing and our proposed models that do *not* learn on training data. We discuss important observations about these models below:

Comparison w.r.t. our best model: In Fig. 4(a), we demonstrate that our best performing non-learning model $_{lin}M_{u,p}^E$ completely outperforms (statistically significant) the two existing non-learning models (B_u and $_{0.6}M_u^A$) in SSGs in terms of predicting beliefs for any defender strategy. Furthermore, although for existing models, Maximin and SUQR are hardest to estimate due to their non-intuitiveness (as is evident by comparing their performances on Maximin and SUQR data against their performances on Proportional and Uniform data), our best model's performance on Maximin and SUQR defender strategies is similar to intuitive strategies such as Uniform and Proportional. Our model's performance further highlights the impact of using the empirical strategy instead of actual mixed strategy, a linear discounting function to capture the adversary's decreasing reliance on their prior beliefs, and a weighted combination of uniform and proportional prior so as to perform well against an unknown adversary who can belong to either one of these two groups.

Informed prior and similarity based updating improves performance: In Fig. 4(a), we observe that the performance of the previously existing Bayesian model B_u is significantly worse (20.73 to 16.11 for Maximin data) as compared to our proposed model B_i^s. This emphasizes the

benefit of starting with an informative prior and updating similar pure strategies when faced with limited observations. The informed rational adversary model ${}^{I}B_u$ which assumes that adversaries have prior knowledge about the set of defender mixed strategies, doesn't perform as well as the uninformed adversary models which make no such assumption.

Linear discounting performs best: We show in Fig. 4(d) that a simple linearly decreasing weighting function on the prior belief in the mixture models (${}_{lin}M_u^E$) surprisingly performs similarly or better when compared to models that consider more complex discounting functions such as hyperbolic (${}_{hyp}M_u^E$) and exponential (${}_{exp}M_u^E$). Results for ${}_{lin}M_u^E$ are statistically significant w.r.t. ${}_{exp}M_u^E$ for all strategies but only on Maximin and Uniform datasets w.r.t. ${}_{hyp}M_u^E$.

7.2 Setting with training data

We show performances for previously existing and our proposed learning models in Fig. 4(b).

Clustering significantly improves performance: First, we show that our proposed clustering based model $B\text{-}REACT_c^{wt}$ with $c = 4$ clusters outperforms (statistically significant) the existing learning model in the literature (${}_{learn}log_u$). Second, it also outperforms the best non-learning model ${}_{lin}M_{u,p}^E$ More importantly, it outperforms all other learning models (B_{learn}^s, ${}_{learn}M_u^E$ and ${}_{learn}M_{u,p}^E$). The significant difference between, the non-clustering model (${}_{learn}M_{u,p}^E$) and $B\text{-}REACT_{c=4}^{wt}$ which learns the same model but on different clusters, can be attributed to our earlier observation in Sec. 5.2 about the four distinct groups of adversary belief updates. This is also consistent with our observation about the weights learned for each model for each of the four clusters: (a) Cluster 1: μ^{learn} decreases almost linearly from 0.95 to 0.05 over 10 days, and the fixed weight on proportional prior ($1-\beta$ in Eqn. 13) is high (approx. 0.95 for most datasets), representing a group of adversaries who start with proportional prior and then linearly updates their reliance on the empirical strategy as they observe more pure strategies; (b) Cluster 2: Both μ^{learn} and β are 0, representing adversaries who only update based on the empirical strategy and do not start from any prior; (c) Cluster 3: It represents a group of adversaries who start with a uniform prior and then update their beliefs with more importance on their observations as days progress– the learned model has high β (approx. 0.97 for most datasets) and a μ that is high initially but gradually decreases (approx. 0.98 to 0.23); and, (d) Cluster 4: A high weight on β (approx. 0.97) and a μ that decreases from 0.90 to 0.50 (approx.), thus representing adversaries who start with a uniform prior and update at random on most of the days.

Learning Dirichlet prior improves performance: Learning a Dirichlet prior significantly improves the predictions of the resulting model (B_{learn}^s). For Proportional data, the average rmse for B_{learn}^s is 11.86 as opposed to 17.47 for the original model with no training data (B_u in Fig. 4(a)).

Learning weights in mixture models do *not* help: ${}_{learn}M_{u,p}^E$ is similar in performance to the best non-learning mixture model ${}_{lin}M_{u,p}^E$. This is a surprising observation, especially because we observed significant improvement in performances due to learning for perfectly rational adversary models (B_i^s vs B_u in Fig. 4(a)) . Further investigation reveals that the shape of the learned weighting function is approximately linearly decreasing for majority of the datasets, and hence the similar performance. Although surprising, this is a significant observation because it demonstrates that in the absence of data we could simply apply a linear decreasing weighting function irrespective of the deployed mixed strategy and expect to perform as well as if we had prior data to learn from.

Furthermore, this demonstrates that human adversaries have extremely strong initial biases towards a prior strategy in our game settings and they only linearly decrease their reliance on that bias over days of the game.

7.3 Setting with training and testing data

This section discusses results for models that use both training as well as additional information of test set adversaries to predict future beliefs. In Fig. 4(c) we compare the performances of such models against the best performing model discussed in the previous section ($B\text{-}REACT^{wt}_{c=4}$).

Testing set information does *not* help clustering models: $B\text{-}REACT^{k=1}_{c=4}$, a model that uses past beliefs of test set participants to infer their clusters, has similar performance to $B\text{-}REACT^{wt}_{c=4}$ which does not have this information. An ideal model that assumes complete knowledge about each test participant's exact cluster (${}^{best}B\text{-}REACT^{k=1}_{c=4}$ has rmse of 9.8 for Proportional data) shows improved performance over $B\text{-}REACT^{wt}_{c=4}$ which has an rmse of 10.08. However, the near similar performance of $B\text{-}REACT^{wt}_{c=4}$ to models that assume additional information about test set adversaries demonstrates the validity of our weighted clustering based technique towards making accurate predictions about adversary beliefs even in the absence of additional information.

IBL models perform best: IBL models outperform (rmse for $IBL_{k=3}$ is 8.93 for Proportional dataset) $B\text{-}REACT^{wt}_{c=4}$ (rmse of 10.08), $B\text{-}REACT^{k=1}_{c=4}$ and ${}^{best}B\text{-}REACT^{k=1}_{c=4}$ with statistical significance. This shows that, while clustering based models suffer from abstraction due to clustering, IBL models are able to make more personalized predictions **when information about past beliefs of test set participants is available**.

8. Conclusion

In this paper, we address the lack of empirical evaluation of belief formation models by conducting the first-of-its-kind systematic comparison of existing and new proposed models on belief data collected through human subjects experiments on AMT. We highlight three key observations. First, we observed surprisingly that a linear discounting function best fits adversary behavior in our setting (and it is also the weighting function learned), as opposed to more complex weighting functions, such as hyperbolic and exponential discounting. Second, we demonstrated the benefit of modeling heterogeneous groups of adversaries for improved belief prediction. We observed the presence of four different groups based on their belief formation and update procedure. Third, we show that our models significantly outperform existing models; the difference in performance further increases when using learning models in data-driven settings.

9. Acknowledgements

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

References

An, B., Kempe, D., Kiekintveld, C., Shieh, E., Singh, S., Tambe, M., & Vorobeychik, Y. (2012). Security games with limited surveillance. *Proceedings of the Twenty-Sixth AAAI Conference on*

Artificial Intelligence (AAAI).

Baker, C. L., Saxe, R. R., & Tenenbaum, J. B. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. *In Proceedings of the Thirtieth Third Annual Conference of the Cognitive Science Society*.

Fang, F., Nguyen, T. H., Pickles, R., Lam, W. Y., Clements, G. R., An, B., Singh, A., Tambe, M., & Lemieux, A. (2016). Deploying paws: Field optimization of the protection assistant for wildlife security. *Proceedings of the Twenty-Eighth Innovative Applications of Artificial Intelligence Conference (IAAI)*.

Farmer, J. D., & Geanakoplos, J. (2009). *Hyperbolic discounting is rational: Valuing the far future with uncertain discount rates*. Technical report, Cowles Foundation for Research in Economics, Yale University.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*, 591–635.

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

Kiekintveld, C., Jain, M., Tsai, J., Pita, J., Ordonez, F., & Tambe, M. (2009). Computing optimal randomized resource allocations for massive security games. International Foundation for Autonomous Agents and Multiagent Systems (AAMAS).

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

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

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

Pita, J., Jain, M., Tambe, M., Ordonez, F., & Kraus, S. (2010). Robust solutions to stackelberg games: Addressing bounded rationality and limited observations in human cognition. *Artificial Intelligence*, *174*, 1142–1171.

Samuelson, P. (1937). A note on measurement of utility. *Review of Economic Studies*, *4*, 155–161.

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

Tambe, M. (2011). *Security and game theory: Algorithms, deployed systems, lessons learned*. New York, NY: Cambridge University Press.

Wright, J. R., & Leyton-Brown, K. (2014). Level-0 meta-models for predicting human behavior in games. *Proceedings of the Fifteenth ACM Conference on Economics and Computation*.