TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 12 (2020) 992–1011

ISSN:1756-8765 online
DOI: 10.1111/tops.12513

This article is part of the topic "Best of Papers from the 17th International Conference on Cognitive Modeling," Terrence C. Stewart and Christopher Myers (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

Toward Personalized Deceptive Signaling for Cyber Defense Using Cognitive Models

Edward A. Cranford,[a] Cleotilde Gonzalez,[b] Palvi Aggarwal,[b] Sarah Cooney,[c] Milind Tambe,[d] Christian Lebiere[a]

[a]*Department of Psychology, Carnegie Mellon University*
[b]*Social and Decision Sciences Department, Carnegie Mellon University*
[c]*USC Center for AI in Society, University of Southern California*
[d]*Harvard Center for Research on Computation and Society, Harvard University*

Received 13 January 2020; received in revised form 6 May 2020; accepted 8 June 2020

Abstract

Recent research in cybersecurity has begun to develop active defense strategies using game-theoretic optimization of the allocation of limited defenses combined with deceptive signaling. These algorithms assume rational human behavior. However, human behavior in an online game designed to simulate an insider attack scenario shows that humans, playing the role of attackers, attack far more often than predicted under perfect rationality. We describe an instance-based learning cognitive model, built in ACT-R, that accurately predicts human performance and biases in the game. To improve defenses, we propose an adaptive method of signaling that uses the cognitive model to trace an individual's experience in real time. We discuss the results and implications of this adaptive signaling method for personalized defense.

Keywords: Cyber deception; Cognitive models; Instance-based learning; Knowledge-tracing; Model-tracing; ACT-R; Stackelberg security game; Signaling

Correspondence should be sent to Edward A. Cranford, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. E-mail: cranford@cmu.edu

1. Introduction

Cybersecurity often involves passive defense strategies which fail to discover a threat before major damage is done to a network. However, recent work within the domain of cybersecurity has focused on developing active defense strategies based on cognitive principles of deception (Cooney et al., 2019; Cranford et al., 2018; Huang & Zhu, 2019). Deception is a form of persuasion where one intentionally misleads an agent into a false belief, to gain an advantage over the agent and achieve one's goals (Rowe & Rushi, 2016). In this line of research, the goal for security is to assist human administrators to defend networks from cyber-attacks (Gonzalez, Ben-Asher, Oltramari, & Lebiere, 2014). Limited defense resources cannot simultaneously protect all targets. Therefore, in the event of an attack, truthful signals that divulge the protection status of a target can deter some attacks on protected targets. However, defenders can use a combination of truthful and deceptive signals to improve protection of the unprotected resources.

Game-theoretic principles have been employed to optimize the allocation of limited defense resources and determine how often to send a deceptive signal before it loses its effectiveness (Xu, Rabinovich, Dughmi, & Tambe, 2015). While deception may reduce attacks on uncovered targets compared to no deception, the algorithms are static and tailored to an entire population. They fail to take into account the individual and their particular set of knowledge, experiences, and biases. Such algorithms are easily learned and exploited by attackers. The goal of this paper is to develop a personalized signaling strategy that can outperform traditional static methods.

Cranford et al. (2018) developed an instance-based learning (IBL) cognitive model (Gonzalez, Lerch, & Lebiere, 2003) of attackers in a cybersecurity signaling game. This model accurately predicts human decision-making from experience. We propose that such a model can be used to trace an individual's knowledge and experiences, and exploit their biases, to determine on-the-fly the best signal given the situation, to further reduce attacks.

The following section presents a line of research on game-theoretic models that prove to optimize deceptive signaling for perfectly rational adversaries, and more recent efforts toward optimizing for boundedly rational adversaries. We then describe an online game developed to investigate attacker behavior against deceptive signaling algorithms and a cognitive model that accurately predicts attacker's behavior. Next, we describe a method for deceptive signaling that uses the cognitive model to drive adaptive signaling, personalized to the individual attacker. We highlight its applicability for optimizing defense by tracking human knowledge, experience, and biases. Finally, we discuss the implications of this line of research and avenues for future research.

2. Deceptive signaling for cybersecurity

Research on Stackelberg Security Games (SSGs) led to the development of algorithms that have greatly improved physical security systems (e.g., protecting ports, scheduling

air marshals, and mitigating poachers) through the optimal allocation of limited defense resources (Pita et al., 2008; Shieh et al., 2012; Sinha, Fang, An, Kiekintveld, & Tambe, 2018; Tambe, 2011). Xu et al. (2015) extended these models by incorporating elements of *signaling*, in which a defender (sender) strategically reveals information about their strategy to the attacker (receiver) to influence the attacker's decision-making (Battigalli, 2006; Cho & Kreps, 1987). Their solution, the Strong Stackelberg Equilibrium with Persuasion (peSSE), improves defender utility against a perfectly rational attacker compared to strategies that do not use signaling. For a given target, the peSSE finds the optimal combination of bluffing (sending a deceptive message that the target is covered when it is not) and truth-telling (sending a truthful message that the target is covered) so the attacker continues to believe the bluff.

The goal of the peSSE is to reduce attacks on uncovered targets. Attackers earn a reward for successful attacks, suffer a loss for failed attacks, and earn zero for withdrawing. When a target is covered, the peSSE will always send a truthful signal. When uncovered, the peSSE will send a deceptive signal with a probability that brings the attacker's expected value of attacking, given a signal, to zero. This makes it equal to the utility of withdrawing the attack and, based on standard game-theoretic assumptions of perfect rationality, the attacker will break ties in favor of the defender and withdraw.

The peSSE is suitable for cyber defense where optimizing the probability of sending a deceptive signal can mitigate attacks on uncovered targets with little overhead. However, it is based on the assumption of perfect rationality while humans exhibit, at best, bounded rationality (Simon, 1956). To address this weakness of the peSSE, researchers have begun to develop signaling algorithms for security against boundedly rational attackers (Cooney et al., 2019). However, these algorithms do not offer substantial improvement over the peSSE in terms of reducing attacks and minimizing defender loss. The main reason is that those algorithms assume rational behavior as specified by game theory optima such as Nash equilibria. However, human behavior systematically deviates from those theoretical descriptions. Human subjects exhibit learning curves—they do not generally compute the equilibria based on perfect knowledge of the interaction but rather they have to painstakingly accumulate the information through experience, then make satisficing decisions by limited cognitive means (e.g., Juvina, Saleem, Martin, Gonzalez, & Lebiere, 2013). Those deviations typically manifest themselves through systematic cognitive biases reflecting the interaction between limited cognitive mechanisms and the statistics of the task (e.g., Lebiere et al., 2013). Additionally, human behavior is dominated by individual differences in knowledge and capacity that manifest themselves into substantial variations in behavior (Lovett, Reder, & Lebiere, 1999).

To further address the weakness of peSSE, Gonzalez, Aggarwal, Cranford, and Lebiere (2020) proposed a research framework for dynamic, adaptive, and personalized deception for cyber defense. This framework implements SSG algorithms for distribution of limited defense resources with signaling theory (e.g., peSSE) to gain insights about human behavior from human-in-the-loop experiments, and cognitive modeling using instance-based learning theory (IBLT) to create personalized defense algorithms.

In what follows, we describe an IBL cognitive model that accurately predicts human attacker behavior playing against the peSSE in a laboratory experiment. We propose that a personalized deceptive signaling scheme based on insights from the IBL model, in combination with model-tracing mechanisms that record human actions and infer their knowledge (e.g., as used in cognitive tutors; Anderson, Corbett, Koedinger, & Pelletier, 1995), can be used to adapt defense signaling to the individual experiences of attackers at each point in time.

3. Cognitive models of human attackers playing against deceptive signaling algorithms

The insider attack game (IAG) was designed to investigate the interaction between a human attacker and defense algorithm in a cybersecurity scenario (Cranford et al., 2018). As shown in Fig. 1, players take the role of the attacker (a company employee) and their goal is to score points by "hacking" computers to steal proprietary data. There are six potential computers to attack, but only two security analysts (defenders controlled by a computer algorithm) that can monitor one computer each. If the player attacks a computer that is monitored, they lose points, but if the computer is not monitored, then they win points. Each computer shows its reward for winning, penalty for losing, and the probability that the computer is being monitored (reflecting the SSE for the game). On each turn, the player must select a computer to attack; after which, the signaling algorithm determines whether to send a truthful signal or a deceptive signal (with the signal, the player is presented the probability that the given signal is deceptive). The player must decide whether to continue their attack or withdraw and earn zero points. Players play four rounds of 25 trials each (after an initial five trials of practice). The payoff structures and monitoring probabilities of the targets are different in each round. Coverage and signaling of targets were precomputed for each trial. Therefore, each individual player experiences the same coverage and signaling schedule.

3.1. Attacker cognitive model

Cranford et al. (2018) developed an IBL cognitive model of the attacker using the ACT-R cognitive architecture (Anderson et al., 2004; Anderson & Lebiere, 1998). Following collection of human attack behavior against the peSSE defense algorithm, we modified the IBL model to more accurately represent human behavior playing the IAG. In accordance with IBLT, the model makes decisions by generalizing across past experiences, or instances, that are similar to the present situation. For the IAG, instances are represented by the contextual features of the selected target, the decision, and the outcome. The context includes the monitoring probability [0.0, 1.0], reward [1, 10], penalty values [−1, −10], and warning signal [present, absent]. The possible decisions are attack or withdraw, and the outcome is the reward or penalty based on the decision. In a given situation, for each possible decision, an associated utility (i.e., expected outcome) is

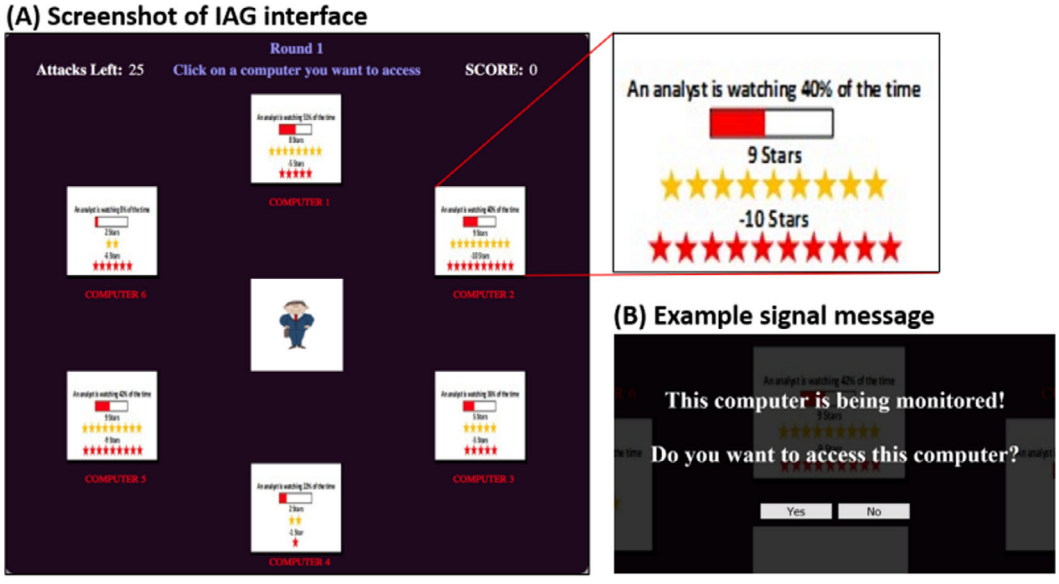

Fig. 1. (A) Screenshot of the insider attack game (IAG). The attacker is in the center surrounded by six targets. The zoomed inset shows that each target displays the monitoring probability (as a percentage in text and represented visually by red bars), the potential reward (represented by the yellow stars), and the potential penalty (represented by the red stars). (B) An example signal message that claims the computer is being monitored (if no signal is presented, the first line of the message is omitted).

computed through *blending*: an average across past outcomes weighted the probability of memory retrieval, which depends on contextual similarity to past instances. The decision with the highest expected outcome is executed. In the present game, there are two decisions: attack or withdraw. However, withdrawing always results in zero points. Therefore, the model only needs to determine the expected outcome of attacking to make a choice.

According to ACT-R's *blending* mechanism, the retrieval of past instances is based on the activation strength of the relevant instance in memory and its similarity to the current context. The activation A_i of an instance i is determined by the following equation:

$$A_i = \ln \sum_{j=1}^{n} t_j^{-d} + MP \times \sum_k \mathrm{Sim}(v_k, c_k) + \varepsilon_i \,. \tag{1}$$

The first term reflects the power law of practice and forgetting, where t_j is the time since the jth occurrence of instance i and d is the decay rate of each occurrence which is set to the default ACT-R value of 0.5. The second term is a partial matching process reflecting the similarity between the current context elements (C_k) and the corresponding context elements for the instance in memory (V_k), scaled by a mismatch penalty (MP; but which was set to the ACT-R default of 1.0). A variance parameter ϵ_i introduces stochasticity in retrieval and is a random value from a logistic distribution with a mean of zero

and variance parameter s of 0.25 (ACT-R default). Similarities between numeric slot values are computed on a linear scale from 0.0, an exact match, to −1.0. Symbolic values are either an exact match or maximally different, −2.5, a relatively large value which minimizes similarities between different actions and signal types.

A Boltzmann softmax equation determines the probability of retrieving an instance P_i based on its activation strength:

$$P_i = \frac{e^{A_i/t}}{\sum_j e^{A_j/t}}. \tag{2}$$

A temperature parameter t can be used to scale probabilities according to the activation such that low temperatures result in greater proportion assigned to the highest activated instances and high temperatures result in proportions being more randomly distributed regardless of activation strength. The current model set temperature to 1.0, which results in retrieval probabilities reflecting the original probability distribution, unbiased toward or against the most active instances.

The IBL model uses ACT-R's *blending* mechanism (Gonzalez et al., 2003; Lebiere, 1999) to generate an expected outcome of attacking a target based on similarity to past instances. The expected outcome is the value V that best satisfies the constraints of all matching instances i weighted by their probability of retrieval, where satisficing is defined as minimizing the dissimilarity between the consensus value V and the actual answer V_i contained in instance i:

$$\arg\min_V \sum_i P_i \times (1 - \mathrm{Sim}(V, V_i))^2. \tag{3}$$

When the values are numerical and the similarity function is linear, the process simplifies to a weighted average by the probability of retrieval $V_t = \sum_{i=1}^{n} P_i \times V_{it}$. Therefore, in summary, the outcomes of past instances are weighted by their recency, frequency, and similarity to the current instance (i.e., probability of memory retrieval) to produce an expected outcome via blending. After that expected outcome is generated, a straightforward decision rule is applied: If the value is greater than zero, then the model attacks; otherwise it withdraws.

3.2. IBL model procedure

To begin the IAG, the model is initialized with seven instances: Five represent a simulated practice round, and two represent knowledge gained from instructions (one instance had a signal value of *absent* and an outcome of 10, representing that attacking when a signal is absent will result in a reward; another instance had signal value of *present* and an outcome of 5, representing that attacking when a signal is present could result in either a penalty *or* a reward). To make a decision through blending on the first trial, rather than

making a random decision, the model requires an initial set of instances that bound the decision space and/or reflect the payoff expectations that human participants likely have knowledge of following practice (Gonzalez & Dutt, 2011; Lejarraga, Dutt, & Gonzalez, 2012); otherwise it would fail to retrieve anything. The current method provides the advantage of allowing each model run to begin with a unique set of experiences, much like the human participants, because the initial instances were randomly sampled from a uniform distribution of possible experiences during the practice round.

On each trial, the model first selects a target to attack as depicted in the left side of Fig. 2. The model cycles through each target and, for each, generates an expected outcome of attacking via blending. The model then selects the target with the highest expected outcome. Although humans tend to remember not only the actual experience but also their expectations prior to the experience (Gonzalez et al., 2003), we did not save the six instances generated during the target selection process. First, it is doubtful that players devote that much attention to the selection phase, and second we did not expect these instances to greatly affect future target selection decisions and the focus of the model is on the attack decision (i.e., understanding the effects of the signal on attack behavior).

As depicted in the right side of Fig. 2, after selecting a target, the context is augmented with the value of the signal (i.e., present or absent) and the model decides whether to attack or withdraw by generating a new expected outcome via blended retrieval. An early version of the model included both the signal and the target features in this decision, but it produced lower probabilities of attack on high monitored targets and high probabilities of attack on low monitored targets, whereas humans produced a more evenly

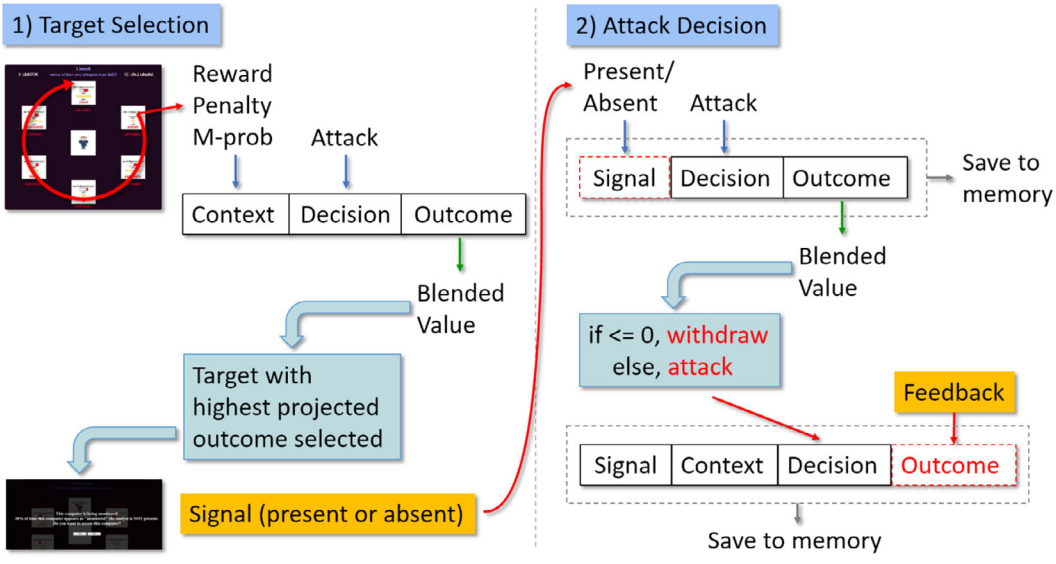

Fig. 2. Instance-based learning cognitive model procedure.

distributed probability of attack across targets. Additionally, because the signal message in the experimental interface occludes the targets on the screen, we inferred that humans base their decisions only on the value of the signal and ignore, forget, or otherwise do not use the occluded target information. Therefore, the similarities to past instances are based solely on the value of the signal (i.e., monitoring probability, reward, and penalty values are ignored) which produces a more evenly distributed probability of attack across targets.

After generating an expected outcome, a decision is made to either continue the attack or withdraw. According to IBLT's feedback process, the action and outcome slots of the current instance are updated to reflect the action taken by the model and the observed outcome (i.e., feedback is immediate). Although we did not save the instances of the target selection evaluations, during the attack decision we saved one instance that represents the model's expectation given the signal and another instance that represents the ground truth outcome. These two instances independently influence future decisions. The model continues for four rounds of 25 trials each and its behavior reflects its experiences. If an action results in a positive/negative outcome, then its future expectations will be increased/decreased, and the model will be more/less likely to select and attack that target in the future. Also, the impact of a particular past experience on future decisions strengthens with frequency and weakens with time. The stored expectations serve as a source of confirmation bias, in which one's preconception of winning/losing can increase the likelihood of attacking/withdrawing on future trials (i.e., generating positive/negative expected outcomes). In fact, a version of the model that did not store the expectations resulted in a mean attack rate of about 49%, which is far less than humans and the current model (~79% as shown below). For example, as depicted in left side of Fig. 3, when the expectations are not stored, a negative experience following a positive experience would likely lead to a subsequent withdraw action (the example just shows the raw average, but the point remains). However, when the expectations are stored, as depicted in the right side of Fig. 3, the model is more likely to persist in attacking, a pattern of behavior observed in this task as well as in previous paradigms of decision-making under risk and uncertainty (Erev et al., 2010).

3.3. IBL model evaluation against human players

The attacker IBL model was compared to human behavior in the IAG. In a laboratory experiment, human participants (i.e., "attackers") played against the peSSE signaling scheme. Participants were recruited via Amazon Mechanical Turk. All participants resided in the United States. For completing the experiment and submitting a completion code, participants were paid $1 plus $0.02 per point earned in the game, up to a maximum of $5.50. Four participants were removed from analysis because they had incomplete data (e.g., data recording errors) or restarted the experiment after gaining experience, resulting in a final sample size of 100.

The data were analyzed for the probability of attack and the number of points earned by attackers across rounds. The probability of attack was calculated as the mean

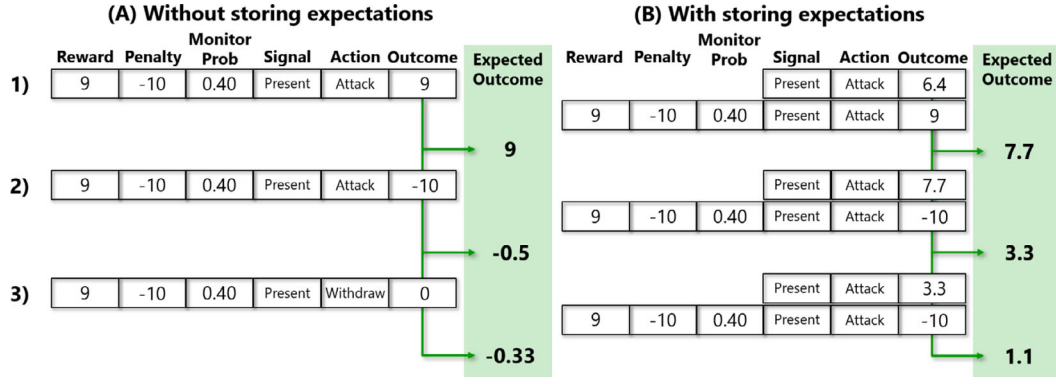

Fig. 3. Example blending calculation across a series of decisions for (A) a model that does not store expectations versus (B) a model that does store expectations.

proportion of attacks that were continued (as opposed to withdrawn) and was examined within players and across trials between players. Points were also separated into mean losses and gains per round. Losses/gains were calculated as the total number of points lost/gained per round by attacking targets that were/were not monitored.

The stochasticity of retrieval mechanisms in ACT-R led to differences in the expected outcomes generated for each run. After any given number of trials, each model run has accumulated a different set of experiences, notwithstanding the initial instances which are randomized between model runs. Therefore, the model played the IAG 1,000 times to generate stable predictions of the probability of attack and total number of points obtained per round. At the end of each run, the model was reset to its initial state and its memory cleared. Like humans, the model displays a range of behaviors which are based on the influence of unique individual experiences over time, and can therefore represent a diverse population of human attackers without the need to parameterize for individual differences.

Fig. 4, top left side, shows the mean probability of attack across trials and rounds for humans, black, compared to the model, gray. The dashed, gray line, horizontal at 0.33 on the *y*-axis, represents the predicted mean probability of attack according to the peSSE, under assumptions of perfect rationality. As indicated, both humans and the model attack far more often than predicted for a perfectly rational attacker. Furthermore, the model is an excellent predictor of human performance. RMSE and correlations, comparing the model to human data, are included at the bottom of the graph. The model is sensitive to the schedule of coverage, just as humans are, which produces the spiking pattern across trials (i.e., if a more popular target is signaled on a particular trial, then the mean probability of attack goes down, but if it is not signaled then the probability of attack goes up). The model not only matches well the mean probability of attack but also the full distribution of human behavior, as indicated by the histogram in the bottom left side of Fig. 4.

Fig. 4, right side, shows the mean points per round on the top and the average gains/losses on the bottom, for the humans compared to the model. Humans attack at a high

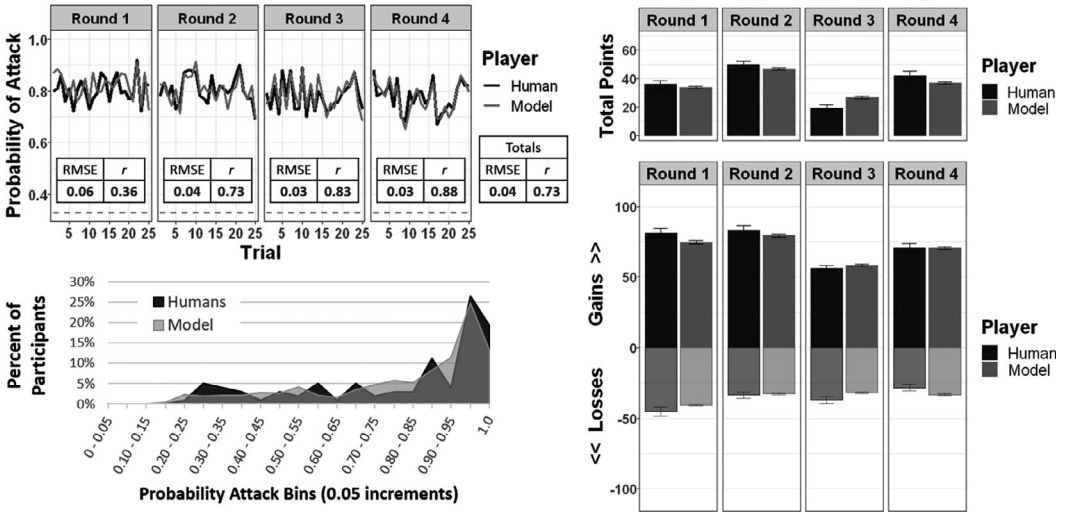

Fig. 4. Probability of attack across trials and rounds (left side, top) and distribution of mean total probability of attack for participants (left side, bottom), and mean points per round (right side, top) and gains/losses per round (right side, bottom) for the humans compared to the IBL model. For probability of attack, RMSE and correlations (r) between human and model data are displayed under each round, and the aggregate values across the entire game are on the right under the legend. For mean gains/losses, gains refer to points earned from attacking uncovered targets and losses refer to points lost from attacking covered targets.

rate, earning many points from attacks on uncovered targets (i.e., gains), while incurring fewer losses from attacks on covered targets (i.e., losses), resulting in an overall gain each round. Moreover, the model accurately predicts this behavior. The peSSE suffers because human biases (e.g., recency, frequency, and confirmation) lead them to attack at a higher rate, resulting in more experiences of wins than losses and thus a propensity to continue attacking. The IBL model captures these biases and, therefore, can feasibly be used as a predictive tool for personalizing deceptive signals for an individual attacker.

4. Toward personalized deception

To personalize deception, we can run the IBL model alongside the human to predict an individual's behavior and adjust the rate of deceptive signals to maximize belief in the signal and minimize the probability of attack. However, due to the stochastic nature of the cognitive model, its behavior would likely deviate further from the human's over time because the probability of randomly sampling the model run from the distribution in Fig. 4 that best matches the human is extremely low. Therefore, the predictive accuracy would likely be low, and thus the personalized signaling scheme would likely be ineffective. To make accurate predictions of an individual, two methods have proven useful to align the model behavior with the human's decisions: model-tracing and knowledge-tracing (Anderson, Boyle, & Yost, 1986; Anderson et al., 1995). Model-tracing aligns the

model's *actual* actions and outcomes to match those of the human, which can be overtly *observed* from the actions made and points earned. Knowledge-tracing, on the other hand, aligns the *expected* outcomes to match those of the human, but these values must be *inferred* because, based on the available data, we only know if the human's expectations were positive or negative prior to making a decision.

4.1. Model-tracing

Model-tracing is a method used to align a model's behavior with that of the human and is commonly used to adjust feedback provided to the student in intelligent tutoring systems (see Anderson et al., 1995). The alignment helps in a way that future model predictions are adapted and optimized to the interaction with the human. For example, geometry tutors use model-tracing to keep track of where errors are made so that the learning experience can be tailored to the individual (Anderson et al., 1986).

We use model-tracing to synchronize the IBL model with the human's *observed* actions and experience in the IAG task. After each trial, the instance saved in memory that represents the model's decision and outcome is changed to reflect the human's action and outcome (i.e., the action and outcome slots are changed to match the human's). Therefore, on the next trial, the model makes predictions based on the exact experience of the human and not on what it would have done based on its own past instances. With more trials, the model is expected to make more accurate predictions of a particular human's actions, as the model's memory aligns better with that of the human. Model-tracing changes the instances representing the *observed* ground truth decision and outcome. However, in order to generate accurate predictions, we must also align the model's expectations to those of the human.

4.2. Knowledge-tracing

While model-tracing relies on observed data, such as the actions humans take (attack or withdraw) and the outcomes obtained, some of the human's knowledge/experience cannot be observed and we must then *infer* what knowledge is likely stored in memory. The model produces instances that represent the expected outcome of attacking, which contributes to confirmation bias, and these must also be changed. *Knowledge-tracing* can be used to *infer* the expectations humans had prior to making a decision that would contribute to confirmation bias. For example, if the model and human both decided to attack (or both withdraw), then nothing need change and the expected outcome generated by the model can be used to infer the human's expectation. However, if the model expects a positive outcome for attacking, but the human withdrew the attack, then we can *infer* that the human expected to lose (or vice versa). For these instances, the expected outcome slot is modified to match the expectations of the player. This value cannot be inferred precisely, so as an estimate of the expected outcome we used either the reward or the penalty of the selected target, depending on whether the human expected to win or lose, respectively.

4.3. Model predictions with model- and knowledge-tracing

To test the effectiveness of model- and knowledge-tracing for predicting human decision-making, the model was run alongside human data in the peSSE condition. On each trial, the model simply makes a prediction, which is recorded and compared to the human's decision to generate a probability of agreement between the model and human. The model is then updated via model-tracing and knowledge-tracing, based on the human decision, to align the model's memory with that of the human for the next trial. The probability that the model attacks should therefore align with the humans, rather than deviate, and should do so consistently across attackers while producing highly accurate predictions for a particular individual. Fig. 5 shows the cumulative probability of agreement across trials between each human and the model (see the thin lines for a sense of the variance in individual agreement; darker lines indicate lower mean agreement). The mean cumulative probability of agreement, shown as the red line, for rounds 1–4 is 86.4% (SD = 12.3%), 90.8% (SD = 11.4%), 89.6% (SD = 12.4 %), and 86.8% (SD = 15.5%), respectively.

While predictive accuracy is more variable in the earlier trials when the model has few experiences on which to base its predictions, the overall trial-to-trial agreement is highly accurate. In fact, even at the first trial the model shows a mean agreement of 83.3%. Moreover, the model adapts well to the individual's probability of attack, becoming more accurate and stable as more data are gathered on the individual's decision-making. However, due to human stochasticity, there are some individuals that the model predicts less accurately, although the overall agreement is still higher than 60% as the model continues to adapt through round 4. Fig. 6 shows the overall probability of attack of individual model runs compared to the human it traced. The model is exceptionally accurate in adapting to the human, r^2 = .95. Using techniques of model-tracing and

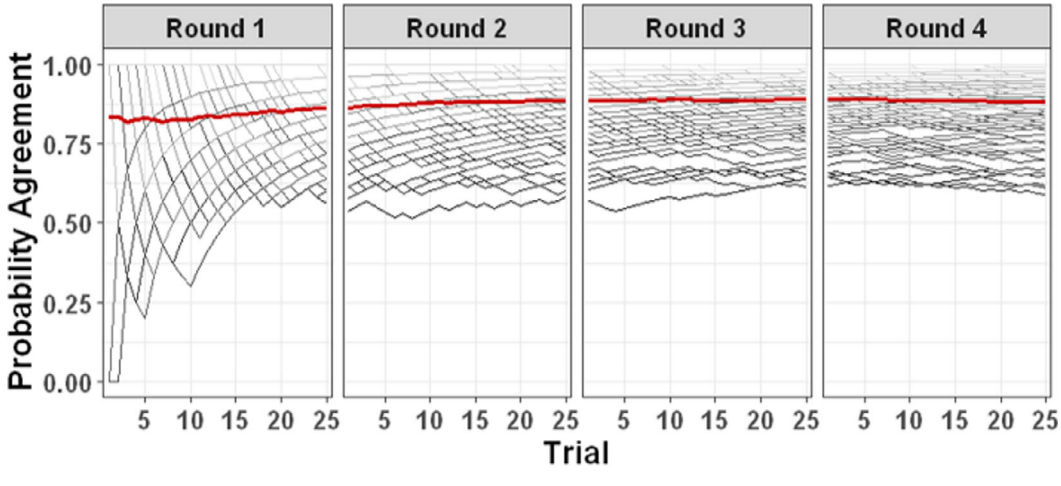

Fig. 5. Individual cumulative probability of agreement between the model and human across trials; darker lines indicate lower mean agreement. The red line shows the mean cumulative probability agreement across participants.

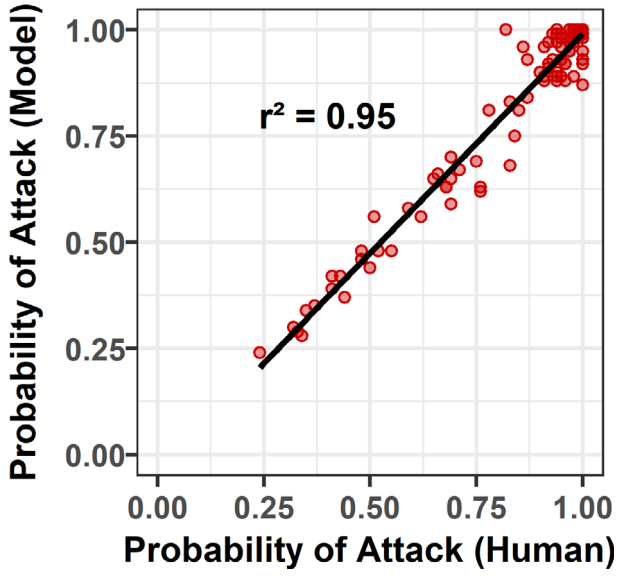

Fig. 6. Overall mean probability of attack comparing individual humans to the model run that traced them, in the peSSE condition using personalized signaling.

knowledge-tracing, the model makes very accurate predictions of the expected probability of attacking and could feasibly be used in designing a personalized signaling scheme.

5. A personalized deceptive signaling scheme

While the goal is to minimize the probability of attack, the peSSE signaling scheme uses deceptive signals on uncovered targets but not on covered targets. These schemes invite attacks with impunity when no signal is given. Therefore, a broader and more symmetrical approach may be warranted, as has been explored in recent game-theoretic research (Cooney et al., 2019). Using deception by *not* signaling when a target is covered reduces the overall frequency of signals, which is hypothesized to lead to fewer attacks when a signal is presented. Meanwhile, the possibility of losing when attacking given no signal should instill uncertainty and lead to fewer attacks. The following signaling scheme also uses deception when a target is covered.

Formally, one can write the goal of minimizing the probability of attack $P(A)$ given the presence of a warning signal (S) or its absence ($\bar{S}$) as:

$$\begin{aligned} \min_{W} P(A) &= P(S) \times P(A|S) + P(\bar{S}) \times P(A|\bar{S}) \\ &= P(A|\bar{S}) + P(S) \times [P(A|S) - P(A|\bar{S})]. \end{aligned} \tag{4}$$

Table 1
Expected impact of the presence or absence of a signal when a target is covered or not

		Target Coverage	
		Covered	Not Covered
Signal	Present	$P(A\|S)\ \searrow$	$P(A\|S)\ \nearrow$
	Absent	$P(A\|\bar{S})\ \searrow$	$P(A\|\bar{S})\ \nearrow$

One can then consider minimizing the probability of attack as a function of the probability of the warning signal, which yields the following equilibrium equation:

$$\begin{aligned}\frac{\partial P(A)}{\partial P(W)} = 0 \ &\Rightarrow P(A|S) - P(A|\bar{S}) = 0 \\ &\Rightarrow P(A|S) = P(A|\bar{S}).\end{aligned} \tag{5}$$

Thus, if the goal is to minimize the probability of attack as a function of the probability of the warning signal, then we must reach an equilibrium where the probability of attack given a signal is equal to the probability of attack given no signal. A signal must therefore be generated at a rate that preserves this equality. We can examine the impact of the presence or absence of a signal in various situations (see Table 1). Specifically, since no significant change occurs in the absence of an attack due to the lack of new information, Table 1 focuses on the change in future probability of attack resulting from the various circumstances of an attack.

For example, given an attack, if a target is covered, the attacker will lose, and their future probability of attack will be lower. If a target is uncovered, the attacker will win, and their future probability of attack will be higher. Each outcome thus increases or decreases one of the attack probabilities. In particular, the change in attack probability (decrease or increase) is determined by whether the selected target is covered or not, respectively, while the probability impacted (signal or no signal) is determined by the presence or absence of a signal, respectively. This results in the following algorithm for determining whether the signal S should be *present* or *absent*, depending if the selected target T is *covered* or *not covered*:

$$S = \begin{cases} \text{present}, & \text{if } T = \text{covered} & \text{and } P(A|S) > P(A|\bar{S}) \\ \text{absent}, & \text{if } T = \text{covered} & \text{and } P(A|S) \le P(A|\bar{S}) \\ \text{absent}, & \text{if } T = \text{not} - \text{covered} & \text{and } P(A|S) > P(A|\bar{S}) \\ \text{present}, & \text{if } T = \text{not} - \text{covered} & \text{and } P(A|S) \le P(A|\bar{S}) \end{cases} \tag{6}$$

The role of the cognitive model in this algorithm is to determine the components of the hypotheses of the conditional statements (i.e., the probability of attack given a signal is present or absent). We know the model generates expected outcomes of attacking E

and decides to attack if the value is greater than zero. Therefore, the following equivalency holds:

$$P(A|S) > P(A|\bar{S}) \Leftrightarrow E(A|S) > E(A|\bar{S}). \quad (7)$$

Thus, we can simply generate the expected outcome of attacking given the presence or absence of a signal and compare them to compute the conditions used in the algorithm above. An essential point is that those expected values are not the true expected values, but the model's subjective expected value given its limited experience and its reflection of human cognitive biases.

Intuitively, if the selected target is covered, then we decide on whether to generate a signal or not depending on which condition is most likely to lead to an attack. This corresponds to trying to catch the attacker when the target is covered, lowering the future probability of attack. Conversely, if the selected target is not covered, select the condition (signal or not) least likely to lead to an attack. Again, the accuracy of the cognitive model is essential in this approach to capture the subject's intention to attack or not. We can use the current model to track an individual's decisions and generate predictions of their probability of attack given the situation.

5.1. Effectiveness of personalized signaling scheme

To generate predictions of the effectiveness of this personalized signaling scheme, we ran the IBL model through the IAG while using the personalized signaling scheme described above to make predictions about the expected outcome of attacking, given a signal and given no signal. Based on those predictions and the underlying coverage of the selected target, the scheme determined whether to give a signal on each trial.

Fig. 7 shows the proportion of signals presented to participants as a frequency distribution. As can be seen, the personalized signaling scheme presents fewer signals on average than the peSSE scheme when a target is covered, while still presenting nearly the same proportion of signals when a target is uncovered. Overall, the personalized signaling scheme presents fewer signals while maintaining the rate of deception on uncovered targets. Interestingly, for some model runs, the personalized signaling scheme presented a large proportion of deceptive signals (75% or even higher).

Fig. 8 shows the frequency distribution of the probability of attack when a signal is presented or not when the target is covered or uncovered. As can be seen, when a signal is presented, compared to the peSSE the personalized signaling scheme is expected to maintain the probability of attack when a target is uncovered, and even slightly increase the probability of attack when a target is covered (a desired behavior if a goal is to catch an attacker). Likewise, when a signal is not presented, the personalized scheme is expected to reduce the probability of attack on uncovered targets.

Fig. 9, top left, shows the probability of attack across trials for the humans in the peSSE compared to the model predictions against the personalized signaling scheme. Compared to the human performance in peSSE, the personalized signaling method is

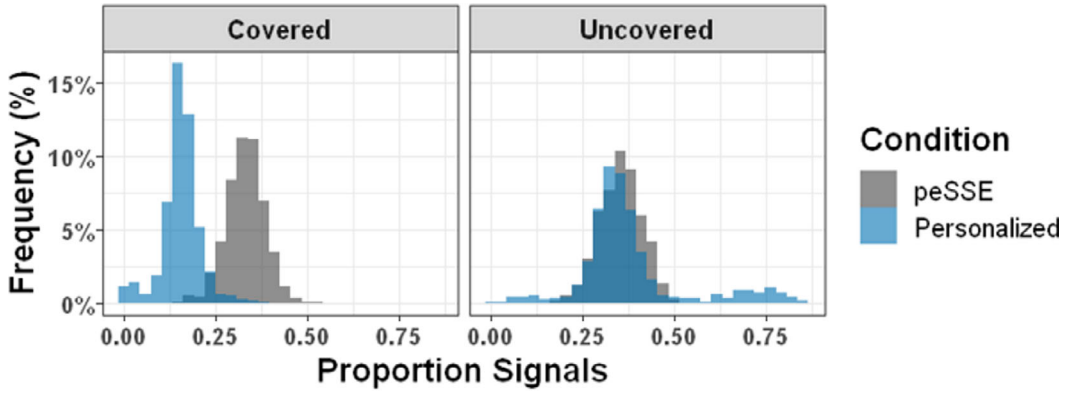

Fig. 7. Frequency distribution of the proportion of signals presented to humans against the peSSE compared to that of the model against the personalized signaling scheme.

expected to reduce the probability of attack by an average of 2.7% (RMSE = 6.6%). The relatively flat line for the personalized signaling scheme emerges because the signaling schedule is not static across participants, which controls for the variance seen in the peSSE. Although the expected reduction in probability of attack is modest, the expected gains in defender utility are valuable, as can be seen in the top right of Fig. 9. Defender utility is calculated as $-1 \times$ the number of attacks on uncovered targets. The improvement in defender utility is expected because, as the bottom right of Fig. 9 shows, the personalized signaling will result in fewer gains from attacks on uncovered targets and more losses from attacks on covered targets. While the effects look small, even modest gains

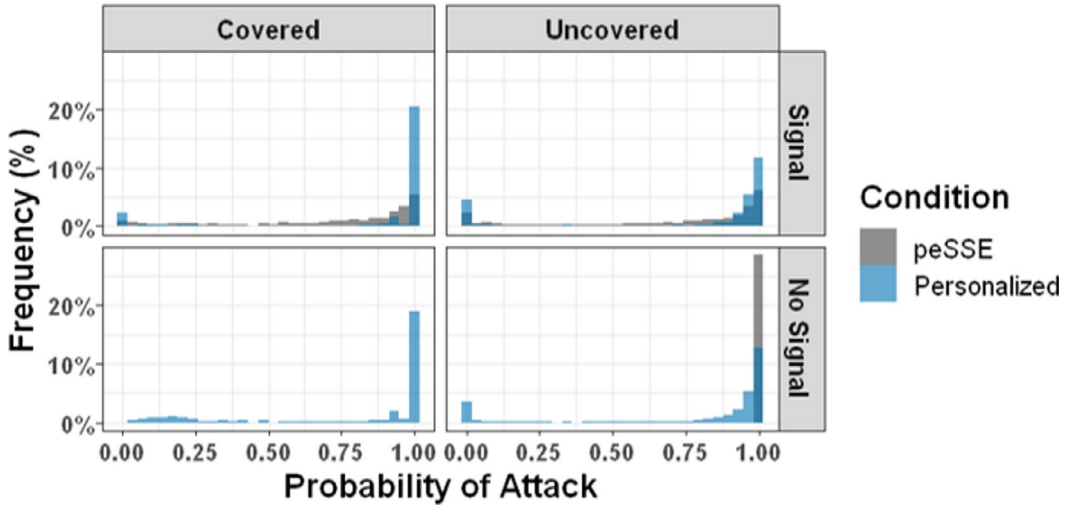

Fig. 8. Frequency distribution of the probability of attack for humans against the peSSE compared to that of the model against the personalized signaling scheme.

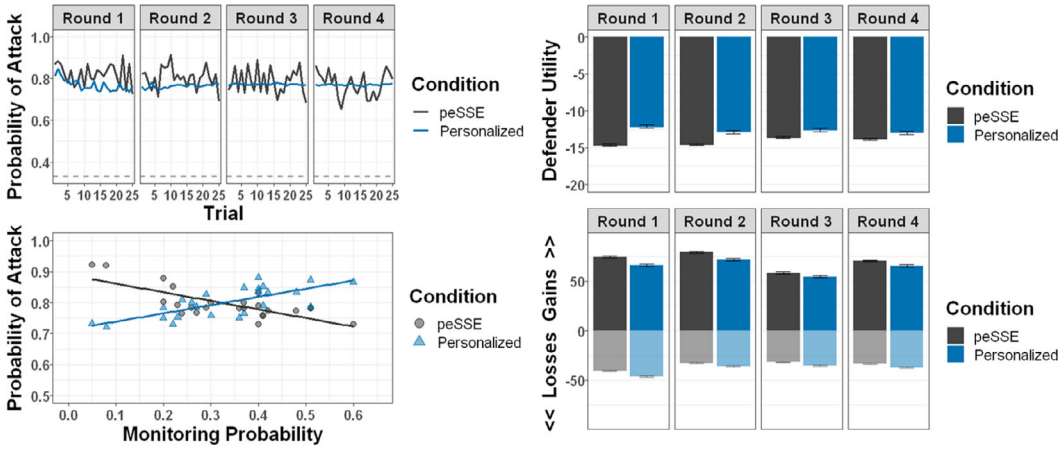

Fig. 9. Behavioral results for the humans against the peSSE compared to model predictions against the personalized signaling scheme. The top left shows the mean probability of attack across trials, the top right shows the mean expected defender utility across rounds, the bottom right shows the mean gains/losses across rounds, and the bottom left shows the mean probability of attack across targets, by their monitoring probability.

in defender utility and reduced attacks on uncovered targets are beneficial for improving defenses. Looking further into the data, the bottom left of Fig. 9 plots the probability of attack across the various targets, based on their monitoring probability. Compared to human performance, the personalized signaling method seems to shift the distribution of attacking toward targets with a higher monitoring probability, and this is why the IBL model incurs more penalties from attacks on covered targets while the overall probability of attack is only somewhat reduced.

6. Conclusions

The present research shows that we can leverage the predictive power of a generalizable IBL model to infer an individual's knowledge, trace their experience, and exploit their biases to design an adaptive signaling scheme that is personalized for an individual. Techniques of model-tracing and knowledge-tracing proved valuable to align the cognitive model with individual human decisions and improve the accuracy of model predictions of human behavior. As the results showed, the model predictions improve with time as more instances of human decisions are added to the cognitive model's declarative memory.

In agreement with IBLT (Gonzalez et al., 2003), human decisions are influenced by their prior expectations and their actual past experiences. The recency, frequency, and similarity of such instances contribute to the emergence of cognitive biases such as confirmation bias. As prior research showed, human decision-making in the IAG can be

explained through IBLT. The cognitive model indicates that the availability of positive experiences leads to a confirmation bias in which humans persist in attacking even after experiencing a loss. The personalized deceptive signaling scheme was designed to track the emergence of these biases and, therefore, exploit them in benefit of the defender.

The current method is an initial attempt toward developing a personalized deceptive signaling scheme for cyber defense. Although the simulations showed that the current scheme is not predicted to greatly reduce the probability of attack compared to the peSSE, it predicts a shift in attack behavior toward the targets that are highly monitored, resulting in more attacks on covered targets and fewer attacks on uncovered targets. Thus, overall defender utility showed an improvement compared to the peSSE. These effects were obtained even while reducing the total number of signals presented. Thus, we provide proof of concept of the model- and knowledge-tracing techniques that will be essential to generate personalized and adaptive signaling schemes.

Overall, the personalized signaling scheme we propose is predicted to improve defenses compared to static schemes such as the peSSE. The adaptive and personalized nature of the scheme allows for more accurate predictions of human behavior, allowing for greater optimization of deceptive signaling against an individual attacker. Of course, any personalized signaling scheme is limited by the ability to track human behavior; a particular problem in the cyber domain where it is difficult to identify an individual and attribute attacks to any one person. Future research will investigate the applicability of personalized signaling in more realistic situations. However, if an individual's behavior cannot be tracked, one can personalize the signaling scheme to a group of attackers or over a time period of attacks. Even such a scheme should show improvement in defenses compared to static schemes. Additionally, we also plan to adapt the coverage of targets according to attacker's behavior to further improve defenses (Cranford et al., 2020).

Most importantly, future research will test the personalized signaling scheme against human attackers. Insight gained from human experiments will provide information about how to modify the signaling logic to create a more effective scheme. Compared to other static signaling schemes that use deception by not providing signals sometimes when a target is covered, the current signaling scheme does not offer a substantial improvement in terms of reducing the probability of attack and increasing defender utility. However, we can expect that a combination of improved signaling logic and model- and knowledge-tracing techniques can further improve defenses. After all, if we can accurately predict human behavior, then an adaptive, tailored signaling scheme should prove better than a static scheme that makes idealistic assumptions about human behavior.

Acknowledgments

This research was supported by the Army Research Office and accomplished under grant number W911NF-17-1-0370.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036.

Anderson, J. R., Boyle, C. F., & Yost, G. (1986). The geometry tutor. *Journal of Mathematical Behavior*, *5*, 5–20.

Anderson, J. R., Corbett, A. T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of Learning Sciences*, *4*, 167–207. https://doi.org/10.1207/s15327809jls0402_2.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. New York: Psychology Press. https://doi.org/10.4324/9781315805696

Battigalli, P. (2006). Rationalization in signaling games: Theory and applications. *International Game Theory Review*, *8*(01), 67–93. https://doi.org/10.2139/ssrn.635244

Cho, I.-K., & Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, *102*(2), 179–221. https://doi.org/10.2307/1885060

Cooney, S., Vayanos, P., Nguyen, T. H., Gonzalez, C., Lebiere, C., Cranford, E. A., & Tambe, M. (2019). Warning time: Optimizing strategic signaling for security against boundedly rational adversaries. In E. Elkind & M. Veloso (Eds.), *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1892–1894). Richland, SC: IFAAMS.

Cranford, E. A., Gonzalez, C., Aggarwal, P., Cooney, S., Tambe, M., & Lebiere, C. (2020). Adaptive cyber deception: Cognitively informed signaling for cyber defense. In T. X. Bui (Ed.), *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 1885–1894). Maui, HI: ScholarSpace.

Cranford, E. A., Lebiere, C., Gonzalez, C., Cooney, S., Vayanos, P., & Tambe, M. (2018). Learning about cyber deception through simulations: Predictions of human decision making with deceptive signals in Stackelberg Security Games. In C. Kalish, M. Rau, J. Zhu & T. Rogers, *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 258–263). Madison, WI: Cognitive Science Society.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S., Hau, R., Hertwig, R., Stewart, T., West, R., & Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making*, *23*(1), 15–47. https://doi.org/10.1002/bdm.683

Gonzalez, C., Aggarwal, P., Cranford, E. A., & Lebiere, C. (2020). Design of dynamic and personalized deception: A research framework and new insights. In T. X. Bui (Ed.), *Proceedings of the 53rd Hawaii International Conference on system sciences* (pp. 1825–1834). Maui, HI: ScholarSpace.

Gonzalez, C., Ben-Asher, N., Oltramari, A., & Lebiere, C. (2014). Cognition and technology. In C. Kott, A. Wang, & R. Erbacher (Eds.), *Cyber defense and situational awareness* (Vol. 62, pp. 93–117). Cham, Switzerland: Springer International Publishing.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, *118*(4), 523–551. https://doi.org/10.1037/a0024558

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635. https://doi.org/10.1007/978-3-319-11391-3_6

Huang, L., & Zhu, Q. (2019). Dynamic Bayesian games for adversarial and defensive cyber deception. In E. Al-Shaer, J. Wei, K. W. Hamlen, & C. Wang (Eds.), *Autonomous cyber deception*(pp. 75–97). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-02110-8_5

Juvina, I., Saleem, M., Martin, J. M., Gonzalez, C., & Lebiere, C. (2013). Reciprocal trust mediates deep transfer of learning between games of strategic interaction. *Organizational Behavior and Human Decision Processes*, *120*(2), 206–215. https://doi.org/10.1016/j.obhdp.2012.09.004

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, *8*, 5–19. https://doi.org/10.1007/BF03354932

Lebiere, C., Pirolli, P., Thomson, R., Paik, J., Rutledge-Taylor, M., Staszewski, J., & Anderson, J. R. (2013). A Functional model of sensemaking in a neurocognitive architecture. *Computational Intelligence and Neuroscience*, *2013*, 1–29. https://doi.org/10.1155/2013/921695

Lejarraga, T., Dutt, V., & Gonzalez, C. (2012). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*, *25*(2), 143–153. https://doi.org/10.1002/bdm.722

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.008

Pita, J., Jain, M., Ordónez, F., Portway, C., Tambe, M., Western, C., & Kraus, S. (2008). ARMOR Security for Los Angeles International Airport. In D. Fox & C. P. Gomes (Eds.), *Proceeding of the twenty-third AAAI Conference on Artificial Intelligence* (pp. 1884–1885). Menlo Park, CA: AAAI Press.

Rowe, N. C., & Rrushi, J. (2016). *Introduction to cyberdeception*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-41187-3

Shieh, E., An, B., Yang, R., Tambe, M., Baldwin, C., & Meyer, G. (2012). Protect: A deployed game theoretic system to protect the ports of the United States. In W. van der Hoek & L. Padgham (Eds.), *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 13–20). Richland, SC: IFAAMS.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*(2), 129–138. https://doi.org/10.1037/h0042769

Sinha, A., Fang, F., An, B., Kiekintveld, C., & Tambe, M. (2018). Stackelberg security games: Looking beyond a decade of success. In J. Lang (Ed.), *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (pp. 5494–5501). Stockholm, Sweden: IJCAI.

Tambe, M. (2011). *Security and game theory: Algorithms, deployed systems, lessons learned*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511973031

Xu, H., Rabinovich, Z., Dughmi, S., & Tambe, M. (2015). Exploring information asymmetry in two-stage security games. In B. Bonet & S. Koenig (Eds.), *Proceedings of the twenty-ninth AAAI Conference on Artificial Intelligence* (Vol. 2, pp. 1057–1063). Palo Alto, CA: AAAI Press.