Cognitive Models of Dynamic Decisions in Autonomous Intelligent Cyber Defense

Baptiste Prebot Yinuo Du Xiaoli Xi
Cleotilde Gonzalez*

Dynamic Decision Making Laboratory
Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

To support the future cyber battlefield, many technological advances are required in the proposed architecture of autonomous intelligent cyber defense agents (AICA). In particular, AICA's decision making processes will need to be dynamic and adaptive to the actions of autonomous and intelligent attackers. AICA's decision-making algorithms must rely on theories that formalize dynamic decision processes; decisions made from experience; and the capabilities to learn, adapt, and react in the face of uncertainty. During the past decade, researchers of behavioral cyber security have created cognitive agents that are able to learn and make decisions in dynamic environments in ways that assimilate human decision processes. However, many of these efforts have been limited to simple detection tasks and represent basic cognitive functions rather than a whole set of cognitive capabilities required in dynamic cyber defense scenarios. Our current work aims at advancing the development of cognitive agents that learn and make defense dynamic decisions during the cyber attacks by intelligent attack agents. We also aim to evaluate the capability of these cognitive models in "Turing-like" experiments, comparing the decisions and performance of these agents against human cyber defenders. In this paper, we present an initial demonstration of a cognitive model of the defender, that relies on a cognitive theory of dynamic decision making, Instance-Based Learning Theory (IBLT); we also demonstrate the execution of the same defense task by human defenders. We rely on OpenAI Gym, CybORG, and adapt an existing CAGE scenario to generate a simulation experiment using an IBL defender. We also offer a new Interactive Defense Game (IDG), where *human* defenders can perform the same CAGE scenario simulated with the IBL model. Our results suggest that the IBL model makes decisions against two intelligent attack agents that are similar to

*Contact Author: coty@cmu.edu

those observed in a subsequent human experiment. We conclude with a description of the cognitive foundations required to build AICA that collaborate in autonomous cyber defense teams.

1 Introduction

The cyber battlefield of the future will certainly see autonomous systems fighting other autonomous systems (Kott, 2018). These autonomous systems, characterized by some level of freedom in decision making and action, will need to perform in uncertain and complex environments (David & Nielsen, 2016). In cyber-security in particular, the rise of autonomous intelligent malware represents a new threat that cyber-security experts and researchers must address quickly (Thanh & Zelinka, 2019). The global race towards more intelligent autonomous systems that are aware of the environment and the rapidly evolving cyber attacker capabilities presents cyber security researchers with two major challenges: (1) developing intelligent defense systems that are able to learn and understand the dynamic strategies of attackers to efficiently anticipate and counter their decisions, and (2) evaluating the capability of these intelligent defense systems to produce defense behaviors that are comparable to those of expert cyber defenders (Vieane et al., 2016; Dhir et al., 2021; Kott et al., 2020).

Human cognition and our ability to computationally represent the dynamic decision-making process of a cyber analyst are key for the future of cyber security (Gonzalez, Ben-Asher, Oltramari, & Lebiere, 2014; Kott et al., 2020). Cognitive models are dynamic and adaptable computational representations of cognitive structures (Gonzalez et al., 2014); and they can represent the human ability to adapt to changing environments and make decisions under uncertainty (Gonzalez, Lerch, & Lebiere, 2003; Gonzalez, 2022). To achieve fully autonomous cyber defense, a valid computational representation of such human abilities will be required in the future.

In the past decade, cognitive models have been developed in the context of cyber security to represent human defender decisions (Dutt, Ahn, & Gonzalez, 2011), human attacker decisions that can inform cyber defense strategies (E. Cranford et al., 2020; E. A. Cranford, Gonzalez, Aggarwal, Tambe, & Lebiere, 2020; Gonzalez, Aggarwal, Lebiere, & Cranford, 2020), and end-user phishing classification decisions that can help improve cyber defense (E. A. Cranford, Lebiere, Rajivan, Aggarwal, & Gonzalez, 2019; Xu, Singh, & Rajivan, n.d.). All these models are based on the well-known cognitive theory of dynamic decision making, Instance-Based Learning Theory (IBLT) (Gonzalez et al., 2003). Generally, IBLT is a comprehensive account of how humans make decisions based on experience during dynamic tasks, and has been used to represent the dynamic decision-making process in cyber security and many other domains (Gonzalez, 2022). IBL models of the cyber analysts date back to the work of (Dutt et al., 2011), who introduced a model of the recognition and comprehension processes of a security analyst in a simple cyber-attack scenario. The IBL model first recognizes cyber events (e.g., execution of a file on a server)

in the network based on the attributes of the situation and the similarity of the attributes of the events to past experiences (instances) stored in the memory of the analyst. Then, the model reasons about whether a sequence of observed events is a cyber attack or not, based upon instances retrieved from memory and the risk-tolerance of a simulated analyst. Execution of the IBL model generates predictions of the analyst's decisions that are evaluated based on their timeliness.

Although this model of the cyber defender is a good starting point in the development of AICA, it is unclear how well this model performs against autonomous intelligent attackers. The model in (Dutt et al., 2011) was not evaluated against human cyber-defenders, and the cyber-security scenario was quite simple, where an attacker would attempt the access of a company's server indirectly through a web server. In the current work, we advance the development of AICA by providing the following contributions: (1) present an IBL cognitive model of the dynamic decision process of cyber defense in a complex OpenAI gym, called CybORG (Baillie et al., 2020), and a challenging cyber attack scenario against two different attack strategies (Standen et al., 2021); (2) develop a new Interactive Defense Game (IDG) that integrates the attack scenario into an interactive tool where human defenders confront the same attackers in CybORG; (3) provide simulation results of the performance of the IBL model against the two attack strategies; and (4) evaluate the capabilities of the IBL models against human performance in the same scenarios using IDG. Our results suggest that the IBL model can make reasonable predictions against the two strategies of attack and that these predictions are similar to the behavior observed in human defenders.

2 A framework to integrate IBL models in autonomous cyber defense

In a reference architecture provided in (Kott et al., 2020), AICA allow a cyber defense agent (e.g., human or autonomous) to acquire data from the environment and the systems in which it operates, to reach an understanding of the current state of the environment and make effective dynamic decisions against intelligent cyber attackers. To achieve this vision, (Kott & Theron, 2020) distinguish 3 phases for AICA implementation. The first phase consists in developing AI models and compare their performance in suggesting remediation plans to those of experienced Human cyber defenders. The second phase would be to test the resilience of the AICA in more complex scenarios, including dynamic attack strategies. Finally, the third phase would concern multi-agent collaboration, human interactions, and ensuring the stealth and trustworthiness of the agent.

In this paper, we contribute to the advance of the first phase of the AICA vision by proposing an integration of a dynamic and adaptable computational representation of the cyber defense agent and the evaluation against the actions of intelligent attackers (Gonzalez et al., 2014). Figure 1 represents the idea of

integration of IBL models and AICA. The main component in the development of AICA is the generation of the IBL cognitive model of the dynamic decision process for cyber defense that can perform well in complex cyber security environments.

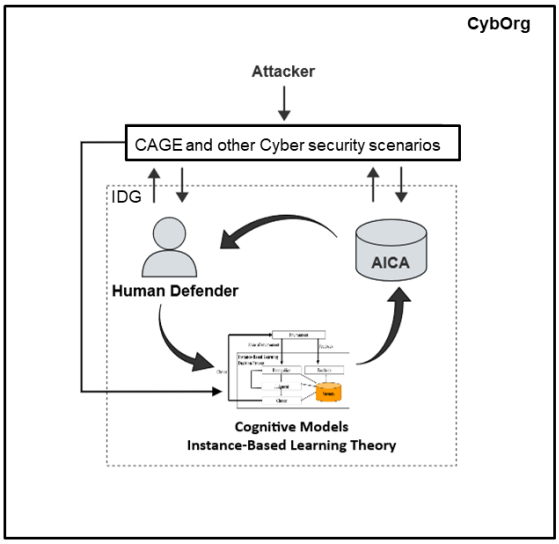

Figure 1: Integration of IBL models in AICA

Such cyber security environments involve both attackers (i.e., red agents) and defenders (i.e., blue agents) in scenarios that co-evolve dynamically. The characteristics of adversarial, evolving, dynamic, and highly complex environments suggest a need for learning and decision-making models that are trained and evaluated across a wide range of adversary behaviours and scenarios (Baillie et al., 2020). To develop adaptive cyber operations, we rely on a recent development of a gym, called CybORG (Baillie et al., 2020). CybORG is an adaptive cyber operations gym appropriate to test learning decision-making models; it is scalable and flexible to simulate adversarial scenarios. To evaluate our IBL defense model, we used a modified CybORG scenario named in a recent modeling competition, called the CAGE challenge (Standen et al., 2021). Furthermore, to evaluate the IBL model, we also propose Interactive Defense Game (IDG) that integrates the attack scenario into an interactive tool where human defenders confront the same intelligent attackers in CybORG, confronted by the IBL model of defense.

In what follows, we present the components of the framework that integrates the IBL models of defense in the AICA framework, and provide an evaluation of this approach.

2.1 CybORG: An Autonomous Cyber Operations Gym

To allow autonomous cyber operations to become an operational capability, researchers are developing AI gyms and platforms that encapsulate cyber elements in integrated environments. Such platforms, aim at enabling the study and development of autonomous adaptive defense agents, by confronting them to dynamic and realistic attack scenarios and network simulations. Our present work aims to exploit such platforms to investigate, develop, and test the IBL-AICA integration proposed above.

We use the CybORG AI gym (Baillie et al., 2020) and adapt the CAGE cyber defense scenario (Standen et al., 2021) to perform experimental simulations using IBL agents and humans as cyber defenders.

The CybORG AI Gym was initially developed and released as an experimental simulation platform to train reinforcement learning agents for cyber defense. It uses the OpenAI Gym interface (Brockman et al., 2016) along with a cyber operation adversarial scenario in a realistic yet simple training environment.

2.2 Adapted CAGE challenge scenario

For the present work, we adapted the Cage Challenge scenario (Standen et al., 2021) (i.e., CAGE) that was implemented in CybORG. In CAGE, a defense agent (blue agent) is tasked to defend a network against an attacker (red agent). Green agents can also be included to simulate normal activity generated by regular users performing *Scans* on a system. These three types of agents interact with the scenario alternatively by performing high-level actions in a game-like episode with a fixed number of steps.

Red agents can choose to do reconnaissance, exploitation, privilege escalation, and pivoting. Blue agents are enabled to conduct network monitoring, host analysis, malicious code removal, and system recovery from backups.

The flexibility of CybORG opens up a wide range of scenarios, and we took advantage of it to adapt the CAGE scenario that allows us to immerse agents (AI or humans) in situations of different complexity. From the CAGE scenario, we were able to manipulate different aspects of the simulation: the type of attacker, and the presence or absence of the 'normal' activity of the Green agent and the size of the network, as illustrated in Figure 2.

The type of attacker can represent different deterministic strategies. Two of them, *Beeline* and *Meander*, were provided in the initial Cage Challenge scenario, to resemble the diversity of attackers and to evaluate the capability of the defender agent more comprehensively. They differentiate by their prior knowledge of the network and the way they route through the network accordingly. Beeline can be assimilated to an agent performing a blitz, targeted attack, while Meander is seeking to gain privileged access on all hosts, stealthy establishing a long term presence in the network. Although these are two types of attackers provided in the CybORG scenario, other AI models of attackers can also be plugged in, to simulate more dynamic, nondeterministic, realistic attackers and to study the adaptation capacities of defenders.

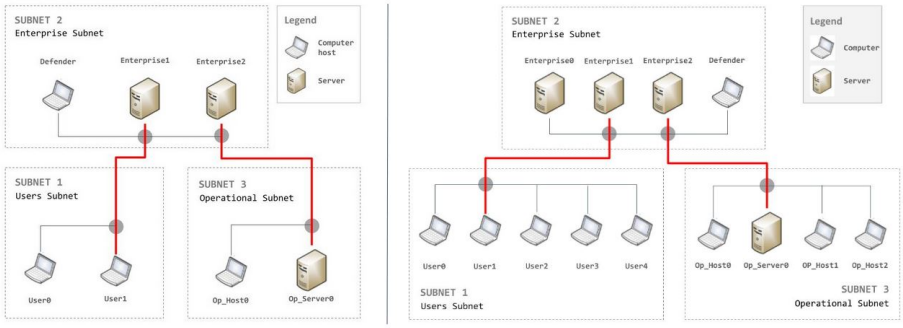

Figure 2: Example of two network sizes created in the CybORG scenario

Manipulating the Green agent allows one to study its impact on the defense strategy and to develop models that would be able to discriminate an attacker and its strategy, from the activity of a normal user.

The same goes for the size of the network. The number of subnets and hosts is a proxy of the complexity and realism of the simulation, and its impact on the performance of cyber defenders is studied.

On the defense side, the simulation environment allows for testing of multiple IBL agents, built with different instances structures or based on different cognitive models, as well as other types of AI agents, for example, RL agents, used to to compare their performance in predicting Human behavior. However, in order to do so, an interactive interface dedicated to Human testing, had to be developed.

In terms of experimental settings, the number of episodes to be simulated and their duration (i.e., the number of steps of each episode) can also be manipulated. For initial experiments, the scenario runs on 25 steps-long episodes on a small network.

2.3 Interactive Defense Game

We created a new Interactive Defense Game (IDG) which is a Django-based web application. IDG offers a web-based graphical user interface to allow human participants to perform the task proposed in our adapted CAGE scenario.

The IDG interface shown in Figure 3 consists of a central interactive table representation of the network and the related information on each host or server: IP Address, name, subnet, last detected activity, and compromised level. In this task, a human defender can select among a set of actions represented in buttons on the bottom right of the screen: *Monitor, Analyze, Remove, Restore.*

Human defenders can select a host by clicking on its row in the table and then choose one of the four actions to perform on that particular host. Then by clicking on the "Next" button, the action selected takes effect, and the defender can see the result (i.e. points lost) from the execution of that action in the "Last

round" value. A new and updated version of the environment is presented to the human defender, demonstrating the status (activity and compromised levels) of the network elements. In order to allow a defense strategy to emerge. The "Last round" outcome provides an immediate feedback regarding the effectiveness of the past action, and the "Total loss" presents the human defender with a cumulative account of the loss during the episode.

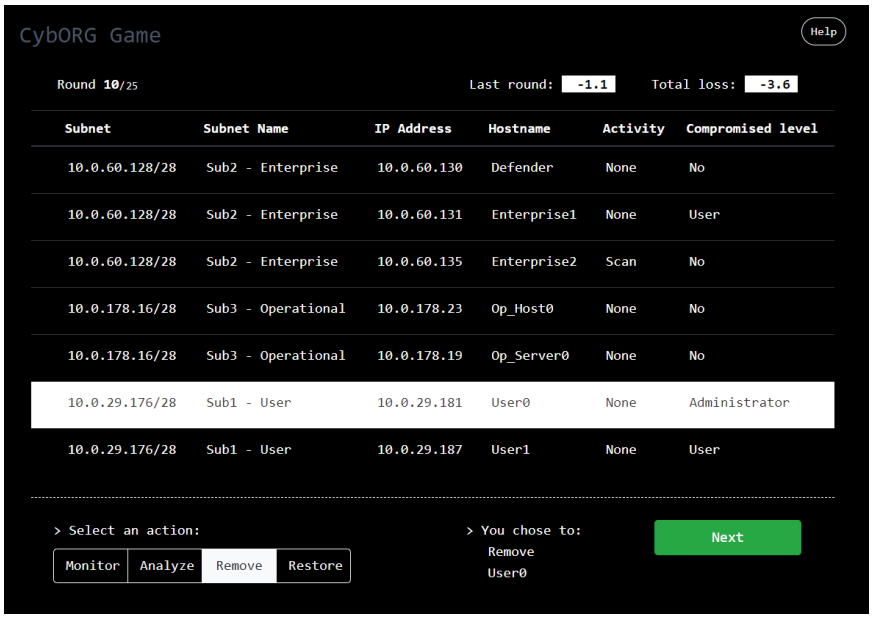

Figure 3: Interactive Defense Game user interface.

2.4 IBLT and IBL model of Blue Agents

IBLT is a well-known cognitive theory of decision making (Gonzalez et al., 2003). The key idea is that decisions are made by acknowledging similar past experiences and integrating them into the generation of the expected utility of the decision alternatives and the selection of the alternative with the maximal expected utility. An IBL model can accurately represent the contents of human memory, recognition, learning, and recall of experiences in decision making. The development of cognitive models for adaptive cyber defense, builds on our extensive work of models of cyberdefenders, cyberattackers, and end-users in a cybersecurity context (i.e., Phishing) (e.g., Gonzalez et al., 2020).

Although IBLT's process and mechanisms have been widely published, and they are common regardless of the particular task the models are developed for, we repeat the mathematical formulations of the theory here for completeness. In IBLT, an "instance" is a memory unit that results from the potential alternatives evaluated. These memory representations consist of three elements that are constructed over time: a situation state s that is composed of a set of

characteristics f; a decision or action a taken corresponding to an alternative in state s; and an expected utility or experienced result x of the action taken in a state. Concretely, for an IBL agent, an option $k = (s, a)$ is defined by the action a in the state s. At time t, assume that n_{kt} different instances (k_i, x_{ik_it}) are considered for $i = 1, ..., n_{kt}$, associated with k. Each instance i in memory has a value *Activation*, which represents the ease with which that information is available in memory (?, ?). Here, we consider a simplified version of the activation equation which only captures recency, frequency, and noise in memory.

$$\Lambda_{ik_it} = \ln \left(\sum_{t' \in T_{ik_it}} (t - t')^{-d} \right) + \sigma \ln \frac{1-\xi_{ik_it}}{\xi_{ik_it}}, \tag{1}$$

where d and σ are the decay and noise parameters, respectively, and $T_{ik_it} \subset \{0, ..., t-1\}$ is the set of the previous timestamps in which the instance i was observed. The rightmost term represents a noise for capturing individual variation in activation, and ξ_{ik_it} is a random number drawn from a uniform distribution $U(0, 1)$ at each step and for each instance and option.

The activation of an instance i is used to determine the probability of retrieving an instance from memory. The probability of an instance i is defined by a soft-max function:

$$P_{ik_it} = \frac{e^{\Lambda_{ik_it}/\tau}}{\sum_{j=1}^{n_{kt}} e^{\Lambda_{jk_jt}/\tau}}, \tag{2}$$

where τ is the Boltzmann constant (i.e., the "temperature") in the Boltzmann distribution. For simplicity, τ is often defined as a function of the same σ used in the activation equation $\tau = \sigma\sqrt{2}$.

The expected utility of option k is calculated based on *Blending* as specified in the choice tasks.

$$V_{kt} = \sum_{i=1}^{n_{kt}} P_{ik_it} x_{ik_it}. \tag{3}$$

The choice rule is to select the option that corresponds to the maximum blended value. When the agent receives delayed results, the agent updates expected utilities using a credit assignment mechanism (Nguyen, McDonald, & Gonzalez, 2021).

IBL Model of Blue Agents

We developed an IBL cognitive model of cyber defense and demonstrate the model's predictions in a simple but realistic scenario against two types of attack strategies (Beeline and Meander) and under various conditions of noise and feedback frequency regarding the attacker's actions (Du, Prebot, Xi, & Gonzalez, 2022).

The contextual features are constructed to resemble the information that would be presented to a human defender in the scenario. Specifically, there

are two slots for each host or server, representing the observed activity and the known compromised status of that host at a certain step in an episode. The order of *(Activity, Compromised Status)* pairs for each host is fixed to encode the identity of each host, that is, the *Host name*, *IP address* and *Subnet*. The *Step Index* slot is included to resemble the step counter within each episode. The decision is for the IBL agent to choose a host to protect and the tool to protect it with. Each action consists of a host and a command in the format of *cmd host*.

The model makes decisions by storing the instances in memory and following the general mechanisms of IBLT described above. The complete description of this model and the simulation results are presented in (Du et al., 2022). The results of the simulations with this model illustrate the expected impact on defense losses when attackers are more knowledgeable and directed in their attacks compared to when an attacker meanders around. Results show how losses increase in the presence of normal activity and how a defender can benefit from not knowing information too frequently to reduce losses.

3 Simulation and Human Experiments

We have built and used an IBL model that represents a human defender performing the adapted CAGE scenario in CybORG. The details of this IBL model and the setting for the simulation experiment are further described in (Du et al., 2022). In the simulation experiment, we designed four conditions involving different strategies of the red agent and the presence or absence of green agents. The Beeline Red agent without Green agents; the Beeline Red agent in the presence of Green agents; the Meander Red agent without Green agents; and the Meander Red agent in the presence of Green agents. We ran 40 IBL simulated defenders, each in 2000 episodes of 25 steps in each condition.

As an illustration, Figure 4 shows the performance and learning over the 2000 episodes for the Blue IBL agents, when confronted with the two Red agents, Beeline and Meander. These results are obtained in the absence of Green agents. The complete set of results of these simulations is shown in (Du et al., 2022). The results are shown in terms of the loss suffered by the IBL agents during the execution of the scenario in each episode. As observed, the IBL agents confronted with the Beeline Red agents show a larger loss initially but are able to learn over episodes. The IBL agents confronted against the Meander Red agents, suffer a less severe loss initially, and again are able to learn by reducing such loss over episodes.

A first experiment has been conducted involving human defenders performing the same scenario as in the simulation experiment. The human experiment involved participants who defend the small network shown in Figure 2, against one of the two types of attackers, Beeline or Meander, using IDG.

120 subjects carried out the task of defending the network against one of the two attackers, with the goal of minimizing the loss of defense points. In the scenario, there were no Green agents, each participant performed 7 episodes of

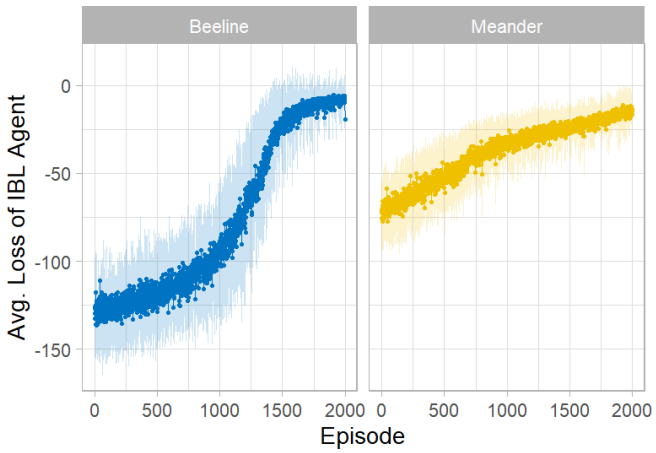

Figure 4: IBL agents average Loss against two different attackers strategies, without Green agent on a small network.

25 steps in each episode.

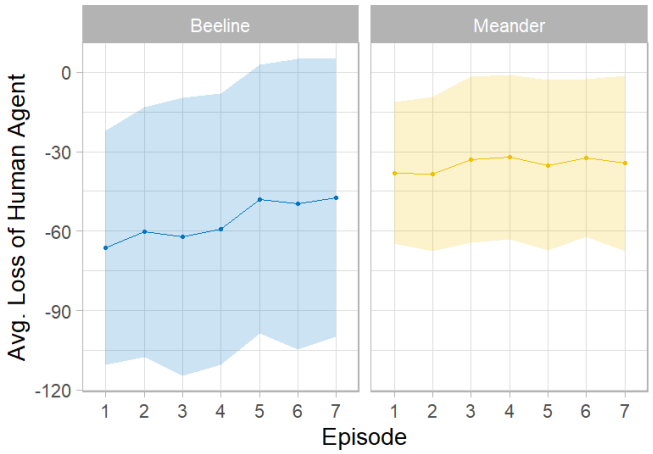

Figure 5: Human defender average Loss against two different attacker strategies, without Green agent on a small network.

Such human data provide some initial evaluations of the IBL model of defense. As observed in the figures above, human participants in the IDG perform worse against the Beeline than the Meander attacker. But also, just like IBL models, humans are able to learn and improve their performance against Beeline more than against Meander attackers.

4 Conclusions and Future Work

In the battlefield of the future, cognitive models will play a critical role. Cognitive models will help emulate the behavior of defenders, attackers, and users, and they would allow to achieve a level of dynamic, adaptive, and personalized decision making that no other technology can achieve. Cognitive models in their goal of emulating human behavior are capable of informing game-theoretic, Machine Learning, optimization algorithms, and other advanced technologies with predictions about human dynamic decision making. This process has been demonstrated in recent work and provides great potential for the future of AICA (Aggarwal et al., 2022)

To develop and test AICA collaboration capabilities, simulation environments such as CybORG will need to be advanced and used in experimentation. We developed a research environment where IBL defense agents can perform a cyber defense task. Using the IBL model we conducted a simulation experiment to test the performance of the IBL model against the two attacker agents of the adapted CAGE scenario. We also developed an Interactive Defense Game to demonstrate the behavior of human defenders.

The cyber defense of the future will involve AICA that collaborate with other computational agents and cyber experts in a team (Theron & Kott, 2019). This aspect of future cyber operations requires the development of intelligent autonomous systems that are able to actively coordinate with other defenders (Human and AI) to act appropriately and optimize the global performance of the team (National Academies of Sciences & Medicine, 2021; Metge, Maille, & Le Blanc, 2021). Our next step is to develop more complex human-machine scenarios. In particular, a dynamic IBL Attacker is being developed to study the impact of a dynamic autonomous agent fighting another adaptive autonomous agent. We also plan to update the IGD to perform a collaborative real-time experiment between human and IBL-enhanced agents.

Using this advanced research framework, we will address questions concerning Human-AICA collaboration. In particular, when considering the structure of collaborative work between humans and autonomous agents, one would necessarily need to consider the level of representation of teammates (Sulistyawati, Wickens, & Chui, 2009; Cuevas, Fiore, Caldwell, & Strater, 2007). In a context of Human-AI collaboration, it is therefore important to provide teammates with the tools to build a mutual understanding (McNeese, Demir, Cooke, & Myers, 2018; Jiao, Zhou, Gebraeel, & Duffy, 2020; Prebot, 2020), and cognitive architectures will play a key role in achieving this goal (Gonzalez et al., 2020).

REFERENCES

Aggarwal, P., Thakoor, O., Jabbari, S., Cranford, E. A., Lebiere, C., Tambe, M., & Gonzalez, C. (2022). Designing effective masking strategies for cyberdefense through human experimentation and cognitive models. *Computers & Security*, 102671.

Baillie, C., Standen, M., Schwartz, J., Docking, M., Bowman, D., & Kim, J. (2020). Cyborg: An autonomous cyber operations research gym. *arXiv preprint arXiv:2002.10667*.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). Openai gym.

Cranford, E., Gonzalez, C., Aggarwal, P., Cooney, S., Tambe, M., & Lebiere, C. (2020). Adaptive cyber deception: Cognitively informed signaling for cyber defense. In *Proceedings of the 53rd hawaii international conference on system sciences*.

Cranford, E. A., Gonzalez, C., Aggarwal, P., Tambe, M., & Lebiere, C. (2020). What attackers know and what they have to lose: Framing effects on cyber-attacker decision making. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 64, pp. 456–460).

Cranford, E. A., Lebiere, C., Rajivan, P., Aggarwal, P., & Gonzalez, C. (2019). Modeling cognitive dynamics in (end)-user response to phishing emails. *Proceedings of the 17th ICCM*.

Cuevas, H. M., Fiore, S. M., Caldwell, B. S., & Strater, L. (2007). Augmenting team cognition in human-automation teams performing in complex operational environments. *Aviation, space, and environmental medicine*, *78*(5 Suppl), B63–B70.

David, R. A., & Nielsen, P. (2016). *Defense science board summer study on autonomy* (Tech. Rep.). Defense Science Board Washington United States.

Dhir, N., Hoeltgebaum, H., Adams, N., Briers, M., Burke, A., & Jones, P. (2021). Prospective artificial intelligence approaches for active cyber defence.

Du, Y., Prebot, B., Xi, X., & Gonzalez, C. (2022). Towards autonomous cyber defense: Predictions from a cognitive model. *Under review*.

Dutt, V., Ahn, Y.-S., & Gonzalez, C. (2011). Cyber situation awareness: Modeling the security analyst in a cyber-attack scenario through instance-based learning. In Y. Li (Ed.), *Data and applications security and privacy xxv* (pp. 280–292). Berlin, Heidelberg: Springer Berlin Heidelberg.

Gonzalez, C. (2022). Learning and dynamic decision making. *Topics in Cognitive Science*.

Gonzalez, C., Aggarwal, P., Lebiere, C., & Cranford, E. (2020). Design of dynamic and personalized deception: A research framework and new insights. In *Proceedings of the 53rd hawaii international conference on system sciences.*

Gonzalez, C., Ben-Asher, N., Oltramari, A., & Lebiere, C. (2014). Cognition and technology. In *Cyber defense and situational awareness* (pp. 93–117). Springer.

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

Jiao, J. R., Zhou, F., Gebraeel, N. Z., & Duffy, V. (2020). Towards augmenting cyber-physical-human collaborative cognition for human-automation interaction in complex manufacturing and operational environments. *International Journal of Production Research*, *58*(16), 5089-5111. doi: 10.1080/00207543.2020.1722324

Kott, A. (2018). Bonware to the rescue: the future autonomous cyber defense agents, a keynote at the conference on applied machine learning for information security. Washington DC. doi: 10.13140/RG.2.2.34126.31042

Kott, A., & Theron, P. (2020). Doers, not watchers: Intelligent autonomous agents are a path to cyber resilience. *IEEE Security & Privacy*, *18*(3), 62–66.

Kott, A., Théron, P., Mancini, L. V., Dushku, E., Panico, A., Drašar, M., ... others (2020). An introductory preview of autonomous intelligent cyber-defense agent reference architecture, release 2.0. *The Journal of Defense Modeling and Simulation*, *17*(1), 51–54.

McNeese, N. J., Demir, M., Cooke, N. J., & Myers, C. (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, *60*(2), 262-273. (PMID: 29185818) doi: 10.1177/0018720817743223

Metge, A., Maille, N., & Le Blanc, B. (2021). Operators and autonomous intelligent agents: human individual characteristics shape the team's efficiency. In *Aica 2021.*

National Academies of Sciences, E., & Medicine. (2021). *Human-ai teaming: State of the art and research needs.* Washington, DC: The National Academies Press. doi: 10.17226/26355

Nguyen, T. N., McDonald, C., & Gonzalez, C. (2021). *Credit assignment: Challenges and opportunities in developing human-like ai agents* (Tech. Rep.). Carnegie Mellon University.

Prebot, B. (2020). *Représentation partagée et travail collaboratif en contexte c2: monitoring d'opérateurs en situation simulée de command and control.* (Unpublished doctoral dissertation). Bordeaux.

Standen, M., Lucas, M., Bowman, D., Richer, T. J., Kim, J., & Marriott, D. (2021). Cage challenge 1. In *Ijcai-21 1st international workshop on adaptive cyber defense.* arXiv.

Sulistyawati, K., Wickens, C. D., & Chui, Y. P. (2009). Exploring the concept of team situation awareness in a simulated air combat environment. *Journal of Cognitive Engineering and Decision Making*, *3*, 309 - 330.

Thanh, C. T., & Zelinka, I. (2019). A survey on artificial intelligence in malware as next-generation threats. In *Mendel* (Vol. 25, pp. 27–34).

Theron, P., & Kott, A. (2019). When autonomous intelligent goodware will fight autonomous intelligent malware: A possible future of cyber defense. In *Proceedings - ieee military communications conference milcom* (Vol. 2019-Novem). doi: 10.1109/MILCOM47813.2019.9021038

Vieane, A., Funke, G., Gutzwiller, R., Mancuso, V., Sawyer, B., & Wickens, C. (2016). Addressing human factors gaps in cyber defense. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*(1), 770-773. doi: 10.1177/1541931213601176

Xu, T., Singh, K., & Rajivan, P. (n.d.). Modeling phishing decisions using instance based learning and natural language processing.