*Human Factors in Design, Engineering, and Computing,* Vol. 159, 2024, 305–314
https://doi.org/10.54941/ahfe1005595

# Examining the Role Memory Plays in Cyber Defence Evaluation: Risk and Uncertainty Demystified

**Asmaa Aljohani[1,2] and James Jones[2]**

[1]Taibah University, Saudi Arabia
[2]George Mason University, USA

## ABSTRACT

Decision-making tasks such as hacking may involve risk and/or uncertainty, contingent upon the hackers' knowledge and expectations of the defensive measures in place. Since memory systems can be activated or inhibited depending on whether a decision is made under uncertainty or risk (Lu et al., 2022; Nicholas et al., 2022), we first investigate how such memory settings and systems affect the accuracy of predicting real-world behaviors using datasets collected under implicit and explicit learning schemes. The findings show an effect of memory systems and settings on the predictive abilities of the models. Additionally, we examined how augmenting reinforcement learning agents with similar memory systems and settings shapes their behaviors when they traverse environments with (un)certain observation spaces. The results point out differences in agents' performance which was found to be influenced by many factors, including the memory systems and settings.

**Keywords:** Deception, Uncertainty, Risk, Reinforcement learning, Observation, Memory systems

## INTRODUCTION

A decision-making process involves internal cognitive processes like perception, attention, and memory (Prezenski et al., 2017) and external factors (e.g., salient features (Cranford et al., 2020), which play a role in activating or inhibiting internal cognitive processes, leading to varying degrees of (un)certainty in one's estimates. Although risk entails some uncertainty, it should be noted that risk and uncertainty are two distinct conditions (De Groot and Thurik, 2018). Uncertainty can be defined as either the complete lack of knowledge or the presence of limited knowledge of the future, the past, or the current events, while risk indicates that the decision-maker has accumulated knowledge and experiences sufficient for an informed decision (see Figure 1). Generally, in a real-world scenario, there are three types of unknowns(Rueter, 2013; Gaskett, 2003), risk, uncertainty, and indeterminacy, ordered from the easiest to the hardest to tackle.

As behavior is heterogeneous amongst individuals, developing a defense strategy that is effective against the various types of attackers requires the defense technology to satisfy an anti-compromise security state. One way to achieve this at the technical level is by converting the hacking task from a decision-under-reducible-uncertainty problem to a decision-under-indeterminacy problem. This approach can be operationally expensive and may interfere with normal systems operations. Another alternative is to minimize the indeterminate nature of hackers' tendencies and to use such tendencies to plan deception placement strategically.

The first part of this research explored the impact of various memory systems and settings on predicting future decisions using datasets of participants performing gambling tasks involving implicit (Iowa Gambling Task) and explicit (sure/gamble) learning rules. The second part investigated how RL attackers with various risk-taking tendencies modeled using various memory systems and settings (e.g., impairments) approach hacking tasks. To summarize, this work analyzed the role memories play in human and RL attackers' decisions since how and when past/recent experiences are encoded and retrieved (i.e., what memory gets activated/inhibited, how activation/inhibition occurs) could be detrimental to how future experiences unfold.

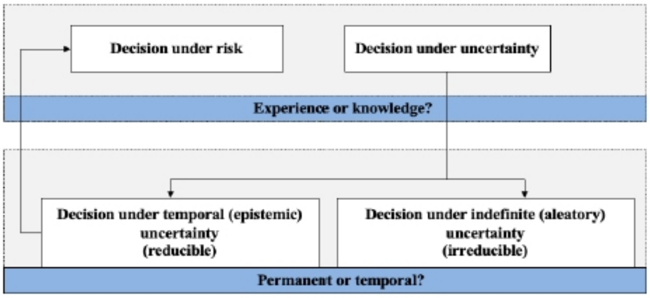
Decision under risk
Decision under uncertainty
Experience or knowledge?
Decision under temporal (epistemic) uncertainty (reducible)
Decision under indefinite (aleatory) uncertainty (irreducible)
Permanent or temporal?

**Figure 1:** A decision-making process. Decisions under temporal uncertainty could be reduced into decisions under risk after an attacker has accumulated enough knowledge and experience.

## PART 1: A PRELIMINARY EXAMINATION OF THE IMPACT OF MEMORY ON PREDICTING REAL-WORLD BEHAVIOR

To analyze the effect of memory systems and settings on predicting real-world behavior, we examined the predictive ability of models calibrated to simulate different memory systems using real-world psychological datasets.[1] The preliminary analysis of the role memory systems and settings play in shaping real-world behavior was conducted using gambling datasets because the tasks used to generate such datasets share some similarities with hacking activities. That is, hackers perform hacking tasks under an assumption of risk and/or uncertainty, depending on their knowledge and expectations about cyber defenses.

[1] The memory settings and systems are adopted from https://github.com/doerlbh/HumanLSTM (Lin et al., 2022) and https://github.com/qihongl/learn-hippo (Lu et al., 2022).

## Method

We examined the role of different memory models in predicting the behavior of human participants using datasets obtained from two different gambling tasks; we note that in the Iowa gambling task, the game rules are implicit and could be learned over time, while the other task is non-dynamic and provides explicit information about expected outcome probabilities.

## Results and Discussion

Since the gambling datasets were relatively small, we utilized a 5-fold cross-validation approach. Because samples were selected randomly, a cross-validation approach can provide a reliable measure of the fit of the models (Hawkins et al., 2003). All gambling datasets were split into 70% for training and 30% for testing. The metric used to evaluate the models is the RMSE, a commonly used metric to evaluate regression problems.

## The Case of Implicit Learning

**Effects of Memory Models:** It is established that individuals' tendencies toward risk and uncertainty differ across unique populations (Busemeyer and Diederich, 2010; Wallis, 2007). Since memory plays a role in shaping the behavior of individuals, we analyzed the ability of models augmented with different memory systems and settings to predict subsequent advantageous and disadvantageous actions and examined how different populations approach such actions. See Figure 2.

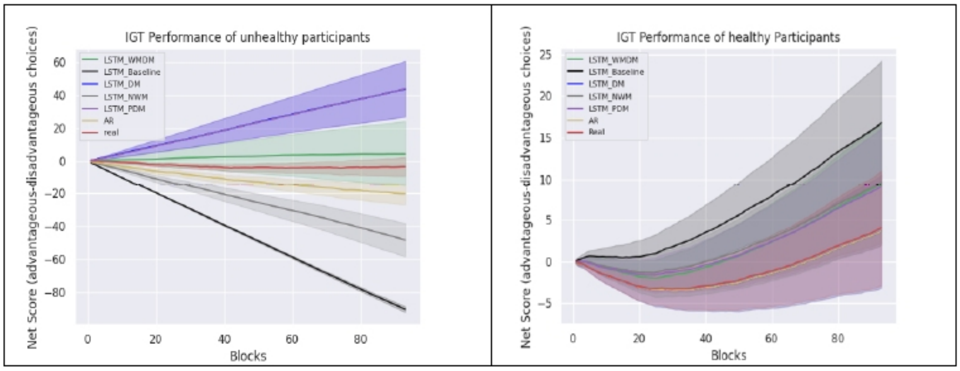

**Figure 2:** The performance of healthy and pathological gamblers.

Table 1 shows the results of evaluating different memory models trained using the Iowa gambling datasets. As can be seen in Table 1, an approximately 57% improvement in the prediction accuracy of the DM memory system, compared to the baseline model was observed. I consider a 30% decrease in the RMSE values between a baseline and a new model to be significant, following (Nau, 2014). This result is consistent and holds for healthy participants only. When the source of a decision impairment is a deficit in memory functions, the difference between healthy and unhealthy (non-addict) participants' use of past and recent experiences is more pronounced (Busemeyer and Stout, 2002). Such a finding might explain why a model augmented with distant memory, as opposed to recent memory, performed better on healthy participants' datasets but not on pathological gamblers' datasets since memory deficits may cause individuals to rely more on recent memories (Busemeyer and Stout, 2002).

For the pathological gamblers' datasets, only the first-order vector auto-regressive model performed better than other memory models. Table 1 shows a significant improvement in the predictive ability of the VAR model compared to the baseline, with an approximately 38.65% decrease in the RMSE value. None of the memory systems or settings where the memory gate value was fixed (unconditional) could predict pathological participants' behaviors. Likewise, the predictive abilities of the models were consistent even when the settings were adjustable in the WMDM condition, resulting in the vector auto-regressive model being the best performing for pathological gamblers' datasets but not for the healthy individuals' datasets.

## The Case of Explicit Learning

**Effects of Memory Models:** We expected the vector auto-regression model to outperform other non-linear models due to its important characteristics (i.e., its reliance on the most recent values) (Brooks and Sokol-Hessner, 2020; McCormick and Telzer, 2018). However, the performance of the vector auto-regression model was the worst compared to all other memory models and settings examined. One reason that could explain the performance degradation is the choice of the variables used to fit the model, which left out important characteristics of risky decisions. That is, in a risky decision-making task, participants are expected to process gains and losses differently. In the present study, we used two variables (gamble and sure), which provided limited information on how the magnitudes of losses and gains would affect decisions.

**Table 1.** RMSEs of models equipped with different memory systems and settings.

| Implicit Learning | | | Explicit Learning | | |
|---|---|---|---|---|---|
| Mem. System | Episodic Mem Gate | RMSE | Mem. System | Episodic Mem Gate | RMSE |
| Healthy Participants | | | DM Capacity: 4 items | | |
| Baseline (LSTM) | - | 0.181 | Baseline (LSTM) | - | 0.290 |
| DM | 0.25 (open) | **0.077** | DM | 0.25 (open) | 0.407 |
| WMDM | 0.0 (closed) | 0.114 | PDM | 0.25 (open) | 0.294 |
| WMDM | 0.25 (open iff t<=50). | 0.134 | NWM | - | 0.359 |
| PDM | 0.25 (open) | 0.1 | VAR(1) | - | 0.457 |
| NWM | - | 0.104 | WMDM | 0.0 (closed) | 0.219 |
| VAR(1) | - | 0.154 | WMDM | 0.25 (open) | **0.126** |
| Participants with Pathological Tendencies (addicts) | | | DM Capacity: 14 Items | | |
| Baseline (LSTM) | - | 0.414 | Baseline (LSTM) | - | 0.297 |
| DM | 0.25 (open) | 0.401 | DM | 0.25 (open) | 0.209 |
| WMDM | 0.0 (closed) | 0.391 | PDM | 0.25 (open) | 0.339 |
| WMDM | 0.25 (open iff t<=50). | 0.403 | NWM | - | 0.353 |
| PDM | 0.25 (open) | 0.414 | VAR(1) | - | 0.450 |
| NWM | - | 0.340 | WMDM | 0.0 (closed) | 0.314 |
| VAR(1) | - | **0.246** | WMDM | 0.25 (open iff t<=50). | **0.252** |
| | | | WMDM | 0.25 (open) | 0.339 |

## PART 2: AN ANALYSIS OF THE ROLE OF MEMORY IN SHAPING RL AGENTS' BEHAVIOR IN (UN)CERTAIN ENVIRONMENTS

In this part, we aim to analyze how the agents' behavior when equipped with similar memory systems and settings changes if they interact with environments characterized by reduced, noisy, masked, and controlled observation spaces.

### Method

We used a modified version of CyberBattleSim (CBS), a high-level simulation environment, that facilitates evaluating attack strategies through simulating multi-stage attacks on network graphs (Microsoft, 2021). To analyze the role memory plays in an RL agent's decision, we replaced the simple linear layers of a Deep Q-learning (DQN) algorithm with an LSTM layer, where memory systems and settings are adjusted to model unique behaviors. The implementation of memory models and settings follows the same procedures used in implementing the non-linear models discussed in Part 1. Additionally, past experiences are sampled from a replay buffer in a randomized or prioritized fashion. While experiences are replayed uniformly (i.e., without any criterion) in the case of randomized sampling, with prioritized sampling, samples with high TD errors (e.g., unfamiliar or new experiences) get replayed (Schaul et al., 2016). For prioritized and randomized sampling, training with minibatches of size 32 is performed over 400 episodes and 5000 steps; the DM capacity is capped at ten. For the WMDM/RM model, retrieval from DM is allowed if and only if the timestep is a multiple of 16.

In the present study, we note that noisy, masked, and reduced observations were applied to a portion of the network. Additionally, a major difference between reduced, masked, and noisy observation spaces is that in an environment with a reduced observation space, while observations are unpredictable, the agents are not penalized. On the other hand, noisy and masked observations are unpredictable, but the agents are penalized if the observation does not lead to a true state. Moreover, the main difference between masked and noisy observations is that noisy observations are probabilistic, while masked observations are not. That is, when observations are masked, part of the observation returned is always hidden, forcing the agent to try different combinations until it succeeds.

### Results and Discussions

#### The interplays between attacker's types, memory models, types of experiences, and observation spaces

We assumed models with the ability to learn or preserve long-term and/or short-term dependencies and models that cannot learn to vary in their approach to the unique observation spaces. More specifically, we expected models that are capable of learning or preserving short-term and long-term dependencies to perform better in uncertain environments, since learning from past experiences can improve learning efficiency (Minsky, 1961).

To evaluate the following hypotheses, we used two performance metrics, some of which are typically recommended for evaluating reinforcement learning algorithms (Chan et al., 2020; Colas et al., 2019): **1**) the average cumulative reward, which describes the agent's overall performance, **2**) the winning rate, which is the percentage of episodes in which the agent achieved the winning requirements. We used Welch's t-test to compare the average cumulative rewards between different types of agents interacting with different observation spaces. The winning rate is used as an additional descriptive metric to evaluate the performance of agents.

**Hypothesis:** Agents equipped with episodic memory and recent memory (WMDM) perform better than similar agents where retrieval from episodic memory is disabled.

When sampling is randomized, the success rate of an agent equipped with episodic and recent memory was higher (60.25%) than that of an agent equipped with recent memory only (24.5%) when the agents traversed an environment with a noisy observation space. If the observation space is controlled, an agent equipped with episodic and recent memory performed worse (the success rate = 38.75%) compared to an agent that is equipped with recent memory only (the success rate = 63.5%).

On the other hand, when traversing an environment with a controlled observation space and if sampling is prioritized, an agent equipped with episodic and recent memory performed better (the success rate = 75%) than an agent equipped with a recent memory only (the success rate = 42.25%). While prioritized sampling improved the performance of agents in a controlled environment, it did not enhance the performance of agents, irrespective of the memory settings (WM with or without episodic memory) if the agent traversed an environment with a noisy observation space.

**Finding:** For a satisficing agent interacting with a noisy observation space, episodic and recent memories have a positive effect on **the success rate** but a negative effect when an agent is interacting with a controlled observation space only if experiences are drawn randomly.

**Table 2.** Statistical test results.

| **The Impact of a Maximizing Strategy** | | |
|---|---|---|
| Sampling | Obs space | WMDM (retrieval enabled X retrieval disabled) |
| randomized | | Welch's t-test of the average cumulative rewards |
| | noisy | −176.651*** |
| | control | 16.990*** |
| | masked | 15.128*** |
| | reduced | −1.136 |
| prioritized | | |
| | noisy | −238.146*** |
| | control | 13.689*** |
| | masked | −68.065*** |
| | reduced | 1.776 |

(Continued)

**Hypothesis:** The type of experiences the agent learns from has an impact on the agent's performance. We note that the type of experiences utilized affects agents' risk propensities as noted in (Schaul et al., 2016).

**Table 2.** Continued

| The Impact of a Satisficing Strategy | | |
|---|---|---|
| randomized | | |
| | noisy | 121.308*** |
| | control | −18.539*** |
| | masked | 15.832*** |
| | reduced | −1.854 |
| prioritized | | |
| | noisy | −1.399 |
| | control | 63.527*** |
| | masked | 15.634*** |
| | reduced | 2.919*** |

*p<0.1, **p<0.05, ***p<0.01

**Finding:** Drawing from unseen and novel experiences (i.e., prioritized sampling) degrades the performance (in terms of **the averaged cumulative rewards**) of models equipped with episodic memory, specifically if the agent is interacting with an environment characterized by a noisy observation space.

**Table 3.** Statistical test results – Effects of experiences.

| Mem System | Obs Space | Randomized X Prioritized |
|---|---|---|
| DM (maximiser) | | Welch's t-test |
| | noisy | 8.673*** |
| | control | 6.761*** |
| | masked | –66.495*** |
| | reduced | –0.396 |
| DM (satisficer) | | |
| | noisy | 4.096*** |
| | control | 3.093*** |
| | masked | –32.795*** |
| | reduced | –0.539 |
| WMDM (maximizer) | | |
| | noisy | –2.359** |
| | control | 18.624*** |
| | masked | 64.044*** |
| | reduced | 0.4955 |
| WMDM (satisficer) | | |
| | noisy | 322.778*** |
| | control | –48.764*** |
| | masked | –39.346*** |
| | reduced | –0.625 |
| PDM (maximizer) | | |
| | noisy | – |
| | control | –3.683*** |
| | masked | –25.280*** |
| | reduced | 0.175 |
| PDM (satisficer) | | |
| | noisy | 28.215*** |
| | control | –33.089*** |
| | masked | –14.190*** |
| | reduced | 0.7451 |

*p<0.1, **p<0.05, ***p<0.01

The results above showed that performance improvement is contingent upon multiple factors. More specifically, the type of strategy the agents follow (i.e., satisficing or maximizing) and the underlying attributes of the observation spaces influenced the agents' behavior in unique ways. Another factor affecting agents' performance is the type of memories stored in episodic memory, which includes either seen or unseen experiences. We note that episodic memory consists of a set of cell states whose content is scaled based on the relationships between the current cell state, the retrieved cell state, and the action selected. Thus, adding the value of the action or using it as a criterion to retrieve relevant memory may change the results dramatically. While past experiences are expected to improve learning, it should be mentioned that some studies noted a possible negative effect of prior knowledge on agents' performance, where a random agent is expected to perform better (Doshi-Velez and Ghahramani, 2011; Dubey et al., 2018). Our findings, however, showed that equipping agents with prior knowledge in the form of episodic memory does not have a consistent effect across the different types of observation spaces.

## CONCLUSION

Existing security studies analyze risky behaviors by either engaging healthy populations, populations whose health conditions are unknown, or by using models capable of predicting a typical human behavior. However, cyber deviance or cybercrimes are found to be prevalent in atypical populations whose risk propensities differ from that of the general population, emphasizing the importance of considering such populations.

The present study's findings revealed that the condition under which a decision is made (i.e., decision-under-risk vs. decision-under-uncertainty) and the population involved are significant factors that need to be considered if better predictive models are to be designed. We demonstrated that predicting pathological gamblers' behavior and healthy participants' behavior accurately requires unique memory systems and settings. Likewise, the behaviors of individuals engaged in implicit or explicit learning schemes cannot be accurately predicted using a singular model or a memory system.

For cybersecurity research, examining the inter-plays between attackers' strategies, the root causes of attackers' behaviors (e.g., memory impairments, motivations), and the environments with which attackers interact is highly challenging, especially if human attackers are to be involved. Thus, we analyzed the impact of augmenting reinforcement learning agents, whose underlying learning strategies resemble that of humans, with similar memory systems and settings. The findings showed that agents' behavior is shaped by the underlying learning strategy, the observation space traversed, and the type of experiences and memory systems and settings used to guide the learning process.

## ACKNOWLEDGMENT

This project was supported by resources provided by the Office of Research Computing at George Mason University (URL: https://orc.gmu.edu) and funded in part by grants from the National Science Foundation (Awards Number 1625039 and 2018631).

## REFERENCES

Brooks, H. R., & Sokol-Hessner, P. (2020). Quantifying the immediate computational effects of preceding outcomes on subsequent risky choices. Scientific Reports, 10(1), 9878. https://doi.org/10.1038/s41598-020-66502-y

Busemeyer, J. R., & Diederich, A. (2010). Cognitive Modeling. SAGE Publications.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. Psychological Assessment, 14(3), 253–262. https://doi.org/10.1037/1040-3590.14.3.253

Chan, S. C. Y., Fishman, S., Canny, J., Korattikara, A., & Guadarrama, S. (2020). Measuring the Reliability of Reinforcement Learning Algorithms. ICLR. ICLR.

Colas, C., Sigaud, O., & Oudeyer, P.-Y. (2019). A Hitchhiker's Guide to Statistical Comparisons of Reinforcement Learning Algorithms. ICLR. ICLR.

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. 64th Human Factors and Ergonomics Society (HFES) Annual Conference, 5.

De Groot, K., & Thurik, R. (2018). Disentangling Risk and Uncertainty: When Risk-Taking Measures Are Not About Risk. Frontiers in Psychology, 9, 2194. https://doi.org/10.3389/fpsyg.2018.02194

Doshi-Velez, Finale, and Zoubin Ghahramani. 2011. "A Comparison of Human and Agent Reinforcement Learning in Partially Observable Domains." in Proceedings of the Annual Meeting of the Cognitive Science Society.

Dubey, Rachit, Pulkit Agrawal, Deepak Pathak, Thomas L. Griffiths, and Alexei A. Efros. 2018. "Investigating Human Priors for Playing Video Games." in ICML. arXiv.

Gaskett, C. (2003). Reinforcement learning under circumstances beyond its control. Proceedings of the International Conference on Computational Intelligence for Modelling Control and Automation, 12.

Hawkins, D. M., Basak, S. C., & Mills, D. (2003). Assessing Model Fit by Cross-Validation. Journal of Chemical Information and Computer Sciences, 43(2), 579–586. https://doi.org/10.1021/ci025626i

Lin, Baihan, Djallel Bouneffouf, and Guillermo Cecchi. 2022. "Predicting Human Decision Making with LSTM." p. 9 in 2022 International Joint Conference on Neural Networks.

Lu, Qihong, Uri Hasson, and Kenneth A. Norman. 2022. "A Neural Network Model of When to Retrieve and Encode Episodic Memories." eLife 11: e74445. DOI: 10.7554/eLife.74445.

McCormick, E. M., & Telzer, E. H. (2018). Contributions of default mode network stability and deactivation to adolescent task engagement. Scientific Reports, 8(1), 18049. https://doi.org/10.1038/s41598-018-36269-4

Microsoft Defender Research Team. (2021). CyberBattleSim. GitHub. https://github.com/microsoft/cyberbattlesim

Minsky, M. (1961). Steps toward Artificial Intelligence. Proceedings of the IRE, 49(1), 8–30. https://doi.org/10.1109/JRPROC.1961.287775

Nau, R. (2014). Linear regression models. https://people.duke.edu/~rnau/compare.htm

Nicholas, Jonathan, Nathaniel D. Daw, and Daphna Shohamy. 2022. "Uncertainty Alters the Balance between Incremental Learning and Episodic Memory." eLife 11: e81679. DOI: 10.7554/eLife.81679.

Prezenski, S., Brechmann, A., Wolff, S., & Russwinkel, N. (2017). A Cognitive Modeling Approach to Strategy Formation in Dynamic Decision Making. Frontiers in Psychology, 8, 1335. https://doi.org/10.3389/fpsyg.2017.01335

Rueter, J. (2013). Diagnosing & Engaging with Complex Environmental Problems.

Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2016, February 25). Prioritized Experience Replay. The International Conference on Learning Representations, Article arXiv:1511.05952. ICLR 2016.

Wallis, J. D. (2007). Orbitofrontal Cortex and Its Contribution to Decision-Making. Annual Review of Neuroscience, 30(1), 31–56. https://doi.org/10.1146/annurev.neuro.30.051606.094334