



DDMLAB YEAR IN REVIEW

2024 - 2025



CONTENTS

4-5	<u>ANNUAL SUMMARY FROM COTY GONZALEZ</u>
6-7	<u>CHANGING MEMBERS</u>
8-9	<u>GRADUATION HIGHLIGHTS</u>
10-18	<u>RESEARCH UPDATES FROM DDMLAB MEMBERS</u>
19	<u>RESEARCH HIGHLIGHT: AI-SDM</u>
20-21	<u>RESEARCH HIGHLIGHT: IARPA RESCIND</u>
22	<u>PUBLICATIONS AND CONFERENCE PAPERS</u>
23	<u>BEST PAPER AWARD: TAILIA MALLOY</u>
24	<u>KEEPING IN TOUCH ALL YEAR!</u>

8-9

[A Year of Cutting Cakes](#)

Saying goodbye to the DDMLab's newest PhDs.

19

[Making a Difference: The DDMLab and AI-SDM](#)

Marking two years with the National Science Foundation
AI Institute for Societal Decision Making.



LETTER FROM THE DIRECTOR

Dear Friends and Collaborators:

We are delighted to share highlights from another dynamic year at the Dynamic Decision-Making Laboratory (DDMLab), 2024–2025. Thank you for allowing us to keep you updated—what a year it has been!

The academic year began with intensive efforts on the **IARPA ReSCIND** program, a visionary initiative designed to leverage cognitive errors and limitations to impede human cyber attackers. This work pushed beyond traditional defense strategies toward cognitively inspired approaches, and we were proud to contribute alongside our collaborators at the University of Texas El Paso, University of Washington, and Peraton Laboratories. Unfortunately, a new IARPA administration chose to discontinue the program just as we were concluding Phase I. While this was deeply disappointing, it reflects broader challenges facing science and academia in the U.S., where significant investments can be undone abruptly by shifting priorities.

On a brighter note, our lab continues advancing several exciting projects. With support from the **National Science Foundation**, we are collaborating with colleagues in Mechanical Engineering and Civil and Environmental Engineering to explore how cognitive models can enhance social capital in physical spaces. Through the **U.S.–Australia International MURI (CATCH)**, we focused on building effective cybersecurity teams of humans and AI agents. Our role in the **NSF AI Institute for Social Decision Making (AI-SDM)** is thriving as the Institute enters its third year. The impact of the Institute is becoming increasingly visible through new initiatives, research projects, and collaborations. I encourage you to explore our ongoing activities at AI-SDM's website. One highlight is the upcoming **second workshop on Human–AI Complementarity for Decision Making**, scheduled for September 25–26, 2025—an expanded version of last year's successful event.

Our publication productivity remains strong, and we are especially proud of a **Best Paper Award** at the HCI International 2025 Conference in Gothenburg, Sweden. The award recognized Tai Malloy's excellent work on training users against social engineering attacks.

This year also brought meaningful personal milestones. I was honored to be elected a **Fellow of the American Association for the Advancement of Science (AAAS)**, in recognition of my contributions to Cognitive Science and the development of Instance-Based Learning Theory. Becoming part of the AAAS Class of 2024 and being inducted in Washington, D.C. this past June was a true highlight. I am proud to be the 15th AAAS Fellow in the history of CMU's Dietrich College of Humanities and Social Sciences, and I am deeply grateful to those who nominated me.

This year took me to the mountains, where I combined intellectual and physical challenges. I attended the Annual Summer Interdisciplinary Conference (ASIC 2025) in the breathtaking setting of Saint Gervais les Bains, followed by a memorable drive across Europe to Vienna for a collective intelligence workshop with colleagues and friends.

Beyond academic life, I continue to find joy in other pursuits. My painting has taken a quieter turn, though I did manage some puzzles and a few canvases. Cycling, on the other hand, has been as energizing as ever—I completed a biking trip through Banff and Jasper, covering more than 200 miles amid stunning scenery, wonderful people, and plenty of food!

As we begin a new academic year, I hold hope for the recovery of U.S. science and research funding, for a year filled with discovery, and for new adventures ahead. Thank you for your continued support and collaboration. Together, let us keep working toward meaningful impact and **toward making this a better world!**





I continue my hobbies of biking and painting when I can!



1980

Fienberg is elected a Fellow of the American Association for the Advancement of Science

Stephen E. Fienberg is the first H&SS faculty member elected to the AAAS. Dietrich College has been home to 15 AAAS Fellows pictured above.

I had the honor to be elected fellow of the American Association for the Advancement of Science, the world's largest general scientific society and publisher of the Science family of journals. With this recognition, I became the 15th AAAS Fellow in the Dietrich College at CMU.



SAYING HELLO TO!



Xiaohong Cai earned her Ph.D. in Psychological and Brain Sciences from

[Indiana University Bloomington](#). Her research examines how context influences human judgment and decision making, with a particular focus on the attraction, similarity, and compromise effects.



Grace Roessling recently earned her PhD in Cognitive Science from

[Rensselaer Polytechnic Institute](#). At the DDM Lab, her research focuses broadly on human-machine collaboration, with an emphasis on Human-AI complementarity in two domains: behavioral cybersecurity and disaster management.



Marko Morrison is a Ph.D. student in Societal Computing through the school of Computer Science.

His research interests include automated cyber-attack/defense systems, multi-agent systems, computational cognition, and reinforcement learning. He received his B.S. in Computer Science from the [University of Washington at Bothell](#).



Vlad Miloserdov is an undergraduate research assistant from Carnegie

Mellon University. He began as a summer intern and will continue working with the lab in the fall. Vlad has helped implement advanced simulations designed to model and counter network threats.

SAYING GOODBYE TO!



Roderick Seow has started a Data Scientist position with PayPal in New York City and **Tailia Malloy** has started their new position in Europe at the University of Luxembourg.

Yinuo Du has taken a research faculty position at UTEP, **Chase McDonald** is working as a Research Scientist at Riot Games, and **Erin Bugbee** is now an Applied Scientist at Amazon's Machine Learning University.



Student research assistants **Saketh Mudundi**, **Sophie Ding**, **Selena Zheng** have continued with their coursework after productive time at the DDMLab.

GRADUATION HIGHLIGHTS





READ MORE FROM OUR GRADUATES!

- [Bugbee, E. H., & Gonzalez, C. \(2025\). Setting and adjusting thresholds in an optimal stopping task: Model predictions and empirical results. In *Proceedings of the Annual Meeting of the Cognitive Science Society* \(Vol. 47\).](#)
- [Du, Y., Prébot, B., Malloy, T., Fang, F., & Gonzalez, C. \(2025\). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. *Computers in Human Behavior: Artificial Humans*, 100148.](#)
- [McDonald, C., & Gonzalez, C. \(2025\). Controllable Complementarity: Subjective Preferences in Human-AI Collaboration. *arXiv preprint arXiv:2503.05455*.](#)

We proudly recognize our new Ph.D.'s Erin Bugbee, Chase McDonald and Yinuo Du. The passion, commitment, and dedication of these three exceptional students has left a lasting impact on the DDMLab that will be felt for years to come. We will miss them all greatly!



RESEARCH UPDATES BY DDMLAB MEMBERS

From Tailia Malloy

This was my third and final year as a postdoctoral researcher in the Dynamic Decision Making Lab. This year we concluded the research project in anti-phishing training using an integration of a Large Language Model (LLM) and Instance Based Learning (IBL) model. We submitted this work, titled 'Improving Online Anti-Phishing Training Using Cognitive Large Language Models' to Computers in Human Behavior which was recently accepted and will be published soon. The main conclusions from this work found that our proposed method of integrating an IBL and LLM model improved student educational outcomes on difficult to correctly identify phishing emails.



Graduation party at the Pittsburgh Golf Club

In addition to finalizing this project, I collaborated with members of the AI Institute for Societal Decision Making to develop an online platform to gather data on disaster relief management decision making. This platform

was used in collaboration with the Florida Division of Emergency Management to test how real professional disaster relief managers make decisions about resource allocation and information gathering during disasters. Another collaboration was between myself and David Hagmann, a former DDMLab collaborator, as well as another project with Roderick Seow, where both of these projects modeled how humans learn from delayed feedback, which has applications in wide ranging areas like stock portfolio management and hiring for human resource managers.

Conferences I attended this year include [Empirical Methods in Natural Language Processing](#) where I presented work in integrating cognitive models with LLM embeddings to improve predictions of student learning progress on our online anti-phishing training platform. Additionally, I presented related work at the [Human Computer Interaction International](#) conference in June, for which we received the best paper award. I also presented our work collaborating with David Hagmann at the [Annual Meeting of the Cognitive Science Society](#), and my work collaborating with Roderick Seow at the [Reinforcement Learning in Decision Making conference](#).

My next steps after working with the DDMLab is a position as a postdoctoral researcher in the [Interdisciplinary Centre for Security, Reliability and Trust](#) at the University of Luxembourg. While in Luxembourg I will continue the work I began at the DDMLab on how humans use LLMs and how they can be improved through the integration of cognitive models like IBL to improve human-computer collaboration. I am still invested in the past work I have done at the DDMLab and I am excited to continue working with members of the lab in the future.

From Maria Ferreira

It has been two years since I joined the DDMLab, and during this time, I have acquired extensive knowledge and insights from my colleagues. My primary research focuses on the Binary Choice Task, where I conducted a study integrating real-time model tracing findings to deliver personalized interventions that enhance participants' task adaptability. I am in the process of drafting a journal paper to present the outcomes of two studies. Our results indicate that the IBL-Personalized (best-fitted decay) approach is most effective in increasing scenarios, while the IBL-Guided (default decay) approach is superior in decreasing scenarios. These findings underscore the significance of tailored cognitive strategies in enhancing adaptability across various contexts. Additionally, I have maintained collaborations with Tailia Malloy, Palvi Aggarwal, and Arunesh Sinha on phishing tasks, resulting in several published research papers and others currently under review.



A significant portion of my time has been dedicated to the ReSCIND project, where I collaborated with Anu Aggarwal, Jeffrey Flagg, and partners from UTEP, UW, and Peraton Labs. We have conducted two studies validating

the cyber isomorph scenarios we developed to reflect the key elements of traditional bias methodologies. Our findings were presented at the [AD&D 2025 workshop](#), demonstrating our success in applying Frequency, Recency, Endowment, and Loss Framing biases to cybersecurity contexts. Our findings show that participants employ more deliberate reasoning in cybersecurity-related problems, while they tend to rely on intuition in general scenarios. We are also collecting data for a new study utilizing a cyber Kill Chain approach that incorporates these cyber scenarios. On a parallel ReSCIND project with Anu, Carolina Carreira, and Alejandro Cuevas Villalba from the Software and Societal Systems Department at CMU, we investigated how cognitive biases impact adversarial decision-making by identifying behavioral signatures of these biases in capture-the-flag (CTF) competitions. Furthermore, we developed PicoCTF challenges based on our cyber isomorph scenarios to assess their performance in realistic cyber tasks. We currently have a publication under review and are conducting data analysis on the challenges we created.

There's a lot of exciting work ahead as I delve into the combination of cognitive strategies and bias methodologies within cybersecurity. This research will help deepen our understanding of participant behavior and improve the design of realistic cyber challenges.

Where have our members been this year?

Maria Ferreira presented a talk reporting experimental findings on binary choices in dynamic environments at [MathPsych/ICCM 2024](#)! The Society for Mathematical Psychology promotes the advancement and communication of research in mathematical psychology and related disciplines.

Listen to her talk on our [YouTube channel](#)!

Adaptation to Change in Binary Choice: Effects of Interventions and Direction of Change

[Maria José Ferreira](#) and Cleotilde Gonzalez



From Chase McDonald

This year marked the conclusion of my time at Carnegie Mellon University and the DDMLab. In April, I successfully defended



my dissertation, On Complementarity in Human-AI Interaction, and earned my doctoral degree. The dissertation brought together three core projects developed over the past several years, all centered on how AI can complement human preferences and learning.

In the first chapter—currently under review at Behavior Research Methods—we introduced two open-source libraries to advance research in human-AI interaction. The first, CoGrid, simplifies the creation of grid-based multi-agent simulation environments. The second, Interactive Gym, enables researchers to run standard simulation environments

directly in browser-based human experiments.

The second chapter built on these tools to propose a novel algorithm, Behavior Shaping, which allows participants to directly influence the behavior of reinforcement learning agents. Through experiments in a collaborative game, we demonstrated that people value both behavioral control and predictability in their AI partners.

The final chapter explored how AI can facilitate human learning in competitive settings by designing effective training partners. Using the two-player fighting game Footsies, we showed that different opponent-design mechanisms (e.g., empowerment maximization, dynamic difficulty) led to distinct strategy development among human learners.

Following graduation, I joined Riot Games as a Senior Research Scientist on the Technology Research team, where I continue to conduct research in AI.

Where have our members been this year?

Chase McDonald presented his work on human-AI complementarity at numerous conferences including the [SJDM Annual Meeting](#) in NYC, [AAAI Conference on Artificial Intelligence](#) in Philadelphia, and [Cognitive Science Society's annual conference](#) in San Francisco.

His findings highlight the need to design AI that prioritizes both task performance and subjective human preferences. By aligning AI behavior with human preferences, we demonstrate how human-AI complementarity can extend beyond objective outcomes to include subjective preferences.



From Roderick Seow

During my second year in the DDMLab, my time was primarily split between wrapping up projects from the previous year and exploring exciting directions

at the intersection of decision making, human-AI teaming, machine learning, and (computational) cognitive modeling.

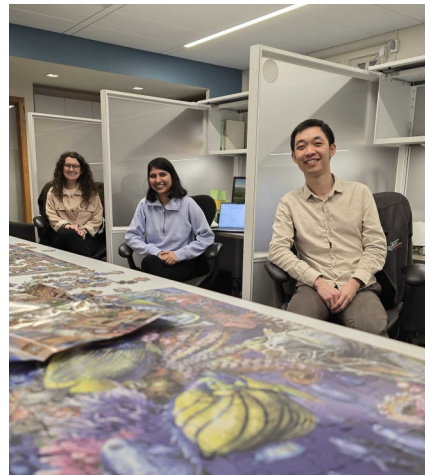
Continuing the collaboration with Hoda Heidari on integrating Theory of Mind capabilities in AI partners to facilitate Human-AI coordination, both simulations and experiments demonstrated that successful coordination by an AI partner relies on accurate predictions of the human agent's actions, which in turn relies on both an appropriate and fine-tuned (Instance-Based Learning) model of the human agent and adopting a complementary k-level of Theory of Mind. Part of these results were presented as a poster at the 2024 meeting of the Psychonomic society and an updated paper was submitted for review to the CogSci 2025 conference.

As an alternative approach to Human-AI coordination, we developed AI agents that integrated algorithms for preference inference, which is another crucial element of Theory of Mind, into the k-level framework. Our proposed AI agent concurrently maintains two internal models of its partner - one as a self-maximizing agent, and another as an agent trying to infer and accommodate the AI's preferences, and estimates which model better describes its partner's coordination strategy based on their actions. The goal, which remains to be tested with actual human partners, is for the proposed agent to adapt to individual differences in both selfish preferences and coordination strategy.

I also got the opportunity to present our project on personalized time-series



forecasters with cognitive models at both the 2024 meeting of the [Society for Judgment and Decision Making](#) and the [6th Workshop on Health Recommender Systems](#). Along the same theme of applying cognitive models to model and predict health-related decisions and outcomes, we are collaborating with David Creswell and Janine Dutcher from CMU Psychology to explore modeling CMU undergraduate students' sleep-related decisions using Instance-Based Learning models. Although this work is in its preliminary stages, through applying IBL modeling, we found that students' decisions to get less or more sleep can be partially accounted for by individual differences in how sensitive they are to various contextual factors such as the difficulty they faced in taking care of themselves on the immediately preceding day.

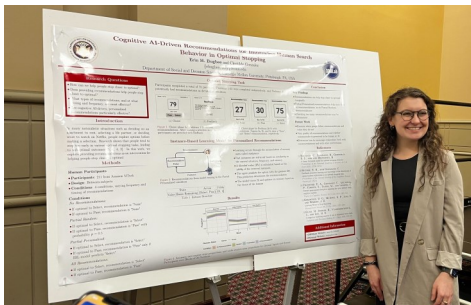


Erin, Anu, and Roderick finally got that puzzle completed!

Finally, I would like to thank everyone in the DDMLab for creating such a warm and welcome environment. Although my time in the lab has been short, I am especially grateful for having met and worked with a group of humble, supportive, and capable researchers. I look forward to reading about the amazing work that will continue to come from the DDMLab and staying in touch with everyone.

From Erin Bugbee

This year marked my final chapter in the DDMLab and at Carnegie Mellon University. On April 8, I defended my dissertation, “Balancing Exploration and Exploitation: Sequential Decision Making in Humans and Machines,” with Coty as the chair and Russell Golman and Sudeep Bhatia as committee members. Having my family and friends there made the day incredibly special, and the moment I was first called “Dr. Bugbee” is one I will never forget. Graduation in May was similarly remarkable, surrounded by family and friends, and capped by the lab celebration at the Pittsburgh Golf Club shared with Chase and Yinuo.



Presenting work at the [2025 AAAI Conference on Artificial Intelligence](#)!

Throughout the year, I presented my research in a variety of venues. I gave a talk at [posit::conf \(2024\)](#) titled “To Explore or To Exploit: Decoding Human Decision Making with R and Python” and a talk at [SJDM](#) titled “A Behavioral and Model-Based Integration for Sequential Decisions from Experience in Optimal Stopping Tasks.” My paper, “Setting and Adjusting Thresholds in an Optimal Stopping Task: Model Predictions and Empirical Results,” was published at [CogSci 2025](#). I also presented on cognitive AI-driven recommendations at the [AAAI Bridge on Cognitive AI and Modeling of Humans](#) in Philadelphia and discussed my work at the [Institute 2024 on Artificial and Human Intelligence](#) at Schloss Dagstuhl in Wadern, Germany.

Leaving Pittsburgh was bittersweet — I will miss the DDMLab, SDS, CMU, and the incredible

friends I made here. I have now moved to Seattle, where I began my next chapter as an Applied Scientist at Amazon on the AWS Machine Learning University team. I create and teach courses focusing on emerging AI and machine learning technologies while continuing to study human decision making. I am deeply grateful for my years in the lab and will carry the learnings and memories with me wherever the next chapters lead.

From Yinuo Du

This year marks my fifth and final year at DDMLab. I successfully defended my dissertation, titled “Human and AI Decision-making in Cybersecurity: A Multiagent Perspective”.



Three chapters of my dissertation have been accepted by [ACM transactions of social computing](#), [Computers in Human Behavior: Artificial Humans](#), and [Acta Psychologica](#). The final chapter is under review at Psychological Review. Thanks to my advisors for shaping the dissertation and my committee members for the constructive feedback.

During my final year, I was fortunate to work with talented undergraduate students like Rony and Vlad to carry some of my projects forward. I am excited that both efforts are coming to submission milestones. I am immensely grateful for my advisors’ guidance and insights. These projects are impossible without the help from Tai, Maria, and Baptiste, who also sat through my dreadful practice talks and mock interviews during this harsh job search season. I am forever in debt to them.

I will start as a Research Assistant Professor at the University of Texas at El Paso this fall. I plan to deepen my interdisciplinary research that brings together cognitive models and AI agents for cybersecurity teamwork. I am excited to continue ongoing projects nurtured by DDMLab and to expand my collaboration network. Looking forward to contributing more impactful research that benefits society.

From Anu Aggarwal

This year at the DDMLab has been exceptional. I worked on an IARPA project in cyber psychology, focusing on cognitive biases in cyber attackers. This project involved collaboration with Maria from DDMLab, along with teams from UTEP, UW, Peraton Labs – the ReSCIND team, and S3D Lab-Carolina, Alejandro and Saketh, where we analyzed cognitive biases in picoCTF datasets and created new CTF challenges to measure the same.

Last year, we successfully collected data for our first experiment, replicating established literature on biases, termed "cyber-isomorphs." Our key findings included the replication of biases like availability, endowment, loss framing, and the recency effect. Notably, individuals with higher practical knowledge and risk-seeking behavior exhibited reduced biases, and participants displayed more deliberate reasoning in cyber scenarios than in traditional contexts. These insights led to my first paper being accepted at the [AD&D Workshop 2025](#), with significant support from Maria, Coty, and the rest of the team, which I presented virtually on July 4th, 2025. Building on this, we designed two additional experiments: Experiment 2 translated cyber-isomorphs into an attack kill chain, enhancing our understanding of biases in realistic cyber environments, while Experiment 3 utilized the CyberVAN testbed with penetration testers and



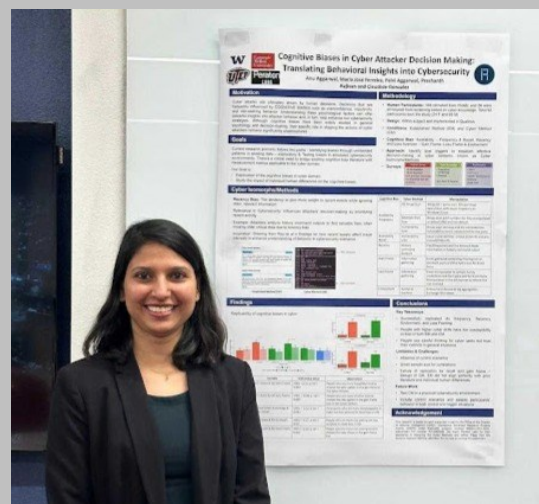
skilled students as participants, refining my skills in experimental design and data collection management. Additionally, a site visit to El Paso allowed me to engage in discussions about our experiments and present a poster on our workshop paper, which boosted my confidence and provided valuable insights from industry experts.

My collaboration with the S3D lab team allowed me to explore cognitive biases in the picoCTF dataset, revealing practical insights for CTF competitions. Working with Carolina, we submitted an accepted paper for HICCS 2025 and are addressing the reviewers' comments. Our study analyzes over 500,000 logs from picoCTF, uncovering behavioral signatures linked to the availability/default effect bias and the sunk cost fallacy. We propose a framework for utilizing these biases in adaptive cyber defenses. Having identified these biases, our primary goal is to design CTF challenges that trigger them, enabling us to test their effectiveness and measure cognitive biases in a similar population on the picoCTF platform. We have completed data collection for this experiment and are now analyzing the results. In addition to my projects, I had the opportunity to complete a course on Experimental Design for Behavioral and Social Sciences, where I earned an A grade. During my time at DDMLab, I significantly improved my skills. I am excited to continue working on my current projects and to explore many more opportunities that await me in the upcoming year.

Where have our members been this year?

Anu Aggarwal had her paper "Cognitive Biases in Cyber Attacker Decision-Making: Translating Behavioral Insights into Cybersecurity" accepted at the [2025 Active Defense & Deception / IEEE EuroS&P Workshop](#)! The AD&D workshop aims at promoting Active Defense as an effective extra security layer, with the goals of moving the research forward and of encouraging its adoption by the industry.

Listen to her talk on our [YouTube channel](#)!



From Xiaohong Cai

My doctoral research examined how the presence and configuration of alternatives shape human judgment and decision making, with a focus on the attraction, similarity, and compromise effects. This



work revealed a general context effect, in which the relational structure among alternatives systematically influences judgments, extending beyond the three classical context effects. As a postdoctoral research associate in the DDMLab, I am expanding this research to human–AI interaction, investigating how contextual structures shape collaborative decision processes and applying these insights to design systems that help humans make better decisions.

My doctoral research systematically sampled stimuli across the entire attribute space to investigate how relationships among options influence the magnitude and direction of these effects. Using Bayesian multilevel modeling, I quantified these relationships and identified a general context effect in which the relational structure among options systematically shapes judgments, extending beyond the three classical effects.

As a postdoctoral research associate in the Dynamic Decision Making Lab, I aim to extend this framework to human–AI interaction. My upcoming work will investigate how contextual structures influence decision processes when humans collaborate with AI partners, including how option relationships affect trust, performance, and joint decision outcomes dynamically. Ultimately, my goal is to apply these insights to the design of decision-support systems that leverage context to improve the quality and effectiveness of human–AI collaboration.

From Grace Roessling

I am an incoming postdoctoral fellow at the DDM Lab, and have recently received my PhD in Cognitive Science from Rensselaer Polytechnic Institute. My thesis examined how drivers



use both online visual information and spatial knowledge to guide high-speed steering—investigating how humans execute rapid steering adjustments informed by prior knowledge of the spatial layout. Building on this foundation, my work at the DDM Lab will shift to higher-level decision-making processes in the context of human–machine teaming, with an emphasis on Human–AI complementarity in two domains: behavioral cybersecurity and disaster management. I look forward to expanding upon my expertise by developing cognitive models within the IBL framework, as well as training RL agents to collaborate effectively with humans.



Grace Roessling and Tailia Malloy both come to us from [Rensselaer Polytechnic Institute](https://www.rpi.edu/)!

From Don Morrison

The latest version of PyIBL is now 5.2.1. The largest changes in the last year are the addition of the `aggregate_details` attribution, allowing the extraction



as a Panda DataFrame of the internal computations of a PyIBL Agent; and the `plot()` method facilitating the easy creation of Matplotlib plots of similar information. Smaller features of use in specialized models have also been added or enhanced, as well as some minor bugs fixed. Details of the changes, as well as instructions for easily upgrading, are in the PyIBL documentation at <http://pyibl.ddmlab.com>.

From Marko Morrison

I'll be joining the DDMLab as a PhD student this fall, but have been working with Coty and other DDMLab members. While at UW, I contributed to the IARPA ReSCIND project with Prashanth Rajivan. During this phase, our team designed and executed an experiment measuring the prevalence of cognitive biases in cyberattackers.

In particular, I focused on how our experiment could be translated into some kind of automated defensive tool. To achieve this, I built a Python library called "UAT+", which takes in logs of individual keystrokes from Peraton's "UAT" keylogger and parses them into fully formed commands. The core functionality lies in maintaining timestamps of all keystrokes, including those absent from the final command. With this information, we can compare actions performed in psychological trigger scenarios with those in control scenarios to search for indicators of mental disarray among attackers, which could include frequency of repeated commands, time between commands, and more.

In the future, I plan to work on automated attack and defense systems, as well as developing techniques for cyber deception on each side. Working with everyone has been a lot of fun so far, and I'm looking forward to my time in the DDMLab and at CMU!



From Jeffrey Flagg

Over the past year, I have contributed to many experiments, especially with Anu and Maria on the various IARPA studies. It was very nice to be working with former DDMLab members Palvi Aggarwal and Prashanth Rajivan on ReSCIND. It is great to maintain connections with our former members. It was a pleasure to help Erin and Chase complete their dissertation experiments. I have completed significant updates to our lab manuals and research guides. I continue to serve on CMU's OSF Advisory Board which helps promote best practices in the greater CMU community.



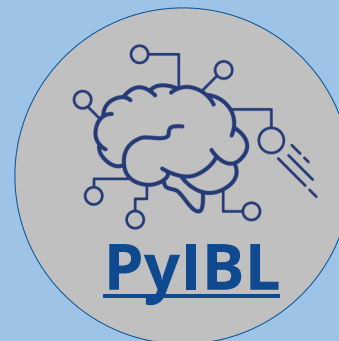
Artwork at the 2025 [CMU Community Picnic](#). That is three years in a row!

I helped coordinate our graduation party at the [Pittsburgh Golf Club](#) for our new Ph.D.'s. This was particularly gratifying as they are the first group of doctoral students that began the program after I joined the DDMLab, so I have been able to see their full careers at CMU. They will be missed!

What is PyIBL?

PyIBL is a Python implementation of a subset of Instance Based Learning Theory. It is made for making computational cognitive models supporting research in how people make decisions in dynamic environments. PyIBL is a module of Python code, useful for creating Python programs.

Learn more [here](#)!



From Selena Zheng

Over the summer at DDMLab, I had the opportunity to work on an exciting project exploring human adaptation in a binary choice task, using the Instance-Based Learning (IBL) model to better understand



the mechanisms behind human behavior. Under the guidance of Maria and Tailia I learned a lot and gained invaluable research experience. We focused on improving people's ability to adapt to continuously changing conditions—a challenge that past research has shown many struggle with. Together, we designed and implemented a new feedback condition and began incorporating additional model features into the IBL model to better reflect human decision-making. Our findings could help design more effective feedback systems and training programs to support better decision-making in dynamic situations. I'm grateful for the chance to learn from great mentors and work with such a supportive team!

From Vlad Miloserdov

Over the course of my internship with the lab this summer, I have worked on implementation of advanced simulations that are designed to model and counter network threats.



Working closely with Yinuo, we have created a platform with futuristic design that will allow researchers around the world to explore the effectiveness of teams with varying team configurations, where each member of the team has a distinct responsibility and has access only to a subset of (network) permissions. This approach provides a controlled yet dynamic environment to study how collaboration, communication, and decision-making shape outcomes under adversarial conditions. My contributions are centered on refining the underlying algorithms and ensuring scalability so that the platform can support increasingly complex threat scenarios. Ultimately, the work lays a foundation for future research on adaptive defense strategies that mirror the realities of modern cybersecurity challenges.

Undergraduate Research

The DDMLab welcomes the participation of undergraduate students who might be interested in gaining research experience in the area of dynamic decision making. All our undergraduate student commit to registering for an [independent study](#) or research for credit for a full semester or more. The number of units vary between 6 and 12, and correspond to the number of hours per week that the student is willing to commit to research within the lab.



From Sophie Ding

This spring semester at DDMLab, I was given the opportunity to respark a subject that has been researched at the lab for quite some time: the dynamic decision-making behavior in "Rock,

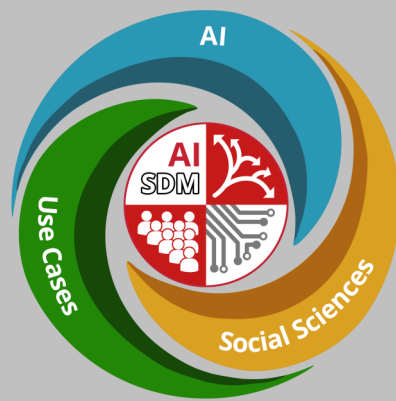


"Paper, Scissors." By studying adversarial behavior of humans when playing RPS against computer models who are designed to compete through different analytics and strategies, we can explore a human's ability to avoid getting exploited, and perhaps even counter-exploit. With the guidance of Roderick and Coty, I have developed thorough understanding in past findings of this area, and learned how to shape my curiosity in this subject in such a way that my proposed research question will efficiently further our knowledge.

Project Highlight: AI-SDM

NSF AI Institute for Societal Decision Making

The National Science Foundation supports the [AI Institute for Societal Decision Making](#), which brings together AI and social sciences researchers to develop human-centric AI for societal good. DDMLab members work with the AI-SDM to explore disaster management scenarios that require complex, often life-saving, decisions to be made under uncertain, dynamic, and resource-constrained circumstances. On September 25-26, the AI-SDM will host an academic workshop focusing on [human-AI complementarity for decision making](#).



The AI-SDM is advancing AI techniques for autonomous decision-making while considering human factors that govern the acceptance of decisions. These AI advancements are not only inspired by human decision-making but are also actively applied to enhance decision-making processes in critical areas.



Coty and Roderick at the 2024 AI-SDM Workshop.

Emergency Response

Try our new [demo](#) developed by Tailia Malloy. In it, you will explore the decision making processes of emergency response managers after a hurricane. These decisions include assessment of damage and allocation of resources. Our goal is to understand the features that emergency response managers use to assess areas in need of additional resources, choose which of the houses to allocate additional resources to, and categorize these houses based on their level of damage across different damage types.



Project Highlight: IARPA ReSCIND

A major grant from the Intelligence Advanced Research Projects Activity (IARPA) Reimagining Security with Cyberpsychology-Informed Network Defenses (ReSCIND) Program helped bring two major DDMLab projects in 2024 and 2025.



The goal of this project was to investigate cyber attackers sensitivity to cognitive biases in capture the flag scenarios. In this study, the DDMLab paired with [picoCTF](#), a platform developed by the experts in security and privacy at Carnegie Mellon University which focuses on providing free educational programs on computer security. picoCTF utilizes a gamified approach built on capture-the-flag framework. It is globally recognized for its capture-the-flag competition held every year, drawing participants from beginner to expert skill levels. These challenges are categorized based on various cybersecurity topics, such as cryptography, reverse engineering, and web exploitation. We focused on two widely studied cognitive biases, availability and sunk cost bias. We hypothesized that participants will prefer commonly encountered settings, ports, vulnerabilities, or strategies over less familiar alternatives. We also theorized that participants will persist in using an initial strategy, even when switching to an alternative would be more efficient.



IARPA ReSCIND picoCTF Team: Carolina Carreira, Coty Gonzalez, Anu Aggarwal, and Maria Ferreira. Also: Alejandro Villalba (not pictured).



In a collaborative project between CMU, the University of Texas at El Paso, the University of Washington, and Peraton Labs, a series of studies were designed to detect cognitive biases among adversaries during a network exploitation process. In the culminating study, we utilized [Cyber Virtual Assure Network \(CyberVAN\)](#), created by Peraton Labs. Peraton Labs works to create foundational technologies for diverse applications to deliver mission critical technology research and high-value solutions to customers across defense, telecom, energy, finance, government, transportation, life sciences and the intelligence community. CyberVAN provides the highest possible fidelity representation of a network by representing the network in a discrete event network simulator and enabling hosts, represented by virtual machines, to communicate over this simulated network. The goal for the attacker in CyberVAN is to exfiltrate sensitive documents from the network. To achieve this goal, an attacker is required to perform network reconnaissance, exploit vulnerabilities to gain access to the systems and perform data exfiltration as a post-exploitation task. The knowledge gained from this study will provide valuable insights into how deceptive strategies, based on cognitive biases, can be used to thwart network reconnaissance from potential adversaries. Data collection for this project has finished in August 2025!

What makes these projects work?

The [IARPA ReSCIND](#) program aims to improve cybersecurity by understanding how human cognition impacts cyber behavior and could affect cyber actors' success while executing malicious activities. The program leverages well-established cognitive vulnerabilities and human limitations to impede cyber attackers. ReSCIND will expand the traditional cyber defense toolkit to help cyber defenders protect National Security Systems and other intelligence community assets across various phases of a cyber attack.



PUBLICATIONS & CONFERENCE PAPERS

In the past year, we published several journal articles and chapters authored by members of the DDMLab and our collaborators. For a full list of publications, please see the publications page on the laboratory's website at

<https://www.cmu.edu/dietrich/sds/ddmlab/publications.html>

Gonzalez, C. (2025). Complex Problem Solving and Dynamic Decision Making: What Is the Difference? In Wendt, N., Holt, D., & Stockhausen, L. (eds.), *Komplexität und Problemlösen: Festschrift für Joachim Funke zum 70. Geburtstag*, Heidelberg: Heidelberg University Publishing, pp. 233–247. <https://doi.org/10.17885/heiup.1067.c23281>.

Bugbee, E. & Gonzalez, C. (2024). Feedback Promotes Learning and Knowledge of the Distribution of Values Hinders Exploration in an Optimal Stopping Task. In *Proceedings of the 46th Annual Conference of the Cognitive Science Society*. pp. 4618–4624. Rotterdam, The Netherlands, July 24–27, 2024. <https://escholarship.org/uc/item/57b644z2>

Cranford, E. A., Lebiere, C., Gonzalez, C., Aggarwal, P., Somers, S., Mitsopoulos, K., & Tambe, M. (2024). Personalized model-driven interventions for decisions from experience. *Topics in Cognitive Science*, 00, 1–24. <https://doi.org/10.1111/tops.127>

Du, Y., Prebot, B., Malloy, T., Fang, F., & Gonzalez, C. (2025). Experimental evaluation of cognitive agents for collaboration in human-autonomy cyber defense teams. *Computers in Human Behavior: Artificial Humans*, 100148. <https://doi.org/10.1016/j.chbah.2025.100148>

Ferreira, M. J., & Gonzalez, C. (2024). Adaptation to change in binary choice: effects of interventions and direction of change. Paper presented at Virtual MathPsych/ICCM 2024. Via <https://mathpsych.org/presentation/1480>.

Malloy, T., Ferreira, M. J., Fang, F., & Gonzalez, C. (2025). Training Users Against Human and GPT-4 Generated Social Engineering Attacks. In *Proceedings of HCI-CPT: 7TH International Conference on HCI for Cybersecurity, Privacy, and Trust* (pp. TBA). HCI International 2025, June 22–27, 2025, Gothenburg, Sweden. https://link.springer.com/chapter/10.1007/978-3-031-92833-8_4

Malloy, T., Seow, R., & Gonzalez, C. (2025). Modeling Attention during Dimensional Shifts with Counterfactual and Delayed Feedback. *Conference on Reinforcement Learning and Decision Making* (2025). <https://arxiv.org/pdf/2501.11161>

McDonald, C., & Gonzalez, C. (2025). Controllable Complementarity: Subjective Preferences in Human-AI Collaboration. <https://arxiv.org/pdf/2503.05455>

Nguyen, T.N., Jamale, K. & Gonzalez, C. (2024). Predicting and Understanding Human Action Decisions: Insights from Large Language Models and Cognitive Instance-Based Learning. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 12(1), 126–136. <https://doi.org/10.1609/hcomp.v12i1.31607>

BEST PAPER AWARD: TAILIA MALLOY

7TH INTERNATIONAL CONFERENCE ON HCI FOR CYBERSECURITY, PRIVACY, AND TRUST

Social engineering attacks such as phishing emails remain a critical method for cybercriminals to exploit sensitive data. Although the threat of AI-generated content in such attacks is growing, current training methods predominantly rely on simplistic human-designed emails. This research introduces a novel experimental paradigm to investigate differences in the detection of human-generated versus AI-generated phishing emails, as well as two different methods by which cyberattackers could use AI as a tool to generate phishing emails. Our behavioral results reveal that emails co-created by humans and Generative-AI models pose a greater challenge to end users compared to emails created by GPT-4 or humans working alone. We also propose a cognitive model that predicts user behavior during training, which offers the potential to be used in future user training to improve training outcomes. Read more [here!](#)



[HCI-CPT](#) intends to help, promote and encourage research in this field by providing a forum for interaction and exchanges among researchers, academics, and practitioners in the fields of HCI and cyber security.



Dynamic Decision Making Laboratory

Carnegie Mellon University
5000 Forbes Avenue
Social and Decision Making Sciences
Porter Hall 223G
Pittsburgh, PA 15213
e-mail: ddmlab@andrew.cmu.edu



Stay Connected All Year!



<http://www.ddmlab.com>



<https://www.facebook.com/ddmlab.cmu>



<https://bsky.app/profile/ddmlab.bsky.social>



<https://www.linkedin.com/showcase/27154823>



<https://www.youtube.com/@dynamicdecisionmakinglab>



Carnegie Mellon University