Curriculum Vitae

Joshua Hewson

joshua_hewson@brown.edu
contact@cyberkind.ai

https://thelemonpig.github.io/
https://cyberkind.ai/

Boston, US
London, UK

Contents

Research *Peer-Reviewed*

Papers most aligned with my current research goals shown with ♦

♦ **Joshua T. S. Hewson**. We Urgently Need Intrinsically Kind Machines.

Joshua T. S. Hewson, Sabina J. Sloman, Marina Dubova (2024). One System for Learning and Remembering Episodes and Rules. *In* 7th annual conference on Computational Cognitive Neuroscience.

Joshua T. S. Hewson, Ke Fang (2024). Evaluating internal and external dissonance of belief dynamics in social systems. *In* 10th International Conference for Computational Social Science.

Joshua T. S. Hewson, Younes Strittmatter, Ioana Marinescu, Chad C. Williams, Sebastian Musslick (2023). Bayesian Machine Scientist for Model Discovery in Psychology. *In* NeurIPS 2023 AI for Science Workshop.

Sebastian Musslick, **Joshua T. S. Hewson**, Benjamin W. Andrew, Younes Strittmatter, Chad C. Williams, George T. Dang, Marina Dubova, John Gerrard Holland (2023). An Evaluation of Experimental Sampling Strategies for Autonomous Empirical Research. *In* Proceedings of the annual meeting of the cognitive science society 45 (45).

Sebastian Musslick, **Joshua T. S. Hewson**, Benjamin Andrew, Sida Li, George Dang, John Gerrard Holland (2023). Evaluating Computational Discovery in the Behavioral and Brain Sciences. *In* AAAI 2023 Spring Symposium on Computational Approaches to Scientific Discovery.

SUBMITTED:

♦ Chad C. Williams*, Daniel Weinhardt*, **Joshua T. S. Hewson**, Martyna Beata, Nicolas Langer, Sebastian Musslick. Enhancing EEG Data Classification Across Diverse Contexts Using Generative Adversarial Networks. [*Submitted to Prestigious Journal*]

Sebastian Musslick, Benjamin Andrew, Chad C. Williams, **Joshua T. S. Hewson**, Star Li, Ioana Marinescu, Marina Dubova, George T. Dang, Younes Strittmatter, John G. Holland. AutoRA: Automated Research Assistant for Closed-Loop Empirical Research.

Joshua T. S. Hewson, Alana Jaskir, Michael J. Frank. Comments on "Dopamine encoding of novelty facilitates efficient uncertainty-driven exploration".

PREPRINT:

♦ **Joshua T. S. Hewson**. Combining Theory of Mind and Kindness for Self-Supervised Human-AI Alignment.

Joshua T. S. Hewson, Star Li. Towards Hierarchical Bayesian Symbolic Regression.

Alana Jaskir, **Joshua T. S. Hewson**, Michael Freund, Daniel Scott. Brainstorm Submissions Responses and Report on Hippocampal SWRs.

IN PROGRESS:
Research Projects approaching completion – working titles

♦ **Joshua T. S. Hewson**, Chad Williams, Daniel Weinhardt, Sebastian Musslick. [Channel Recovery and Signal Generation for EEG Data Across Diverse Contexts Using Generative Adversarial Networks]

Alexander Fengler, …, **Joshua T. S. Hewson**, …, Michael J. Frank [Hierarchical Sequential Sampling Modeling Package HSSM]

Academic Experience

October 2023 - Present
Researcher Assistant in Computational Neuroscience
Brown University – in the Lab for Neural Computation and Cognition

June 2022 – October 2023
Research Assistant in Machine Learning for Scientific Discovery in Psychology and Machine Learning for Computational Neuroscience
Brown University – in the Autonomous Empirical Research Lab

Jan 2021 – Feb 2021
Research Assistant in Causal Inference Algorithms
Carnegie Mellon University – under Peter Sprites

Industry Experience

April 2024 - Present
AI Research Scientist
RedCompass Labs

May 2023 – April 2024
AI Research Advisor
RedCompass Labs

May 2022 – January 2024
Data Science Consultant
QSIDE Institute / RedCompass Labs

May 2020 – Aug 2020, Aug 2019 – Dec 2019, July 2018
Data Science Intern in Human Trafficking Detection and Prevention
RedCompass Ltd, QSIDE Institute

Workshop & Conference Presentations

Poster Presentation
We Urgently Need Intrinsically Kind Machines
NeurIPS 2024 IMOL Workshop
Vancouver Conference Center

Poster Presentation
One System for Learning and Remembering Episodes and Rules
7th annual conference on Computational Cognitive Neuroscience (2024)
Massachusetts Institute for Technology

Poster Presentation
Evaluating internal and external dissonance of belief dynamics in social systems
10th International Conference for Computational Social Science (2024)
University of Pennsylvania

Poster Presentation
Bayesian Machine Scientist for Model Discovery in Psychology
NeurIPS 2023 AI for Science Workshop
New Orleans Conference Center

Talk Presenter
Bayesian Machine Scientist for Model Discovery in Psychology
Carney Brain Institute Computational Modeling Workshop (2023)
Brown University

Talk Presenter
On Modern Trafficking Modeling, Detection, and Prevention
QSIDE Institute Colloquium Series 2022
Institute for the Quantitative Study of Inclusion, Diversity, and Equity

Workshop Involvement

Reviewer
NeurIPS 2024 Safe Generative AI Workshop
Vancouver Conference Center

Reviewer
7th annual conference on Computational Cognitive Neuroscience (2024)
Massachusetts Institute for Technology

Program Assistant
Carney Brain Institute's Third Cognitive Effort Workshop (2022)
Brown University

Internal Presentations

June 2024

Carney Brain Institute Brainstorm Challenge Awards Talk

Brown University

July 2022

Bayesian Symbolic Regression for Model Recovery in Psychology

Autonomous Empirical Research Lab

April 2022

Data Modeling Complexity in Biological Systems

Williams College – Math Major Colloquium

Jan 2022

An Overview of Artificial Intelligence in the Context of Past Technological Advancement

Williams College – Neuroscience of Learning Winter Study

July 2020

Clustering Algorithm Approaches for Analyzing Financial Data

RedCompass Labs – Lunch & Learn

Noteworthy Awards

January 2024
Winners of Brainstorm Program's Brainstorm Challenge
Carney Institute for Brain Science

August 2023
Winners of Brainstorm Program's Computational Modeling Workshop Competition
Carney Institute for Brain Science

December 2021
Innovation Award
RedCompass Labs

Additional Certifications

October 2023
Santa Fe Institute's Complexity Interactive
Santa Fe Institute

August 2023
Santa Fe Institute's Introduction to Agent Modeling Summer School
Santa Fe Institute

July 2020
Santa Fe Institute's Complexity Explorer
Santa Fe Institute

Skills

Mathematics:
- Writing and interpreting proofs
- Statistical analysis and interpretation

Neuroscience & Psychology:
- EEG/SEEG/EMG preprocessing, analysis, augmentation, and generation
- Cognitive Modeling, and data analysis

Agent-Modeling:
- Designing, Implementing, and Analyzing Multi-agent Models

Computer Science
- Languages: Python, Java, C++, R, MATLAB
- Research Software Development (i.e. pip packages)
- Computational Modeling

Data Science:
- Packages: Pandas, NumPy, Matplotlib, Seaborn, Plot.ly
- Bayesian Inference
- Data Visualization and Advanced Data Analysis
- Applying Basic to Advance Machine Learning Algorithms

Machine Learning:
- Packages: PyTorch, SciPy, Sci-kit Learn
- Working with, building, and improving neural network models
- Working with and improving transformer models
- Working with, building, and improving symbolic regression models
- Causal Inference Algorithms

Artificial Intelligence:
- Co-developing RAG-based agents
- Creating, designing, and developing Human-AI Interaction Systems
- Designing and applying algorithms to fine-tune generative models

Industry
- Mediating relationships between academics and businesses
- Advising businesses on AI Strategy
- R&D for internal and external AI products

Soft Skills
- Presenting to small and large groups of people
- Facilitating Group Discussions
- Advising Project Strategy

Education

Sep 2017 – Jun 2022
Undergraduate Degree
Williams College
Major: Mathematics
Minor: Cognitive Science
Classes including:
- Neuroscience of Learning
- Topics in Cognitive Science
- Cognitive Psychology
- Casual Inference
- Mathematical Modeling
- Statistics and Data Analysis
- Human AI Interaction
- Theory of Computation
- Algorithms
- Data Structures

Sep 2014 – Jun 2016
A-LEVELs
TWYFORD C OF E HIGH SCHOOL
- Physics A*
- Mathematics A*
- Further Mathematics A
- Chemistry A
- Geology A (AS)

Sep 2012 – Jun 2014
GCSEs
TWYFORD C OF E HIGH SCHOOL
- IGSCE Physics A*
- IGSCE Chemistry A*
- IGSCE Biology A*
- IGSCE Mathematics A*

- Mathematics A*
- Geology A*
- Geography A*
- Spanish A*
- Latin A*
- History A
- English Literature A*
- English Language A*
- Religious Education A*

HIGHSCHOOL RESEARCH PROJECTS
- IMPERIAL COLLEGE LONDON GOLD CREST AWARD
    - Research Project coordinated by Senior Imperial College Researcher
    - Title: An Investigation into the Link between Friction Coefficient and Contact Angle
- UNIVERSITY COLLEGE LONDON EXTENDED PROJECT QUALIFICATION (**A***)
    - Title: Is Nature Inherently Probabilistic?
    - Original Research, 6000-word dissertation

ADDITIONAL QUALIFICATIONS
- Udemy Bootcamp (2019)
- Santa Fe Complexity Explorer (2017)
- Extended Mathematics Qualification (2014)
- SAT Subject Tests Math II (**800/800**) / Physics (**800/800**)

Growing up, I found joy in connecting disparate ideas to form new insights around how people think, which eventually led me to the U.S. for college. I had long been drawn to psychology, but my perspective widened when I encountered cognitive science through the lens of artificial intelligence when reading about AlphaZero in 2018. This moment revealed to me that complex reasoning could extend beyond humans, opening fascinating questions about the parallels and limitations of human and artificial intelligence.

At Williams College, I immersed myself in cognitive science, computer science, and complexity theory to understand the connections between human cognition and AI. As I explored deeper, it became apparent to me that to understand human cognition computationally I would need to learn more neuroscience. Soon into my journey, while taking a class on the neuroscience of learning in late 2021, I became aware of people's worries surrounding AI. It was at this point I shifted from simply having an intellectual curiosity in these topics to caring about ensuring the safe development of AI and the equitable cognitive empowerment of people.

I had inconsistent grades at Williams College, which I eventually discovered was due to mild Autism and have since found that with basic accommodations I am no longer held back by it, as evidenced by my research productivity. This, combined with The Williams Curve (1-2 As per class), meant I only got a GPA of 3.1 (even worse, I got 2.99 because I forgot to remove a class I dropped).

Since graduating from Williams College, I have been working at the Carney Institute for Brain Science at Brown initially under Professor Sebastian Musslick, and now under Professor Michael J. Frank. In those 2.5 years, I was able to get extensive experience in research across a range of disciplines, ranging from closed loop behavioral research design to reinforcement learning, presenting at 5 conferences/workshops, publishing 6 peer-reviewed papers, with 3 more currently under review, (and 2 more nearing completion), with a total of 5 (+2) as lead author. In that time my team won first place in the Carney Institute's Brainstorm Modeling Challenge and the 2nd Annual Brainstorm Challenge.

During the pandemic, I also explored complexity science through courses at the Santa Fe Institute, studying emergence and agent-based modeling. This multidisciplinary lens helped me understand cognition as an emergent property of neural computations, especially when applied to large-scale systems like LLMs and the brain. This reinforced my view that cognition extended beyond reductionism, requiring insights from nondeterministic approaches to be computationally modeled.

Through all this, I came to appreciate how difficult it was to reverse engineer the brain from neural signals, how important computational modeling is in creating testable hypotheses about human cognition, and how remarkably emergent human and machine cognitive capabilities are. Combined with the realization of just how fast AI was developing, I concluded that the most efficient path to understanding and harnessing cognition would involve mapping the architectural and algorithmic biases of the brain onto machine learning models across a wide range of scales. I believe such an approach addresses the emergent nature of cognition, allowing us to explain cognitive capabilities that arguably emerge from processes that arise as the model scale increases. I also believe this allows us to apply knowledge from social psychology and neuroscience–about how prosocial behavior arises and operates–to the AI models that are becoming increasingly autonomous. And finally, I believe this allows us to better translate between AI and human cognition, so that the technology currently being developed can more seamlessly fit with how we think and what we value.

I am currently exploring this by determining whether social and emotional reasoning can be enhanced by explicitly disentangling computations related to sensory processing and decision-making. My research plan is to implement, test, and continue developing a cognitively-inspired transformer model that makes a first attempt at this by separating heads that are trained using reinforcement learning from those trained with next token prediction, while allowing heads in both groups to access the layer-wise outputs of the other. I have written a preprint that goes into more detail about this. My research goals are still in their early stages, and I expect them to change as I learn from and adapt to my next research lab.

I was struck by the interdisciplinary focus and leading faculty in socially aware and human-centered AI at Columbia, which aligns well with my goals to develop cognitively inspired AI systems that understand and engage empathetically with humans. I am particularly excited about the opportunity to work with Professor Zhou Yu, whose expertise in designing intelligent interactive systems that blend multimodal sensing with socially-aware AI aligns perfectly with my interest in creating AI that engages naturally and empathetically with humans. Her focus on non-verbal behaviors and conversational AI complements my vision of socially-aligned cognitive systems. Furthermore, I am inspired by Professor Jeannette Wing's dedication to developing trustworthy AI, as her emphasis on robustness, fairness, and interpretability directly intersects with my commitment to designing AI that is safe, reliable, and values-driven. Columbia's strong emphasis on interdisciplinary collaboration and its leadership in advancing both technical and societal dimensions of AI make it an ideal environment for me to grow as a researcher and contribute meaningfully to the field.

Alongside my research, I have also founded CyberKind, an advocacy group focused on the development of compassionate AI and the equitable empowerment of people through AI and neurotechnology. I envision Columbia as a place where my academic work will not only advance my research goals but also strengthen my ability to do this advocacy work.

My long-term vision is to foster a symbiotic relationship between humans and AI, where both benefit and grow together. I aspire to help develop AI systems that can understand and empathize with humans, with the hope that this will help with value alignment. I plan to apply my research to benefit both academia and broader society, developing technologies that empower individuals and foster a collaborative future between humans and AI. I am excited by the prospect of engaging with the collaborative and interdisciplinary environment at Columbia, as I believe it is the ideal setting to grow as a researcher and advocate for a kinder, fairer future.