

# Adam Morris

thatadamorris.com

thatadamorris@gmail.com | 908.642.1536 | 4908 Cedar Ave, Philadelphia, PA

## EDUCATION

### HARVARD UNIVERSITY

PHD IN COGNITIVE SCIENCE

May 2022 | Cambridge, MA

GPA: 4.0

### BROWN UNIVERSITY

BS IN PSYCHOLOGY WITH HONORS

MAGNA CUM LAUDE

May 2015 | Providence, RI

GPA: 4.0

## COURSEWORK

### GRADUATE

Computational cog. sci.

Statistics

### UNDERGRADUATE

Machine learning

Computational prob. & stats

Computational cog. neuro

Game theory

Linear algebra

Differential equations

## SKILLS

### RESEARCH

Leading projects • analyzing papers / conducting lit reviews • designing and executing experiments • writing scientific publications

### DATA SCIENCE

Data cleaning • visualization • analysis • computational modeling

### PROGRAMMING

R • Python • MATLAB • Javascript

## AWARDS

Kirschstein-NRSA F32 NIH Fellowship (2022 - present)

Certificate of Teaching Distinction, Harvard Bok Center (2020, 2018, 2017)

Prize for Best Student Paper, Society for Philosophy and Psychology (2018)

George Goethals Teaching Prize (2018)

National Defense Science & Engineering Graduate Fellowship (2017 - 2021)

Harvard Presidential Scholarship (2015)

Magna Cum Laude, Phi Beta Kappa (2015)

## VALUE ADDED

- I bring **deep technical ability** in research & data science alongside **exceptional soft skills** for collaborating, writing, and communicating.
- I **pick up new skills/knowledge extremely quickly**, and quickly grok the underlying logical structure of ideas, research literatures, and projects.
- I have **extensive experience leading & collaborating on complex research projects**, from conception to publication.
- I **doggedly pursue creative solutions** to technical & social obstacles.

## RESEARCH EXPERIENCE

### PRINCETON UNIVERSITY | POSTDOCTORAL RESEARCH FELLOW

July 2022 - present | Crockett Lab, Princeton, NJ

- Investigating people's ability to introspect on their own choice mechanisms, and whether introspection can be improved through attentional training.
- Developed rigorous methods for quantifying introspective accuracy, using computational modeling to characterize people's choice mechanisms.
- Received a Kirschstein-NRSA F32 NIH fellowship to support this research, and an Editor's Choice award from the American Psychological Association.

### HARVARD UNIVERSITY | DOCTORAL STUDENT

Sep 2015 - May 2022 | Moral Psychology Research Lab, Cambridge, MA

- Studied the algorithms underlying human decision making, with a focus on reinforcement learning models. Won several awards for this research, including an NDSEG Fellowship and a Best Student Paper prize.
- Led numerous successful research projects. Proposed new models of decision making, grounded in machine learning and evolutionary game theory; designed, programmed, and ran experiments to test these models.
- Developed numerous computational models as part of these projects, and fit those models to empirical data using contemporary Bayesian methods.
- Published results in top journals, with >15 papers and >700 citations.

## AI-RELATED EXPERIENCE

### FAITHFUL INTROSPECTION IN LLMs | INDEPENDENT PROJECTS

Fall 2024 - present

- Investigating the faithfulness of LLMs' reports about their own internal processes in chain-of-thought reasoning and post hoc reports. Applying methods I developed in my research on humans.
- Found that frontier LLMs can at baseline report complex internal processes with moderate accuracy, and can be trained to report them more accurately.

### AI SAFETY FUNDAMENTALS COURSE | BLUE DOT IMPACT

Spring 2023

- Completed BlueDot's AI Safety Fundamentals course, technical track.
- Read and discussed numerous papers on contemporary machine learning models, mechanistic interpretability research, RLHF, and scalable oversight.

## SELECT PUBLICATIONS

1. Cushman & Morris (2015). Habitual control of goal selection in humans. *PNAS*.
2. Morris et al. (2017). Evolution of flexibility & rigidity in punishment. *PNAS*.
3. Morris & Cushman (2019). Model-free RL or action sequences? *Frontiers*.
4. Morris, A. (2024). Invisible gorillas in the mind. *PsyArXiv*.