

# Guy Davidson

guy.davidson@nyu.edu | +1 415-510-9167 | <https://github.com/guydav> | <https://guydavidson.me>

I am an artificial intelligence and cognitive science researcher excited about understanding the human mind and leveraging ideas from cognition to develop more human-like artificial intelligence. In my work, I focus on identifying computational problems people solve, developing models to tackle the same tasks, and applying insights from modeling humans to understanding and advancing artificial methods. I am particularly interested in goal generation and structured, program-like goal representations. I have also led work on cognitively-inspired machine learning and contributed to more theoretical psychology efforts.

## EDUCATION

### NYU CENTER FOR DATA SCIENCE

Ph.D. candidate

New York, NY

2019–Present

Advised by Professors Brenden Lake and Todd Gureckis, my thesis research attempts to advance the understanding of goals in human cognition. My work proposes a framework to represent goals as reward-producing programs and supports this idea by combining evidence from rich human experiments and computational models. My research offers a novel treatment of goals in cognitive science and aims to contribute to the development of richer goal representations for artificial agents.

### MINERVA UNIVERSITY

BSc in Computational Sciences

San Francisco, CA

2015–2019

I graduated summa cum (GPA: 3.98/4.0) with a concentration in Machine Learning. In my capstone project, I investigated the scaling behavior of different meta-learning algorithms: how quickly new tasks are learned as a function of previous training, using a novel benchmark paradigm inspired by visual question answering (externally advised by Mike Mozer, published at CVPR 2020).

## PUBLICATIONS AND PRESENTATIONS

**Davidson, G.,** Todd, G., Togelius, J., Gureckis, T. M., Lake, B. M. (2024). Goals as Reward-Producing Programs. *Accepted in principle, Nature Machine Intelligence*: <https://arxiv.org/abs/2405.13242>

**Davidson, G.,** Gureckis, T. M. (2024). Toward Complex and Structured Goals in Reinforcement Learning. *Finding the Frame Workshop @ RLC 2024 (Oral)*.

**Davidson, G.,** Orhan, A. E., Lake, B. M. (2024). Spatial Relation Categorization in Infants and Deep Neural Networks. *Cognition*.

Sharma, S. **Davidson, G.,** Khetarpal, K., Kanervisto, A., Arora, U., Hofmann, K., Momennejad, I. (2024). Toward Human-AI Alignment in Large-Scale Multi-Player Games. *Wordplay: When Language Meets Games Workshop @ ACL 2024*.

**Davidson, G.,** Todd, G., Gureckis, T. M., Togelius, J., Lake, B. M. (2023). Generating Human-Like Goals by Synthesizing Reward-Producing Programs. *Intrinsically Motivated Open-ended Learning Workshop @ NeurIPS 2023*.

**Davidson, G.,** Gureckis, T. M., Lake, B. M. (2022). Creativity, Compositionality, and Common Sense in Human Goal Generation. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society, CogSci 2022*.

Bennett, D., **Davidson, G.,** Niv, Y. (2022). A model of mood as integrated advantage. *Psychological Review*.

**Davidson, G.,** Lake, B. M. (2021). Examining Infant Relation Categorization Through Deep Neural Networks. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society, CogSci 2021*.

**Davidson, G.,** Lake, B. M. (2020). Investigating Simple Object Representations in Model-Free Deep Reinforcement Learning. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society, CogSci 2020*.

**Davidson, G.,** Lake, B. M. (2020). Systematically Comparing Neural Network Architectures in Relation Learning. *Object-Oriented Learning (OOL): Perception, Representation, and Reasoning Workshop @ ICML 2020*.

**Davidson, G.,** Mozer, M. C. (2020). Sequential mastery of multiple visual tasks: Networks naturally learn to learn and forget to forget. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

**Davidson, G.,** Radulescu, A., Niv, Y. (2019). Contrasting the effects of prospective attention and retrospective decay in representation learning. *The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making*.

Bennett, D., **Davidson, G.,** Niv, Y. (2019). Momentum and mood in policy-gradient reinforcement learning. *The 4th Multidisciplinary Conference on Reinforcement Learning and Decision Making*.

## ADDITIONAL RESEARCH EXPERIENCE

### META FUNDAMENTAL AI RESEARCH

Visiting Researcher

New York, NY

09/2024–Present

I am visiting Meta's Fundamental AI Research (FAIR) through the Meta-NYU AI Mentorship program. I evaluate LLMs' ability to infer accurate task specifications and user goals from ambiguous instructions. I also develop methods to improve LLM goal inferences and expose latent inferences made by the models. Mentored by Adina Williams in the Alignment team.

## MICROSOFT RESEARCH

### Research Intern

New York, NY  
05/2022-08/2022

Developed methods inspired by the cognitive psychology concept of task-sets (abstract task representations) to analyze and predict behavior in a large-scale gameplay dataset in a multiplayer game. Initial results highlighted consistent differences between players by their propensity to flee or attack in fight-or-flight scenarios. Mentored by Ida Momennejad and Harm van Seijen.

## PRINCETON NEUROSCIENCE INSTITUTE

### Research Intern

Princeton, NJ  
05/2018-08/2018

Interned with Professor Yael Niv's lab to investigate human reinforcement learning (RL) in multidimensional environments. Modeled data from previous experiments to examine dissociable roles of attention and decay in human RL and the efficacy of eye-tracking and fMRI-based attention measures (presented at RLDM 2019). Developed a simulation environment for bandit problems to motivate work framing mood as a momentum variable (presented at RLDM 2019 and published in Psych Review).

## SELECTED INVITED TALKS

CoLab, Princeton University	05/2024
Invited talk, Cognitive Control of Action Workshop, Princeton University	03/2024
CoCoDev Lab, Harvard University	02/2024
CoCoSci Lab, Massachusetts Institute of Technology	02/2024

## TEACHING

### DS-GA 1016: COMPUTATIONAL COGNITIVE SCIENCE

NYU

#### Section leader, grader

01/2021-05/2021, 01/2022-05/2022, 09/2024-12/2024

Served as a section leader and grader (twice) and lead teaching assistant (once) for a graduate-level course on computational approaches to modeling cognition.

### DS-UA 112: INTRODUCTION TO DATA SCIENCE

NYU

#### Section leader

09/2019-12/2019

Served as a section leader for new undergraduate course introducing students to NYU's newly approved Data Science major.

## PROFESSIONAL EXPERIENCE

### AIDOC MEDICAL

Tel Aviv, Israel

#### Research Engineer

05/2017-11/2017

- Implemented research-supporting tools in Azure cloud environment to facilitate and expedite deep learning experimentation, reducing idea-to-experiment turnaround 10x from 1-2 hours to 5-10 minutes.
- Explored deep learning frameworks supporting distributed training, performed experiments and cost-benefit analysis leading to prioritizing tool development.

### AMAZON PRIME AIR

Seattle, WA

#### Software Development Engineer Intern

05/2016-08/2016

- Implemented binary image loading solution for proprietary hardware and software platform using C and assembly.
- Contributed to serialization library used in multiple projects across the Prime Air development group.

### MINERVA PROJECT

San Francisco, CA

#### Software Development Engineer Intern

01/2016-05/2016

- Designed and delivered overhaul of invoicing system, transitioning from a fixed to a line-itemized implementation, to allow for improved flexibility, history tracking, and increased robustness, in a Python/Django/MySQL web-stack.

### SIMILARWEB

Tel Aviv, Israel

#### Software Engineer

06/2015-09/2015

- Integrated products with external partners, debugged and solved multiple production issues, in Node.JS, iOS, and Python.

### ISRAEL DEFENSE FORCES INTELLIGENCE BRANCH

Israel

#### Team leader, training instructor, software engineer

08/2008-10/2014

- Team leader: grew team of four software developers to ten, responsible for developing tools to enhance analysis capabilities and solve production issues. Reduced turnaround time more than 2x, contributed to Israel Defense Prize-winning project.
- Instructor: managed two other instructors, training ten pupils in software engineering and computer networking.
- Software engineer: built in-house data ingestion pipelines and analysis tools. Developed primarily in Python and Java and cultivated debugging and fault analysis expertise.

## SERVICE

### **INTRINSICALLY MOTIVATED OPEN-ENDED LEARNING WORKSHOP**

#### **Co-organizer**

Served as one of the organizers of the 2024 workshop on Intrinsically Motivated Open-ended Learning @ NeurIPS 2024

NeurIPS  
12/2024

### **REVIEWING**

Served as a reviewer for Cognitive Science, Advances in Neural Information Processing Systems (NeurIPS), the Annual Conference of the Cognitive Science Society (CogSci), and several conference workshops.

## SHORT COURSES ATTENDED

### **BRAINS, MINDS, AND MACHINES SUMMER COURSE**

Hosted by the MIT and Harvard Center for Brains, Minds, and Machines

Woods Hole, MA  
08/2021

### **MACHINE LEARNING SUMMER SCHOOL**

Hosted by University College London and Imperial College London

London, England  
07/2019

## SKILLS AND TECHNICAL COMPETENCIES

- Neural network research and development (mostly PyTorch)
- Data science and analysis (Numpy, Pandas, Sklearn, Matplotlib)
- Software engineering, architecture design, and testing
- Human participant research experiment design
- Web experiment development (React, vue, tailwind, Firebase)
- Research environment and stimulus design (Unity, Blender)