

Cameron Robert Jones

Department of Cognitive Science, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0515
(+1) (858) 319-6265 ◆ camrobjones.com ◆ cameron@ucsd.edu

Research Positions

Postdoctoral Scholar

2024 – Present

Dept. of Cognitive Science, UC San Diego

PI: Benjamin K. Bergen

Evaluating persuasion and deception as dangerous capabilities in Large Language Models in interactive conversational settings.

Graduate Researcher, Language and Cognition Lab

2019 – 2024

Dept. of Cognitive Science, UC San Diego

PI: Benjamin K. Bergen

Investigating social intelligence in large language models, including theory of mind benchmarks, and interactive evals like the Turing test.

Dissertation Research, Centre for Language Evolution

2016-2017

Philosophy, Psychology & Language Sciences, University of Edinburgh

PI: Simon Kirby

Design and analysis of in-person experiments investigating how induced biases influence information propagation.

Education

UC San Diego

2019 – 2024

PhD, Cognitive Science (GPA: 4.00)

Advisor: Benjamin K. Bergen

Dissertation: Reading Minds: Social Intelligence and Large Language Models

University of Edinburgh

2016 – 2017

MSc, Evolution of Language and Cognition (Distinction)

Dissertation Advisor: Simon Kirby

Dissertation: The Effect of Biasing Information on a Transmission Chain of Short Texts

University College London

2012 – 2016

BA, Classics (First Class)

Grants and Funding

AI Persuasiveness Evaluation

2024 – 2026

[Open Philanthropy](#)

\$470,731

PI: Benjamin K. Bergen

I was the primary author of the grant, which is being used to fund my postdoctoral research on persuasion and deception with LLMs.

Public Online Turing Test

2024

[Manifund](#)

\$2,000

A crowdfunded grant to extend research on the Turing test.

Publications

Under Review. Rathi, I., Taylor, S., Bergen, B., & **Jones, C. R.** GPT-4 is judged more human than humans in displaced and inverted Turing tests. [\[preprint\]](#)

Under Review. Pi, Z., Vadaparty, A., Bergen, B., & **Jones, C. R.** Dissecting the Ullman Variations with a SCALPEL: Why do LLMs fail at Trivial Alterations to the False Belief Task?. [\[preprint\]](#)

Under Review. **Jones, C. R.** & Bergen, B. People cannot distinguish between GPT-4 and a human in a Turing test. [\[preprint\]](#)

2024. **Jones, C. R.**, Bergen, B., & Trott, S.. Do Multimodal Large Language Models and Humans Ground Language Similarly? *Computational Linguistics* [\[paper\]](#)[\[code\]](#)

2024. **Jones, C. R.**, Trott, S., & Bergen, B. Comparing Humans and Large Language Models on an Experimental Protocol Inventory for Theory of Mind Evaluation (EPITOME). *Transactions of the Association of Computational Linguistics* [\[paper\]](#)[\[code\]](#)

2024 **Jones, C. R.** & Trott, S.. Multimodal Language Models Show Evidence of Embodied Simulation. *LREC-COLING 2024* [\[paper\]](#)

2024. **Jones, C. R.** & Bergen, B. Does GPT-4 Pass the Turing Test? *NAACL* [\[paper\]](#)
[\[experiment\]](#)

2024. **Jones, C. R.** & Bergen, B. Does word knowledge account for the effect of world knowledge on pronoun interpretation? *Language and Cognition* [\[paper\]](#)

2023. Trott, S.*, **Jones, C. R.***, Michaelov, J. A., Chang, T. A., & Bergen, B. Do Large Language Models know what humans know? *Cognitive Science* (Vol. 47, No. 7) [*co-first author] [\[paper\]](#) [\[code\]](#)

2022. **Jones, C. R.**, Chang, T. A., Coulson, S., Michaelov, J. A., Trott, S., & Bergen, B. Distributional Semantics Still Can't Account for Affordances. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 44, No. 44). [\[paper\]](#) [\[code\]](#)

2021. **Jones, C. R.**, & Bergen, B. The Role of Physical Inference in Pronoun Resolution. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43). [\[paper\]](#) [\[code\]](#)

2021. Binder, F. J.*, **Jones, C. R.***, Kaufman, R. A., Lin, N. T., Poole, C. R., & Vul, E. Cognitive cost and information gain trade-off in a large-scale number guessing game. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43). [*co-first author] [\[paper\]](#) [\[code\]](#)

2018. **Jones, C.** & Kirby, S. The Effect of Biasing Information on a Transmission Chain of Short Texts. *2nd Conference of the Cultural Evolution Society*, Arizona, USA.

Talks, Workshops, & Posters

2024. **Jones, C. R.**, Trott, S., & Bergen, B. Comparing Humans and Large Language Models on an Experimental Protocol Inventory for Theory of Mind Evaluation (EPITOME). *Oral & Poster Presentation, ACL 2024*

2024. **Jones, C. R.**, Trott, S., & Bergen, B. Does reading words help you to read minds?. *Oral Presentation, Cognitive Science Society Conference 2024* [\[paper\]](#)

2024. **Jones, C. R.** & Bergen, B. Does GPT-4 Pass the Turing Test? *Oral presentation, NAACL 2024* [\[paper\]](#) [\[experiment\]](#)

2024 **Jones, C. R.** & Trott, S.. Multimodal Language Models Show Evidence of Embodied Simulation. *Poster session, LREC-COLING 2024* [\[paper\]](#)

2023. EPITOME: Experimental Protocol Inventory for Theory Of Mind Evaluation, *Poster Session, Workshop on Theory-of-Mind at ICML 2023* [\[paper\]](#)

2023. Does Matrix Multiplication Have a Theory of Mind?, *Comparative Cognition Lab, Dept of Cognitive Science, UC San Diego*

2023. Language Models for Psycholinguistic Analysis, *Brain and Cognition Lab, Dept of Cognitive Science, UC San Diego* [[notebook](#)]

2022. Language Comprehension Requires Affordances, with Arthur Glenberg. *Workshop on Embodied, Situated and Grounded Intelligence: Implications for AI. The Santa Fe Institute* [[recording](#)]

2022. Distributional Semantics Still Can't Account for Affordances. *Cognitive Science Society Conference 2022*

2022. Situation Modelling in Humans and Machines, with Ronen Tamari. *Language and Vision Workshop. Cognitive Tools Lab, Dept of Psychology, UC San Diego*

2022. Can We Know Words by the Company They Keep? Guest Lecture on Distributional Methods in Computational Linguistics and their Limits. *Cognitive Science, Ilia State University, Georgia*

2022. Working with Language Models in Python. *Methods Training Assistant Workshop, Dept of Cognitive Science, UC San Diego* [[notebook](#)]

2022. Distributional Semantics Still Can't Account for Affordances. *Brain and Cognition Lab, Dept of Cognitive Science, UC San Diego*

2021. The Role of Physical Inference in Pronoun Resolution. *Poster Session, Cognitive Science Society Conference 2021*

2021. Cognitive cost and information gain trade-off in a large-scale number guessing game. *Poster Session, Cognitive Science Society Conference 2021*

2021. World Knowledge Influences Pronoun Resolution Both Offline and Online. *Centre for Research and Language, UC San Diego*

Media

2024. Understanding, Grounding, and Reference in LLMs, with Sean Trott. *The Gradient Podcast* [[link](#)]

2023. It takes a body to understand the world, with Arthur Glenberg. *The Conversation* [[link](#)]

2023. Does GPT-3 Have a Theory of Mind?, *Cognitive Science Society Blog* [[link](#)]

Awards

Diversity Award Nominee	2024
<i>Dept. of Cognitive Science, UC San Diego</i>	\$100
Summer Graduate Teaching Scholarship	2019-2023
<i>UC San Diego</i>	\$1200
Glushko Travel & Research Award	2019-2023
<i>Dept. of Cognitive Science, UC San Diego</i>	\$500 p.a.
Highly Commended Dissertation Award	2017
<i>School of Philosophy, Psychology & Language Sciences, Edinburgh University</i>	

Teaching Experience

Instructor of Record

COGS 9: Introduction to Data Science	Summer 2023
--------------------------------------	-------------

Teaching Assistant

COGS 3: Introduction to Computing	Bardolph, M.	Winter 2020
COGS 3: Introduction to Computing	Bardolph, M.	Spring 2020
COGS 11: Minds & Brains	Boyle, M.	Summer 2020
COGS 15: Uncensored Intro. to Language	Bergen, B. K.	Fall 2020
COGS 102A: Cognitive Perspectives	Johnson, C.	Winter 2021
COGS 100: Cyborgs Now and in the Future	Allen, M.	Summer 2021
COGS 187A: Web Design and Info Architecture	Kirsh, D.	Fall, 2021
COGS 187B: Pro Practicum in Web Design	Kirsh, D.	Winter, 2022
COGS 9: Introduction to Data Science	Shannon, K.	Summer, 2022
DSGN 1: Design of Everyday Things	Meyer, M.	Fall, 2022
COGS 152: Cognitive Science of Mathematics	Nuñez, R.	Winter, 2023
COGS 153: Language Comprehension	Trott, S.	Fall, 2023
COGS 150: Large Language Models	Trott, S.	Winter, 2024
COGS 112: Humor	Walker, D.	Spring, 2024

Academic Service

Graduate Colloquium Organizer 2023 – 2024

Dept. of Cognitive Science, UC San Diego

Organizing departmental events and speaker series, including a debate: *Do Large Language Models understand language?*

Ad Astra Organizer 2022 – 2024

Dept. of Cognitive Science, UC San Diego

Organizing a weekly graduate student meeting to share in-progress research and get constructive feedback from peers.

Graduate Student Mentor 2023 – 2024

Dept. of Cognitive Science, UC San Diego

Leading a mentor pod to foster community and support less experienced graduate students.

Graduate Student Representative 2022 – 2023

Dept. of Cognitive Science, UC San Diego

Represent the interests of graduate students through managing a \$10,000 budget, organizing colloquia, and liaising with faculty.

Methods Training Assistant 2021 – 2022

Dept. of Cognitive Science, UC San Diego

Weekly office hours and quarterly workshops to share information about programming, experimental design, and statistical analysis.

Webmaster 2018 – Present

Cultural Evolution Society

Maintaining the society's website [culturalevolutionsociety.org].

Ex-officio non-voting executive committee member.

Employment History

Data Engineer 2018 – 2019

AtomLeap, Berlin

Full-stack web development of natural language processing tools for competitive intelligence (data extraction, search, report generation).

Data Analytics Consultant 2017-2018

EY, Edinburgh

Development of natural language processing tools to support internal audit teams; data analytics in SQL to support financial audits.

Intern 2015

Wellcome Trust, London

Analysing the efficacy of the Trust's £6m *Neuroscience and Education* Initiative.

Intern 2013-2014

Houses of Parliament, London

Parliamentary Intern to Robert Halfon MP. Research for several campaigns including reducing hospital parking costs and energy bills.

Projects

Turing Test [turingtest.camrobjones.com] 2023

Interactive Turing Test implemented using GPT-4.

Paperbot [<https://github.com/camrobjones/paperbot>] 2023

CLI tool implementing Retrieval Augmented Generation (RAG) to query local repositories of academic paper PDFs using LLMs.

Hackercast [camrobjones.com/hackercast] 2023

Automated generation of a podcast summary of the top articles on Hacker News using GPT-4 and AWS Polly.

Point the Compass [camrobjones.com/pointthecompass] 2021

Web application to gather norms on the political compass. Users vote on the placement of topics along two political axes. Django, Vue.js

Spyke [camrobjones.com/spyke] 2020

Education and research web application to help users build interactive simulations of spiking neural networks. Django, Vue.js, NEURON

Genetic Algorithms [genetic-algorithms.com] 2018

Generate art using a population of agents that evolve according to a genetic algorithm, where fitness is based on user preferences. Django.