

# Cameron Robert Jones

---

Department of Psychology, Stony Brook University, Nicolls Rd, Stony Brook, NY 11974  
(+1) (858) 319-6265 ◆ [camrobjones.com](http://camrobjones.com) ◆ [cameron.jones@stonybrook.edu](mailto:cameron.jones@stonybrook.edu)

## Academic Positions

**Assistant Professor** 2025 – Present

*Psychology Department, Stony Brook University (SUNY)*

Director of the Cognition, Language, Interaction, and Computation Lab. Leading research at the intersection of psychology and AI.

**Postdoctoral Scholar** 2024 – 2025

*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.

## 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

**Code of Practice Compliance Monitoring (Harmful Manipulation)** 2026  
\$1,200,000

[EU AI Office](#)

Consortium member via Equistamp. Technical assistance for risk assessment of general-purpose AI models.

**LLM Persuasion and Theory of Mind** 2025  
\$30,000

[Diverse Intelligences Summer Institute](#)

Investigating what role theory of mind plays in naturalistic persuasion, and evaluating humans and LLMs in both domains.

**AI Persuasiveness Evaluation** 2024 – 2026  
\$470,731

[Open Philanthropy](#)

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  
\$2,000

[Manifund](#)

A crowdfunded grant to extend research on the Turing test.

## Publications

**Under Review.** Jones, C. R. & Bergen, B. Large Language Models Pass the Turing Test. [\[preprint\]](#)[\[data\]](#)

**Under Review.** Jones, C. R., Lombardi, A., Mahowald, K., & Bergen, B. K.. LLMs and people both learn to form conventions--just not with each other. [\[preprint\]](#)

**Under Review.** Moore, J., Overmark, R., Cooper, N., Cibralic, B., Haber, N., & Jones, C. R.. Large Language Models Persuade Without Planning Theory of Mind [\[preprint\]](#)

**Under Review.** Trott, S., Taylor, S., Jones, C., Michaelov, J. A., & Rivière, P. D. Language Statistics and False Belief Reasoning: Evidence from 41 Open-Weight LMs. [\[preprint\]](#)

**Under Review.** Schoenegger, P., Salvi, F., Liu, J., Nan, X., Debnath, R., Fasolo, B., Jones, C.R., ... & Karger, E. Large Language Models Are More Persuasive Than Incentivized Human Persuaders. [\[preprint\]](#)

2026. Bengio, Y., Clare., S., Prunkl, C.... **Jones, C.R.**, et al. International AI Safety Report. [\[paper\]](#)

2025. Bengio, Y., Clare., S., Prunkl, C.... **Jones, C.R.**, et al. First Key update to the International AI Safety Report. [\[paper\]](#)

2025. Moore, J., Cooper, N., Overmark, R., Cibralic, B., Haber, N., & **Jones, C. R.** Do Large Language Models Have a Planning Theory of Mind? Evidence from MindGames: a Multi-Step Persuasion Task, *2nd Conference on Language Modeling* [\[preprint\]](#)

2025. 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? *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47). [\[paper\]](#)

2025.. Rathi, I., Bergen, B., & **Jones, C. R.** Judging the Judges: Displacing and Inverting the Turing test to Investigate the Interrogator. *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47). [\[paper\]](#)

2025. Riviere, P. D., Parkinson-Coombs, O., **Jones, C. R.**, & Trott, S. Does Language Stabilize Quantity Representations in Vision Transformers? *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 47). [\[paper\]](#)

2025. **Jones, C. R.** & Bergen, B. People cannot distinguish between GPT-4 and a human in a Turing test. *ACM Conference on Fairness, Accountability, and Transparency 2025 (FAccT 2025)* [\[paper\]](#)

2025. Rathi, I., Taylor, S., Bergen, B., & **Jones, C. R.** GPT-4 is judged more human than humans in displaced and inverted Turing tests. *Workshop on Detecting AI Generated Content at COLING 2025* [\[preprint\]](#) **[Best Paper Award]**

2025 Schoenegger, P., **Jones, C. R.**, Tetlock, P. E.,& Mellers, B. Prompt Engineering Large Language Models' Forecasting Capabilities [\[preprint\]](#)

2026. **Jones, C. R.** & Bergen, B. Lies, Damned Lies, and Distributional Language Statistics: Persuasion and Deception with Large Language Models, *Artificial Intelligence Review* [\[paper\]](#)

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

2026. LLMs as Simulacra of Human Social Reasoning. *Invited Talk, Social Reasoning and the Ecology of Thought Workshop, IVADO, Montréal.*

- 2026.** LLMs as Simulacra of Human Social Reasoning. *Invited Talk, Stanford NLP Group, Stanford, CA.*
- 2025.** Dissecting the Ullman Variations with a SCALPEL: Why do LLMs fail at Trivial Alterations to the False Belief Task? *Poster (co-presented with Zhiqiang Pi), Annual Meeting of the Cognitive Science Society 2025.*
- 2025.** Judging the Judges: Displacing and Inverting the Turing test to Investigate the Interrogator. *Poster (co-presented with Ishika Rathi), Annual Meeting of the Cognitive Science Society 2025.*
- 2025.** People cannot distinguish between GPT-4 and a human in a Turing test. *Oral Presentation, ACM Conference on Fairness, Accountability, and Transparency 2025 (FAccT 2025).*
- 2025.** Risks from Persuasion and Deception from Large Language Models. *International Association for Safe and Ethical AI 2025.*
- 2024.** Comparing Humans and Large Language Models on an Experimental Protocol Inventory for Theory of Mind Evaluation (EPITOME). *Oral & Poster Presentation, ACL 2024.*
- 2024.** Does reading words help you to read minds?. *Oral Presentation, Annual Meeting of the Cognitive Science Society 2024.*
- 2024.** Does GPT-4 Pass the Turing Test? *Oral presentation, NAACL 2024*
- 2024** Multimodal Language Models Show Evidence of Embodied Simulation. *Poster session, LREC-COLING 2024.*
- 2023.** EPITOME: Experimental Protocol Inventory for Theory Of Mind Evaluation, *Poster Session, Workshop on Theory-of-Mind at ICML 2023.*
- 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. *Annual Meeting of the Cognitive Science Society 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, Annual Meeting of the Cognitive Science Society 2021*.

2021. Cognitive cost and information gain trade-off in a large-scale number guessing game. *Poster, Annual Meeting of the Cognitive Science Society 2021*.

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

## Media

2024. AI passed the Turing Test -- And No One Noticed. *Sabine Hossenfelder* [[link](#)]

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

2023. 1960s chatbot ELIZA beat OpenAI's GPT-3.5 in a recent Turing test study. *Ars Technica* [[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

Best Paper Award

2025

*Workshop on Detecting AI Generated Content @ COLING 2025*

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

## Teaching Experience

### Instructor of Record

COGS 9: Introduction to Data Science	<i>Summer 2023</i>
PSY 201: Statistical Methods in Psychology	<i>Spring 2026</i>
PSY 365: Psychology of Language	<i>Spring 2026</i>

### Teaching Assistant

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

## Mentorship

### Current Students

**John Stallings**, MA student (2025–present). Supervising project on AI-induced psychosis.

### Research Assistants

I have provided mentorship to more than 10 research assistants, many of whom have led their own research projects and published first author papers (including Ishika Rathi and Zhiqiang Pi).

### High School Students

Katherine Zeng (2025–present). Mentoring independent research project on human detection of AI-generated images. Project selected for Regeneron Science Talent Search competition.

## Academic Service

### Reviewing

2019 – Present

Over 40 manuscripts across journals including *Cognitive Science*, *Computational Linguistics*, *Nature Scientific Reports*, *Nature Humanities & Social Sciences Communications*, *Cognitive Research: Principles and Implications*; and conferences including the *Cognitive Science Society* conference and the *Conference on Computational Linguistics (COLING)*.

### Grievance Committee

2025 – Present

*Dept. of Psychology, Stony Brook University*

Serving on a departmental committee to address faculty and student concerns.

### Workshop Co-Organizer

2026

*AI, Manipulation, & Information Integrity (AIMII) Workshop, International Association for Safe and Ethical AI (IASEAI), UNESCO House, Paris.*

### 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](http://culturalevolutionsociety.org)].

Ex-officio non-voting executive committee member.

**Non-Academic Positions**

**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.