

RACHEL C. METZGAR

rachel.metzgar@princeton.edu | rachelmetzgar.psychology.princeton.edu | linkedin.com/in/rachelmetzgar

EDUCATION

- Ph.D. in Psychology**, Princeton University *expected: 2026
Adviser: Michael S.A. Graziano
Dissertation: Attributions of Mind and Consciousness
- M.A. in Psychology**, Princeton University 2023
- B.S. with Distinction in Neuroscience and B.A. in Biological Sciences**, University of Delaware 2017

PUBLICATIONS, PRESENTATIONS, AND TALKS

[Google Scholar](#)

Manuscripts in Preparation

- Metzgar, R.C., Christian, I.R., Morey, R., & Graziano, M.S.A. (Under Review). Behavioral and neural responses to human- and AI-labeled partners during naturalistic conversation.
- Metzgar, R.C. (In preparation). Folk Psychological Structure of Mind Predicts Moral Judgments in Large Language Models.
- Metzgar, R.C. (In preparation). Behavioral and Representational Distinctions Between Human and AI Conversation Partners in Large Language Models.

Publications

- Metzgar, R.C. & Graziano, M.S.A. (Accepted; 2026). Comparing LLM and human responses to human- and AI-labeled partners during naturalistic conversation. *Proceedings of the Cognitive Science Society*.
- Metzgar, R.C. (Accepted; 2026). Folk Psychological Structure of Mind in Large Language Models. *Conference on Cognitive Computational Neuroscience*, New York, NY.
- Oakman, J., Rexford, J., Metzgar, R.C., & d'Oleire Uquillas, F. (2026). A regional and data-centered approach to workforce development in an AI economy. In J. M. Lavista Ferres (Ed), *Degrees of change: What AI means for education and the next generation*. Wiley.
- Metzgar, R., Lombrozo, T., & Graziano, M.S.A. (2025). Lay perspectives on the physical and non-physical nature of consciousness. *Journal of Consciousness Studies*, 32(11–12), 36–71.
- Christian, I.R., Graziano, M., Metzgar, R., & Nastase, S. (2024). Cortical semantic encoding varies with attention and habituation. *Conference on Cognitive Computational Neuroscience*, Boston, MA.
- Metzgar, R., Stoll, H., Grafton, S.T., Buxbaum, L.J., & Garcea, F.E. (2022). Single-case disconnectome lesion-symptom mapping: Identifying two subtypes of limb apraxia. *Neuropsychologia*, 170, 108210.

Invited Talks

- Metzgar, R.C. (2026). Attributions of mind to humans and AI. Cognitive Science Lunchtime Talk Series, Princeton University, March 2026.
- Metzgar, R.C. (2025). Thinking machines or thinking minds? Neural and behavioral responses to beliefs about conversational partners. Cognitive Science Lunchtime Talk Series, Princeton University, March 2025.
- Metzgar, R.C. (2025). Mind or matter? Exploring lay beliefs about consciousness and its physical basis. Cognitive Research Seminar, Princeton University, February 2025.
- Metzgar, R.C. (2023). A link between social cognition and visual consciousness. Cognitive Research Seminar, Princeton University, November 2023.
- Metzgar, R.C. (2023). A link between social cognition and visual consciousness. Cognitive Research Seminar, Princeton University, March 2023.
- Metzgar, R.C. (2022). Attributing awareness to the self and others. Cognitive Research Seminar, Princeton University, April 2022.

Poster Presentations

- Metzgar, R.C. (2026). Representation of Mind in LLMs. Twenty Ninth Annual Meeting of the Association for the Scientific Studies of Consciousness, Santiago, Chile.
- Metzgar, R.C., Christian, I.R., & Graziano, M.S.A. (2025). Thinking machines or thinking minds? Neural and behavioral responses to beliefs about conversational partners. Twenty Eighth Annual Meeting of the Association for the Scientific Studies of Consciousness, Heraklion, Greece.
- Metzgar, R.C., Lombrozo, T., & Graziano, M.S.A. (2024). Public opinion on whether consciousness is physical or nonphysical. Twenty Seventh Annual Meeting of the Association for the Scientific Studies of Consciousness, Tokyo, Japan.
- Metzgar, R.C. & Graziano, M.S.A. (2023). A link between social cognition and visual consciousness. Twenty Sixth Annual Meeting of the Association for the Scientific Studies of Consciousness, New York, NY.
- Metzgar, R.C. & Graziano, M.S.A. (2022). Attributing awareness to the self and others. Twenty Fifth Annual Meeting of the Association for the Scientific Studies of Consciousness, Amsterdam, NL.
- Metzgar, R.C., Bonner, M.F., & Epstein, R.A. (2019). What lies beyond: Representations of the connectivity structure of the local environment. Nineteenth Annual Meeting of the Vision Sciences Society, St. Pete Beach, FL.
- Lompadó, A., Metzgar, R., Stibolt, O., & Hoffman, J.E. (2017). Can targets be semantically primed during the emotional blink? 25th Annual Meeting of the Object Perception, Attention, and Memory Workshop, Vancouver, BC.
- Metzgar, R.C. & Hoffman, J.E. (2017). How does emotion interfere with attention? University of Delaware Undergraduate Research and Service Celebratory Symposium, Newark, DE.
- Metzgar, R.C. & Hoffman, J.E. (2016). How does emotion interfere with attention? University of Delaware Undergraduate Research and Service Celebratory Symposium, Newark, DE.
- Metzgar, R.C., Osborne, B.F., & Schwarz, J.M. (2015). Impact of sex and neonatal infection on neuroimmune and peripheral immune responses in the juvenile rat. University of Delaware Undergraduate Research and Service Celebratory Symposium, Newark, DE.

RESEARCH EXPERIENCE

2021–Present: Graduate Researcher, Princeton University

- Led self-directed multi-experiment investigation into how attributions of mind shape conversational behavior, subjective judgments, and neural processing, using fMRI, computational linguistics, and mechanistic interpretability methods with large language models.
- Designed and conducted psychophysics experiments testing links between social cognition and visual awareness, and large-scale survey studies mapping lay beliefs about whether consciousness is physical or nonphysical.
- Mentored undergraduate students and coordinated data collection across behavioral, survey, and neuroimaging paradigms.

2019–2021: Lab Manager, Moss Rehabilitation Research Institute

- Coordinated a team of multidisciplinary researchers to study the impacts of traumatic brain injury on cognition.

2017–2019: Research Specialist, University of Pennsylvania

- Designed and executed fMRI and VR research on the neural link between visual scene perception and spatial memory.

2016–2017: Research Assistant, University of Delaware

- Conducted independent honors thesis on effects of emotion on visual attention using EEG, behavioral, and eye-tracking methods.

FELLOWSHIPS IN RESPONSIBLE AI & INDUSTRY

6–10 hours / week per experience

June 2026–Present: **AI Sentience Scholars (AISS) Program**, Neuromatch

- Selected for program supporting early-career researchers investigating AI, consciousness, and ethics through empirically grounded inquiry.

Fall 2025–Spring 2026: **Communications Fellow**, Princeton Laboratory for Artificial Intelligence

- Authored blog series synthesizing philosophical, cognitive science, and technical perspectives on machine understanding for general audiences.

Spring–Summer 2025: **Social Impact Fellow**, New Jersey AI Hub

- Co-authored Wiley book chapter on regional AI workforce development; designed student engagement strategy to connect students with industry and government partners.

Fall 2025: **University Administrative Fellow**, Responsible AI Learning Cohort, Princeton GradFUTURES

- Designed a cross-disciplinary series bringing graduate students together to discuss ethical AI practice.

Summer 2024: **Social Impact Fellow**, Responsible AI Institute

- Developed accessible guide on AI explainability/interpretability for non-technical stakeholders.

Summer 2023: **Social Impact Fellow**, All Tech Is Human

- Curated research themes for journal collection on social impacts of AI on youth and children.

OTHER LEADERSHIP & SERVICE

2025–2026: **Head Cognitive Science Fellow**, Princeton University

- Coordinated programming for the cognitive science graduate and undergraduate community, including professional development, research events, and community building.

2024–2025: **Professional Development Committee**, Princeton University Graduate School

- Helped plan events on career development, job market preparation, and skills beyond traditional academic training for Princeton graduate students.

2024–2025: **Student Committee Member**, Association for the Scientific Study of Consciousness (ASSC)

- Co-organized the Professional Development Q&A event at the 2024 and 2025 annual meetings, connecting early-career researchers with established scientists to discuss career paths.

TEACHING EXPERIENCE

Fall 2024: **Assistant in Instruction**, Introduction to Psychology, Princeton University

Spring 2024: **Assistant in Instruction**, Psychotherapy Theories and Skills, Princeton University

Summer 2023: **Teaching Assistant**, Deep Learning, Neuromatch Academy

Fall 2022: **Assistant in Instruction**, Developmental Psychology, Princeton University

HONORS & AWARDS

2025: Cognitive Science Fellow (Head Fellow), Princeton University

2025: Cognitive Science Graduate Student Research Funding, Princeton

2024: Cognitive Science Fellow, Princeton University

2024: Cognitive Science Graduate Student Research Funding, Princeton

2023: Cognitive Science Graduate Student Research Funding, Princeton

2017: Honors Degree with Distinction, University of Delaware

SELECTED COURSEWORK

Princeton University

- Machine Learning: A Practical Introduction for Humanists and Social Scientists (2022)
- AI Safety & Alignment (2023)

Neuromatch Academy

- NeuroAI (2024)

SKILLS

LLM Evaluation & Research Methods: evaluation of LLMs, controlled prompt manipulation studies, cross-model evaluation, dyadic LLM-to-LLM conversation simulation, refusal-rate analysis, naturalistic conversation paradigms, LLM-as-judge frameworks, experimental design, survey design, quality assurance frameworks, cross-functional collaboration, human subjects research, hypothesis testing, text/language analysis, neuroimaging, psychophysics, AI tools to automate workflows.

Statistical & Software: Python, PyTorch, R, MATLAB, Claude Code, GitHub, HPC/Slurm, Qualtrics, Prolific, OpenAI & HuggingFace APIs, open-source LLMs, mechanistic interpretability (linear probing, activation steering, concept vector analysis, representational similarity analysis), prompt engineering, NLP/text analysis, machine learning (ML), deep learning, statistical modeling (regression, mixed-effects, factor analysis, ANOVA, predictive modeling), generative AI, rapid research prototyping, reproducible analysis pipelines, data visualization.

Communication & Leadership: project management, public speaking, technical writing, thought leadership content development, translating complex research into accessible frameworks, cross-sector and cross-functional collaboration, stakeholder and policy engagement.

Research topics: mechanistic interpretability, explainability, RAI, AI safety, AI alignment, theory of mind, attribution of mind, human-AI interaction, digital minds, attention, consciousness, cognitive neuroscience, social cognition, spatial cognition, emotion, scene perception, action production, neuropsychology, navigation