Sarah A. Wu she/her

🌐 https://sarahawu.github.io/   ✉ sarahawu@stanford.edu   ⬡ sarahawu

Education

Stanford University	2020 – 2026
Ph.D. in Psychology	
Advisor: Tobias Gerstenberg	
Diverse Intelligences Summer Institute	2021
Massachusetts Institute of Technology	2016 – 2020
B.S. in Mathematics with Computer Science; Brain & Cognitive Sciences	

Experience

Research Scientist Intern, Human-Centered AI Research, Apple	Summer 2024
Manager: James Rae	
PhD Research Intern, Mosaic, Allen Institute for Artificial Intelligence (AI2)	Summer 2023
Mentors: Sydney Levine, Xiang Ren	

Honors & Awards

Hsieh Family Fellow, Stanford Interdisciplinary Graduate Fellowship (SIGF)	2023 – 2026
Norman H. Anderson Research Fund, Stanford Psychology	2023
Stanford Institute for Research in the Social Sciences Grant	2022
Stanford Institute for Research in the Social Sciences Grant	2021
Best Paper Award, NeurIPS Cooperative AI Workshop	2020
Computational Modeling Prize in Higher Cognition, Cognitive Science Society	2020
NSF Graduate Research Fellowship (GRFP)	2020 – 2023
Phi Beta Kappa	2020

Publications

Journal Articles

Sarah A. Wu and Tobias Gerstenberg (2023). If not me, then who? Responsibility and replacement. *Cognition*, 242, 105646.

Sarah A. Wu*, Rose E. Wang*, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner (2021). Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science*, 13(2), 414-432.

Sarah A. Wu and Edward Gibson (2021). Word order predicts cross-linguistic differences in the production of redundant color and number modifiers. *Cognitive Science*, 45(1), e12934.

Book Chapters

Rose E. Wang*, **Sarah A. Wu***, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner (2021). Too many cooks: Bayesian inference for coordinating multi-agent collaboration. In S. Muggleton and N. Charter (Ed.), *Human-like Machine Intelligence*, pp. 152-170. Oxford University Press.

Peer-reviewed Conference Proceedings

Emily Jin, Zhuoyi Huang, Jan-Philipp Fränken, Weiyu Liu, Hannah Cha, Erik Brockbank, **Sarah A. Wu**, Ruohan Zhang, Jiajun Wu, and Tobias Gerstenberg (2024). MARPLE: A Benchmark for Long-Horizon Inference. *Advances in Neural Information Processing Systems (NeurIPS)*.

Sarah A. Wu, Xiang Ren, Tobias Gerstenberg, Yejin Choi, and Sydney Levine (2024). Resource-rational moral judgment. *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

Sarah A. Wu*, Erik Brockbank*, Hannah Cha, Jan-Philipp Fränken, Emily Jin, Zhuoyi Huang, Weiyu Liu, Ruohan Zhang, Jiajun Wu, and Tobias Gerstenberg (2024). Whodunnit? Inferring what happened from multimodal evidence. *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Sarah A. Wu, Shruti Sridhar, and Tobias Gerstenberg (2023). A computational model of responsibility from counterfactual simulations and intention inferences. *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

Sarah A. Wu, Shruti Sridhar, and Tobias Gerstenberg (2022). That was close! A counterfactual simulation model of causal judgments about decisions. *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Sarah A. Wu*, Rose E. Wang*, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner (2020). Too many cooks: Coordinating multi-agent collaboration through inverse planning. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

Conference Presentations

"Whodunnit? Inferring what happened from multimodal evidence"
  Poster at CogSci 2024

"Resource-rational moral judgment"
  Talk at SPP 2024
  Poster at CogSci 2024
  Poster at NeurIPS 2023 AI Meets Moral Philosophy and Moral Psychology (MP2) Workshop

"A computational model of responsibility from counterfactual simulations and intention inferences"
  Poster at CogSci 2023

"That was close! A counterfactual simulation model of causal judgments about social agents"
  Talk at (E)SPP 2022
  Poster at CogSci 2022
  Poster at RSS 2022 Social Intelligence in Humans and Robots Workshop

"The role of counterfactual reasoning in responsibility judgments"
  Talk at SPP 2021
  Poster at CogSci 2021

"Too many cooks: Bayesian inference for coordinating multi-agent collaboration"
  Spotlight talk & poster at NeurIPS 2020 Cooperative AI Workshop

"Word order predicts cross-linguistic differences in the production of redundant color and number modifiers"
  Talk at AMLaP 2020

"Too many cooks: Coordinating multi-agent collaboration through inverse planning"
  Talk at CogSci 2020
  Talk at AAMAS 2020

Invited Presentations

Workshop on Genericity, Stability, and Structural Interactions, Cal State East Bay	2024
CompCogSci Group, University of Edinburgh	2023
Cognitive Tools Lab, UCSD	2023
Computational Vision and Learning Lab, UCLA	2022
Experimental Jurisprudence Workshop, Michigan Law School	2022
Visual Intelligence Lab, UCLA	2022
Social and Cognitive Computational Neuroscience Lab, Boston College	2020

Teaching

Teaching Assistant

Stanford PSYCH 205 Foundations of Cognition	Spring 2023
Stanford PSYCH 251 Experimental Methods	Fall 2021, Fall 2022, Fall 2023
Stanford PSYCH 252 Statistical Methods for Social & Behavioral Sciences	Winter 2022
Stanford SYMSYS 1 Minds and Machines	Summer 2021
MIT 6.046 Design and Analysis of Algorithms	Spring 2019, Fall 2019, Spring 2020
MIT 6.036 Introduction to Machine Learning	Fall 2018

Instructor

IIS Curie-Sraffa High School, Milan, Italy (through MIT Global Teaching Labs)	2019

Mentoring

Chuqi Hu (Master's student, 2024 –)
Verona Teo (predoctoral researcher, 2024 –)
Shruti Sridhar (undergraduate, 2021 – 2024)
Siying Zhang (research assistant, 2022 – 2023)
Gabe Gaw (undergraduate, 2022 – 2023)

Professional Service & Activities

Reviewing

Journal of Experimental Psychology: Learning, Memory, and Cognition
Cognitive Science Society Conference (CogSci)
NeurIPS Workshops
ICML Workshops
AAAI Workshops

Conference Organizing

"Counterfactuals in Minds and Machines" Workshop, ICML 2023

Department & University Service

Stanford Psychology Admissions Reader	2024 –
Stanford Psychology Advising Coach	2023 –
Stanford Psychology Faculty Meeting Representative	2020 – 2022
MIT Educational Counselor	2020 – 2025

Outreach

Stanford Future Advancers of Science & Technology (`https://fast.stanford.edu/`)	2020 –
President	2023 – 2024
Chief Program Officer	2022 – 2023
Director of Mentor Recruitment	2021 – 2022