Sarah A. Wu she/her

https://sarahawu.github.io/ sarahawu@stanford.edu sarahawu

Education

Stanford University 2020 – 2026
Ph.D. in Psychology
Committee: Tobias Gerstenberg (advisor), Noah Goodman, Judith Fan

Massachusetts Institute of Technology 2016 – 2020
B.S. in Mathematics with Computer Science; Brain & Cognitive Sciences

Industry Experience

Apple, Human-Centered AI Research Team 2024
Research Scientist Intern
Supervisor: James Rae

Allen Institute for Artificial Intelligence, Mosaic Team 2023
PhD Research Intern
Supervisors: Sydney Levine, Xiang Ren

Awards

Best Paper Award, NeurIPS Cooperative AI Workshop 2020
Computational Modeling Prize in Higher Cognition, Cognitive Science Society 2020
Phi Beta Kappa, MIT 2020
Hans Lukas Teuber Award for Outstanding Academics, MIT 2019, 2020
U.S. National Physics Team 2016

Fellowships & Grants

Stanford Interdisciplinary Graduate Fellowship 2023 – 2026
Norman H. Anderson Research Grant, Stanford University 2023, 2025
Institute for Research in the Social Sciences Grant, Stanford University 2021, 2022
Diverse Intelligences Summer Institute Fellow 2021
NSF Graduate Research Fellowship 2020 – 2023
Amgen National Scholar 2018
Singapore-MIT Undergraduate Research Fellow 2017

Publications

Preprints

Neha Balamurugan, **Sarah A. Wu**, Adam Chun, Gabe Gaw, Cristobal Eyzaguirre, and Tobias Gerstenberg (submitted). Spot The Ball: A Benchmark for Visual Social Inference. *arXiv:2511.00261 [cs.CV]*.

Archival Publications

Carlota Parés-Morlans, Michelle Yi, Claire Chen, **Sarah A. Wu**, Rika Antonova, Tobias Gerstenberg, and Jeannette Bohg (2025). Causal-PIK: Causality-based physical reasoning with a physics-informed kernel. *International Conference on Machine Learning (ICML)*.

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 and Tobias Gerstenberg (2024). 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.

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.

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.

Other Publications

Verona Teo, **Sarah A. Wu**, Erik Brockbank, and Tobias Gerstenberg (2025). Leave a trace: Recursive reasoning about deceptive behavior. *Proceedings of the 47th Annual Conference of the Cognitive Science Society.*

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.*

Invited Presentations

Workshop on Genericity, Stability, and Structural Interactions, Cal State East Bay	2024
Experimental Jurisprudence Workshop, University of Michigan Law School	2022

Conference Presentations

CogSci 2025 Poster on "Leave a trace: Recursive reasoning about deceptive behavior"
Verona Teo (presenting), Sarah A. Wu, Erik Brockbank, and Tobias Gerstenberg

CogSci 2025 Poster on "Spot the ball: Inferring hidden information from human behavioral cues"
Neha Balamurugan (presenting), Sarah A. Wu, Cristobal Eyzaguirre, Adam Chun, Gabe Gaw, and Tobias Gerstenberg

ICML 2025 Poster on "Causal-PIK: Causality-based physical reasoning with a physics-informed kernel"
Carlota Parés-Morlans (presenting), Michelle Yi, Claire Chen, Sarah A. Wu, Rika Antonova, Tobias Gerstenberg, and Jeannette Bohg

NeurIPS 2024 Poster on "MARPLE: A Benchmark for Long-Horizon Inference"
Emily Jin (presenting), Zhuoyi Huang, Jan-Philipp Fränken, Weiyu Liu, Hannah Cha, Erik Brockbank, Sarah A. Wu, Ruohan Zhang, Jiajun Wu, and Tobias Gerstenberg

CogSci 2024 Poster on "Whodunnit? Inferring what happened from multimodal evidence"
Sarah A. Wu* (presenting), Erik Brockbank* (presenting), Hannah Cha, Jan-Philipp Fränken, Emily Jin, Zhuoyi Huang, Weiyu Liu, Ruohan Zhang, Jiajun Wu, and Tobias Gerstenberg

CogSci 2024 Poster on "Resource-rational moral judgment"
Sarah A. Wu (presenting), Xiang Ren, Tobias Gerstenberg, Yejin Choi, and Sydney Levine

SPP 2024 Talk on "Resource-rational moral judgment"
Sarah A. Wu (presenting), Xiang Ren, Tobias Gerstenberg, Yejin Choi, and Sydney Levine

NeurIPS 2023 Talk on "Resource-rational moral judgment"
AI Meets Moral Philosophy and Moral Psychology Workshop
Sarah A. Wu (presenting), Xiang Ren, Tobias Gerstenberg, Yejin Choi, and Sydney Levine

CogSci 2023 Poster on "A computational model of responsibility from counterfactual simulations and intention inferences"
Sarah A. Wu (presenting), Shruti Sridhar, and Tobias Gerstenberg

CogSci 2022 Poster on "That was close! A counterfactual simulation model of causal judgments about social agents"
Sarah A. Wu (presenting), Shruti Sridhar, and Tobias Gerstenberg

(E)SPP 2022 Talk on "That was close! A counterfactual simulation model of causal judgments about social agents"
Sarah A. Wu (presenting), Shruti Sridhar, and Tobias Gerstenberg

CogSci 2021 Poster on "The role of counterfactual reasoning in responsibility judgments"
Sarah A. Wu (presenting) and Tobias Gerstenberg

SPP 2021 Talk on 'The role of counterfactual reasoning in responsibility judgments"
Sarah A. Wu (presenting) and Tobias Gerstenberg

NeurIPS 2020 Spotlight talk on "Too many cooks: Bayesian inference for coordinating multi-agent collaboration"
Cooperative AI Workshop
Rose E. Wang* (presenting), Sarah A. Wu*, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner

AMLaP 2020 Talk on "Word order predicts cross-linguistic differences in the production of redundant color and number modifiers"
Sarah A. Wu (presenting) and Edward Gibson

CogSci 2020 Talk on "Too many cooks: Coordinating multi-agent collaboration through inverse planning"
Sarah A. Wu* (presenting), Rose E. Wang*, James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner

AAMAS 2020 Talk on "Too many cooks: Coordinating multi-agent collaboration through inverse planning"
Sarah A. Wu*, Rose E. Wang* (presenting), James A. Evans, Joshua B. Tenenbaum, David C. Parkes, and Max Kleiman-Weiner

Teaching

TA for Stanford PSYCH 205 Foundations of Cognition	Spring 2023
TA for Stanford PSYCH 251 Experimental Methods	Fall 2021, Fall 2022, Fall 2023
TA for Stanford PSYCH 252 Statistical Methods for Social & Behavioral Sciences	Winter 2022
TA for Stanford SYMSYS 1 Minds and Machines	Summer 2021
TA for MIT 6.046 Design and Analysis of Algorithms	Spring 2019, Fall 2019, Spring 2020
TA for MIT 6.036 Introduction to Machine Learning	Fall 2018
TA for MIT 12.000 Solving Complex Problems (Terrascope)	Fall 2017
Instructor at IIS Curie-Sraffa High School, Milan, Italy	Winter 2019

Mentoring

Min Jung (Stanford undergraduate student)	2025 –
Chuqi Hu (Stanford master's student)	2025 –
Verona Teo (Stanford predoctoral researcher)	2024 –
Siying Zhang (Stanford predoctoral researcher)	2022 – 2023
Gabe Gaw (Stanford undergraduate student)	2022 – 2023
Shruti Sridhar (Stanford undergraduate & master's student)	2021 – 2025

Service & Leadership

Journal Reviewing

Open Mind	2025
Journal of Experimental Psychology: Learning, Memory, and Cognition	2024

Conference Reviewing

Cognitive Science Society Conference (CogSci)	2021 – 2025
Association for the Advancement of Artificial Intelligence (AAAI) Workshops	2024
International Conference on Machine Learning (ICML) Workshops	2023 – 2024
Neural Information Processing Systems (NeurIPS) Workshops	2022 – 2023

Organizing

ICML "Counterfactuals in Minds and Machines" Workshop	2023

University Service & Outreach

Stanford Psychology Admissions Reader	2024 –
Stanford Psychology Advising Coach	2023 –
Stanford Psychology Paths to PhD Mentor	2020 – 2023
Stanford Psychology Faculty Meeting Representative	2020 – 2022
MIT Educational Counselor	2020 – 2025
Future Advancers of Science & Technology (`https://fast.stanford.edu/`)	
Outcomes Officer	2024 –
President	2023 – 2024
Chief Program Officer	2022 – 2023
Director of Mentor Recruitment	2021 – 2022