

KEREM OKTAR

<https://keremoktar.com> | oktar.research@gmail.com

Employment

Meta Fundamental AI Research (FAIR) 2025 - Present
Postdoctoral Researcher
Projects: LLM Evaluation, Alignment, & Post-training

Education

Princeton University 2019 - 2025
Ph.D. in Psychology
Advisors: Tania Lombrozo & Thomas L. Griffiths
Thesis: *Cognitive Mechanisms of Social Controversies*

Pomona College 2015 - 2019
B.A. in Economics, Cognitive Science (*cum laude*)

Publications

SELECTED PEER-REVIEWED RESEARCH

* denotes joint authorship

1. Wu, A.*, Liu, R.*, **Oktar, K.***, Summers, T., & Griffiths, T. (2025). Are Large Language Models Sensitive to the Motives Behind Communication? *Advances in Neural Information Processing Systems* [NeurIPS], 38. [\[Link\]](#)
2. **Oktar, K.** & Lombrozo, T. (2025). How Beliefs Persist Amid Controversy: The Paths to Persistence Model. *Psychological Review*. [\[Link\]](#)

PEER-REVIEWED JOURNAL ARTICLES

... elides many co-authors

1. **Oktar, K.***, Handley-Miner, I.*, Kleiman-Weiner, M., Zollman, K., & Young, L. (in press). Intuitive Theories of Truth. *Trends in Cognitive Sciences*. [\[Link\]](#)
2. **Oktar, K.** & Lombrozo, T. (2025). How Aggregated Opinions Shape Beliefs. *Nature Reviews Psychology*. [\[Link\]](#)
3. Sucholutsky, I., ..., **Oktar, K.**, ..., Griffiths, T. L. (2025). Getting aligned on representational alignment. *Transactions on Machine Learning Research* [\[Link\]](#)
4. **Oktar, K.**, Lombrozo, T.*, & Griffiths, T.* (2024). Learning from Aggregated Opinion. *Psychological Science*. [\[Link\]](#)

5. **Oktar, K.**, Sucholutsky, I., Lombrozo, T., & Griffiths, T. (2024). Dimensions of Disagreement: Unpacking Divergence and Misalignment in Cognitive Science and Artificial Intelligence. *Decision*. [\[Link\]](#)
6. **Oktar, K.**, Lerner, A., Malaviya, M., & Lombrozo, T. (2023). Philosophy instruction changes views on moral controversies by decreasing reliance on intuition. *Cognition*, 236, 105434. [\[Link\]](#)
7. **Oktar, K.** & Lombrozo, T. (2022). Deciding to be authentic: Intuition is favored over deliberation when authenticity matters. *Cognition*, 223, 105021. [\[Link\]](#)
8. **Oktar, K.**, Laura, Z., & Leticia, S. (2019). Impact of Increasing Levels of Difficulty on the Generation Effect. *Koc University Undergraduate Psychology Journal*, 4(18), 12-17. [\[Link\]](#)

PEER-REVIWED CONFERENCE PROCEEDINGS

Note: In computer science, conference proceedings are the primary venue for publications.

1. Foster-Hanson, E., **Oktar, K.**, Lombrozo, T., & Kelts, S. (2025). Psychological Measurement of Technology Ethics Education using the REGAIN Empirical Framework. Proceedings of the 2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology (ETHICS), *IEEE ETHICS*. [\[Link\]](#)
2. **Oktar, K.***, Byers, B.*, & Lombrozo, T. (2024). Are Disagreements Just Differences in Beliefs? Proceedings of the 46th Annual Conference of the Cognitive Science Society, *Cognitive Science Society*. [\[Link\]](#)
3. **Oktar, K.**, Summers, T., & Griffiths, T. (2024). A Rational Model of Vigilance in Communication. Proceedings of the 46th Annual Conference of the Cognitive Science Society, *Cognitive Science Society*. [\[Link\]](#)
4. **Oktar, K.** and Lombrozo, T., (2022). Mechanisms of Belief Persistence in the Face of Societal Disagreement. Proceedings of the 44th Annual Conference of the Cognitive Science Society, *Cognitive Science Society*. [\[Link\]](#)
5. Malaviya, M., Sucholutsky, I., **Oktar, K.**, and Griffiths, T. (2022). Can Humans Do Less-Than-One-Shot Learning? Proceedings of the 44th Annual Conference of the Cognitive Science Society, *Cognitive Science Society*. [\[Link\]](#)
6. **Oktar, K.** and Lombrozo, T., (2021). Deciding to be Authentic: Intuition is Favored Over Deliberation for Self-Reflective Decisions. Proceedings of the 43rd Annual Conference of the Cognitive Science Society, *Cognitive Science Society*. [\[Link\]](#)

7. **Oktar, K.** and Lombrozo, T., (2020). You Should Really Think This Through: Cross-Domain Variation in Preferences for Intuition and Deliberation. Proceedings of the 42nd Annual Conference of the Cognitive Science Society, *Cognitive Science Society*. [\[Link\]](#)

COMMENTARIES

1. Sommer, J., & **Oktar, K.** (2025). Maps by Which We May Not Steer: Why Psychologists Should Expect Low Belief-Behavior Correspondence. *Psychological Inquiry*, 36(1), 60-66. [\[Link\]](#)
2. **Oktar, K.** (2024). Psychology Remains Marginally Valid. *Nature Reviews Psychology*. [\[Link\]](#)

Manuscripts in Progress

MANUSCRIPTS UNDER REVIEW

1. **Oktar, K.**, Sumers, T., & Griffiths, T. Rational Vigilance of Intentions and Incentives Guides Learning from Advice. Under Review at *Journal of Experimental Psychology: General*. Preprint: https://doi.org/10.31234/osf.io/khtpy_v1
2. **Oktar, K.***, Byers, B.*, & Lombrozo, T. Disagreement is Not (Just) a Difference in Beliefs. Under Review at *Cognition*.
3. **Oktar, K.***, Gottlieb-Cohen, S.*, & Lombrozo, T. Deciding when we cannot know: First-person priority in transformative choice. Under Review at *Open-mind*.

PRE-PRINTS

1. **Oktar, K.**, Collins, K. M., Hernandez-Orallo, J., Coyle, D., Cave, S., Weller, A., & Sucholutsky, I. (2025). Identifying, Evaluating, and Mitigating Risks of AI Thought Partnerships. *arXiv preprint* [arXiv:2505.16899](https://arxiv.org/abs/2505.16899)
2. Schoenegger, P., ..., **Oktar, K.**, ... & Karger, E. (2025). Large Language Models Are More Persuasive Than Incentivized Human Persuaders. *arXiv Preprint* [arXiv:2505.09662](https://arxiv.org/abs/2505.09662)

Awards

Ethics Education Research Award

National Ethics Project, 2025

Best Poster Prize

Society for Philosophy and Psychology, 2023

Marr Prize for Best Student Paper	Cognitive Science Society, 2022
Graduate School Teaching Award	Princeton University, 2022
Diversity and Inclusion Conference Award	Cognitive Science Society, 2021
Fletcher Jones Foundation Top Thesis Prize	Pomona College, 2019
Distinction in Linguistics and Cognitive Science	Pomona College, 2019
Pomona College Scholar	Pomona College, 2015 - 2019

Grants & Fellowships

Rockefeller Graduate Prize Fellowship	Princeton Center for Human Values, 2023
Title: <i>A Study of Individual Belief Amid Controversy</i>	
Award: \$110,000 (PI) [research funding, stipend, tuition]	
Cognitive Science Research Grant	Princeton University, 2022
Title: <i>Computational Modeling of Belief Persistence</i>	
Award: \$3,000 (Co-PI) [research funding]	
Partisan Animosity Survey Time Grant	Polarization Research Lab, 2022
Title: <i>Political Differences in Belief Persistence</i>	
Award: Nationally representative sample (Co-PI) [2,000 participants]	
Summer Research Fellowship	Pomona College, 2019
Title: <i>Affective Reactions to Inconsistent Information</i>	
Award: \$4,000 (PI) [research funding, travel]	

TRAVEL FELLOWSHIPS

Conference/Symposium	Society for Judgment and Decision-Making, 2024
	Society for Philosophy and Psychology, 2024
	NYU Stern School of Business, 2024
	Institute for Humane Studies, 2024
	Cognitive Science Society, 2022
Summer Institute	Diverse Intelligences Summer Institute, 2023

Presentations

INVITED TALKS

Learning from Opinions in Minds and Machines

- UW Paul Allen School of Computer Science (PI: Yulia Tsvetkov) 2025
- UW Foster School of Business (PI: Kleiman-Weiner) 2025

- Meta Fundamental AI Research (PI: Celikyilmaz) 2025

How Beliefs Persist When Millions Disagree

- Max Planck Institute (PI: Hertwig) 2024
- Carnegie Mellon (ICSD Talk Series) 2024
- Stanford (PI: Zaki) 2024
- UC Berkeley (PI: Ayduk & Mendoza-Denton) 2024
- UC San Diego (PI: Rossano) 2024
- Harvard (PI: Cushman & Greene) 2023
- Harvard Kennedy School (PI: Minson) 2022
- UCL (Judgment and Decision Making Seminar) 2022

Learning from Aggregated Opinion

- UC San Diego (PI: Lin) 2025
- NYU (PI: Ho & Van Bavel) 2024
- MIT Sloan (PI: Rand & Bhui) 2024
- NYU (ConCats Talk Series) 2024
- Princeton (PI: Vélez) 2023
- Princeton (PI: Crockett) 2023

Can Morality be Taught?

- Bogazici University (Psychology Department Colloquium) 2023
- Princeton (PI: Crockett) 2022

Deciding to be Authentic

- Pomona College (Cognitive Science Department Colloquium) 2020

CONFERENCE TALKS

Fall 2025. Are Large Language Models Sensitive to the Motives Behind Communication? Talk at the Conference on Language Modeling in Montreal, CA.

Fall 2024. Cognitive Mechanisms of Social Controversies. Talk at the Annual Meeting of the Society for Judgment and Decision Making in New York, NY.

Summer 2024. A Rational Model of Vigilance in Motivated Communication. Talk at the 46th Annual Meeting of the Cognitive Science Society in Rotterdam, NE.

Spring 2023. Ideological Differences in Belief Persistence. Talk at the Second Annual Meeting on Partisan Animosity in America, New Orleans, LA.

Fall 2023. Learning from Aggregated Opinion. Talk at the Annual Meeting of the Society for Judgment and Decision Making, San Francisco, CA.

Summer 2023. Learning from Aggregated Opinion. Virtual Symposium talk: 45th Annual Conference of the Cognitive Science Society, Sydney, Australia.

Summer 2022. Mechanisms of Belief Persistence in the Face of Societal Disagreement. Conference talk: 44th Annual Conference of the Cognitive Science Society, Toronto, Canada.

Summer 2021. Deciding to be Authentic: Intuition is Favored Over Deliberation for Self-Reflective Decisions. Virtual conference talk: 43rd Annual Conference of the Cognitive Science Society, Vienna, Austria.

CONFERENCE POSTER PRESENTATIONS

Oktar, K.*, Park, S., & Lombrozo, T. (2024, June). Why do Philosophers Disagree about Disagreement? Presentation at the Society for Philosophy and Psychology.

Oktar, K.*, & Lombrozo, T. (2024, February). What Enables Societal Controversy? Presentation at the Society for Personality and Social Psychology.

Oktar, K.*, Byers, J. B. & Lombrozo, T. (2024, February). Is Disagreement Just a Difference in Beliefs?. Presentation at the Society for Personality and Social Psychology.

Oktar, K.*, Sumers, T.R., & Griffiths, T. L. (2023, November). Trust and Doubt: A Rational Model of Epistemic Vigilance Explains How Inferred Motivations Moderate Inferences. Presentation at the MIT Conference on Directions of Polarization, Social Norms, and Trust in Societies: Perspectives from behavioral sciences.

Oktar, K., Lerner, A., Malaviya, M., & Lombrozo, T. (2023, June). Philosophy instruction changes views on moral controversies by decreasing reliance on intuition. Presentation at the 49th Annual Meeting of the Society for Philosophy and Psychology.

Oktar, K. and Lombrozo, T. (2021, June). Intuition is Favored Over Deliberation When Authenticity Matters. Virtual presentation at the 47th Annual Meeting of the Society for Philosophy and Psychology.

Oktar, K. and Lombrozo, T. (2021, June). Intuition is Favored Over Deliberation When Authenticity Matters. Virtual presentation at the 47th Annual Meeting of the Society for Philosophy and Psychology.

Oktar, K. and Lombrozo, T. (2020, July). You Should Really Think This Through:

[APALA](#): Effective actions to reduce our carbon footprint. [French] 2023
[Big Think](#): Why it isn't strategic to always be rational. 2022
[Psychology Today](#): When to Go With Your Gut and Be True to Yourself. 2022

Service

LEADERSHIP

Mentor Project SHORT, 2022 - Present
Ambassador Center for Open Science, 2022
First Year President Associated Students of Pomona College, 2015
President Princeton Graduate Turkish Students Association, 2021 - 2024
Prospective Student Interview Committee Princeton University, 2022
Student Pool Administrator & IRB Reviewer Princeton University, 2019 - 2021
Student Liaison Pomona Cognitive Science Department, 2018
Faculty Search Panel Member Pomona Cognitive Science Department, 2018

AD HOC REVIEWER

Tally marks (e.g., III) indicate number of reviews completed.

Journals

Psychological Science
Nature Machine Intelligence
Cognition (III)
Journal of Experimental Psychology: General (II)
Current Directions in Psychological Science
Psychonomic Bulletin & Review
Personality and Social Psychology Bulletin
Artificial Intelligence Review
Open-Mind
Decision
Scientific Reports
Psychological Reports
Humanities and Social Sciences Communications
International Journal of Psychology

Conferences

The Annual Meeting of the Cognitive Science Society (IIII)
NeurIPS (Conference on Neural Information Processing Systems) Workshop on Human and Machine Decision-making
ICML (International Conference on Machine Learning) World Models Workshop
ICLR (International Conference on Learning Representations) Representational

Alignment Workshop
The ACM Annual Designing Interactive Systems (DIS) Conference

PROFESSIONAL MEMBERSHIPS

Cognitive Science Society
Society for Judgment and Decision-Making
Society for Philosophy and Psychology
The Society for Personality and Social Psychology
The American Psychological Association

MENTORSHIP

Undergraduate Independent Research

Burcu Aydin [Psychology] Bogazici University, 2024 - Present

- Topic: Examining cross-cultural variation in responses to societal dissent, focusing on the Turkish context.

Sungho (Sam) Park [Philosophy] Princeton University, 2023 - 2025

- Topic: Psychological principles of peer disagreements in epistemology. Mentee received a travel grant to present research at SPP 2024.

Joshua Schoenberg [Computer Science] Princeton University, 2022 - 2024

- Topic: Leveraging importance sampling to explain political perception gaps. Now a software engineer at Microsoft.

References

Tania Lombrozo

Arthur W. Marks '19 Professor of Psychology
Princeton University
lombrozo@princeton.edu

Thomas L. Griffiths

Henry R. Luce Professor of Information Technology, Consciousness, and Culture
Princeton University
tomg@princeton.edu

Asli Celikyilmaz

Senior Staff Research Manager
Meta Fundamental AI Research (Meta FAIR)
Affiliate Professor
University of Washington
asli.ca@live.com

Molly Crockett
Professor of Psychology
Princeton University
mj.crockett@princeton.edu