CURRICULUM VITAE THOMAS L. GRIFFITHS

PERSONAL DETAILS

Electronic mail:	tomg@princeton.edu
Physical mail:	320 Peretsman Scully Hall Department of Psychology Princeton University Princeton NJ 08540
Nationality:	Citizen of Australia, the United Kingdom, & the United States of America

PROFESSIONAL POSITIONS

2018 - present	Henry R. Luce Professor of Information Technology, Consciousness, and Culture, Departments of Psychology and Computer Science, Princeton University
2024 - present	Director, Princeton Laboratory for Artificial Intelligence, Princeton University
2023 - 2024	Director, Center for Statistics and Machine Learning, Princeton University
2017 - 2018	Class of 1951 Professor, Miller Institute for Basic Research in Science University of California, Berkeley
2015 - 2017	Professor, Department of Psychology and Cognitive Science Program University of California, Berkeley
2010 - 2018	Director, Institute of Cognitive and Brain Sciences University of California, Berkeley
2010 - 2015	Associate Professor, Department of Psychology and Cognitive Science Program University of California, Berkeley
2006 - 2010	Assistant Professor, Department of Psychology and Cognitive Science Program University of California, Berkeley
2005 - 2006	Assistant Professor, Department of Cognitive and Linguistic Sciences Brown University

EDUCATION

Ph.D. in Psychology, Stanford University, 2005

Dissertation title: *Causes, coincidences, and theories*

Exchange scholar, Brain and Cognitive Sciences Department and Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 2002-2004

M.S. in Statistics, Stanford University, 2002

M.A. in Psychology, Stanford University, 2002

B.A. (Honours) in Psychology, University of Western Australia, 1998

AWARDS AND HONORS

2019 Troland Research Award, National Academy of Sciences.

2017 Fellow of the John Simon Guggenheim Memorial Foundation.

Miller Professorship, University of California, Berkeley.

2013 Early Career Impact Award for the Cognitive Science Society, Federation of Associations in Behavioral and Brain Sciences (FABBS) Foundation.

2012 Outstanding Young Investigator Award, Psychonomic Society.

Distinguished Scientific Award for Early Career Contribution to Psychology, American Psychological Association.

Fellow, Association for Psychological Science.

2011 Janet Taylor Spence Award for Transformative Early Career Contributions, Association for Psychological Science.

2010 Sloan Foundation Research Fellowship (Computer Science).

Young Investigator Program grant, Air Force Office of Scientific Research.

Young Investigator Award, Society of Experimental Psychologists.

2009 Faculty Early Career Development (CAREER) award, National Science Foundation.

William K. Estes Early Career Award, Society for Mathematical Psychology.

2006 "AI Ten to Watch" award from *IEEE Intelligent Systems* magazine, awarded to the ten most promising young scientists performing artificial intelligence research as part of the 50th anniversary of the first artificial intelligence conference.

2002 Stanford University Centennial Teaching Assistant Award.

Department of Psychology Distinguished Teaching Award.

1999 Stanford Graduate Fellowship

1998 Hackett Studentship

J.A. Wood Prize (best student in the Faculties of Arts, Law, and Economics at the University of Western Australia).

Best paper awards

2025 Outstanding Paper Prize from the International Conference on Machine Learning for with "Conformal prediction as Bayesian quadrature" with Jake Snell.

2022 Outstanding Paper Prize from the Neural Information Processing Systems Conference for "Using natural language and program abstractions to instill human inductive biases in machines" with Sreejan Kumar, Carlos Correa, Ishita Dasgupta, Raja Marjieh, Michael Hu, Robert Hawkins, Nathaniel Daw, Jonathan Cohen, and Karthik Narasimhan.

2020 Computational Modeling Prize in Language from the Annual Conference of the Cognitive Science Society for "Generalizing meanings from partners to populations: Hierarchical inference supports convention formation on networks" with Robert Hawkins, Noah Goodman, and Adele Goldberg.

2017 Blue Sky Paper Award from International Symposium on Robotics Research (ISRR) for "Pragmatic-pedagogic value alignment" with J. F. Fisac, M. A. Gates, J. B. Hamrick, C. Liu, D. Hadfield-Menell, M. Palaniappan, D. Malik, S. S. Sastry, and A. Dragan.

2017 Best Paper Prize from the Cognitively Informed Artificial Intelligence Workshop at the Neural Information Processing Systems conference for "Learning to select computations" with Falk Lieder, Fred Callaway, Sayan Gul and Paul Krueger.

2016 Computational Modeling Prize in Perception and Action from the Annual Conference of the Cognitive Science Society for "Adapting deep network features to capture psychological representations" with Josh Peterson and Josh Abbott.

2012 Best Poster award at the Education and Data Mining conference for "Inferring learners knowledge from observed actions," with Anna Rafferty and Michelle Lamar.

2010 Best Article Published in *Psychonomic Bulletin and Review* in 2010, for "Exemplar models as a mechanism for performing Bayesian inference," with Lei Shi, Naomi Feldman, and Adam Sanborn.

Best Application Paper award at the International Conference on Machine Learning for "Modeling transfer learning in human categorization with the hierarchical Dirichlet process," with Kevin Canini and Mikhail Shashkov.

2007 Adam Sanborn received the Outstanding Student Paper prize for “Markov chain Monte Carlo with people” at the Neural Information Processing Systems conference.

2006 Elizabeth Bonawitz received the Marr prize for best student paper for “Modeling cross-domain causal learning in preschoolers as Bayesian inference”at the Cognitive Science Society conference.

2004 Honorable mention for Marr prize for best student paper for “Using physical theories to infer hidden causes” at the Cognitive Science Society conference.

2003 Best student paper prize, natural systems (cognitive science) at the Neural Information Processing Systems conference for “From algorithmic complexity to subjective randomness,” with Joshua Tenenbaum.

Best student paper prize, synthetic systems (machine learning) at the Neural Information Processing Systems conference for “Hierarchical topic models and the nested Chinese restaurant process,” with David Blei, Michael Jordan, and Joshua Tenenbaum.

Distinguished invited lectures

2025 Distinguished Lecture in Data Science, Stanford University.

Distinguished Lecture in Neuroscience, Carnegie Mellon University.

Distinguished Lecture in Data Science, University of Chicago.

2022 The Edinburgh Lectures in Language Evolution, University of Edinburgh.

2021 Crowder Lecture, Yale University.

2019 J. James Woods Lecture Series, Butler University.

2018 Roger N. Shepard Visiting Scholar, University of Arizona.

2016 Mind Lecture, University of Kansas.

2015 Teuber Lecture, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

2012 Distinguished Speakers in Cognitive Science Lecture Series, Michigan State University.

2009 Distinguished Speaker Series, Center for Machine Learning and Intelligent Systems, University of California, Irvine.

GRANTS AND FUNDING

External

2025-2027 “Predicting algorithmic trust at scale,” (with 6 other faculty members, Thomas Griffiths as PI), DARPA ($4,997,132).

2025-2028 “Collaborative Research: HNDS-I: Infrastructure to study human-AI hybrid interactions with behavioral experiments,” National Science Foundation, BCS-2523501 ($400,000).

2024-2027 “Overcoming unexpected failures using neurocognitive multi-abstraction active exploration” (with 6 other faculty members, David Held as PI), Office of Naval Research ($816,976).

2024-2026 “Enhancing human creativity through cognitive psychology and AI: A cross-disciplinary investigation into human-AI collaboration,” Toyota ($299,998).

2023-2026 “Instantiating human inductive biases in machines via metalearning,” Office of Naval Research ($750,000).

2023-2026 “Collaborative Proposal: CompCog: RI: Medium: Understanding human planning through AI-assisted analysis of a massive chess dataset,” (with 2 other faculty members) National Science Foundation ($1,500,000).

2022-2025 “Understanding diverse intelligences via diverse constraints,” (with 3 other faculty members) Templeton World Charity Foundation ($1,500,000).

2022-2027 “HUDDLE: Human Autonomy Teaming in Uncertain and Dynamic Environments,” (with 5 other faculty members, Laurel Riek as PI) Office of Naval Research ($931,130).

2020-2021 “RAPID: The effect of a crisis on intertemporal choice,” (with Jonathan Cohen) National Science Foundation ($125,142).

2019-2022 “Toward a Scientific Understanding of the Human Capacity for Autonomy” John Templeton Foundation (with 4 other faculty members, Jonathan Cohen as PI) ($4,995,106).

2018-2022 “Structured Deep Learning for Modeling and Controlling High-Dimensional Dynamical Systems,” Office of Naval Research (with 2 other faculty members, Ani Majumdar as PI) ($1,999,830).

2018-2021 “CompCog: Helping People Make More Future-minded Decision Using Optimal Gamification,” National Science Foundation, ($499,423).

2018-2021 “Resource Rationality as a Foundation for Augmented Reality Systems,” Facebook Reality Labs ($802,787).

2017-2021 “An Integrated Nonparametric Bayesian and Deep Neural Network Framework for Biologically-Inspired Lifelong Learning” DARPA (with 5 other faculty members, Katherine Heller as PI) ($1,200,000).

2017-2020 “ Discovering optimal strategies for bounded agents,” Air Force Office of Scientific Research, FA9550-18-1-0077 ($694,343).

2017-2020 “RI SMALL: CompCog: Leveraging Deep Neural Networks for Understanding Human Cognition,” National Science Foundation ($448,284).

2016-2018 “Understanding and extending human metacognitive intelligence,” Templeton World Charity Foundation ($199,707).

2016-2021 “Center for human-compatible AI,” Open Philanthropy Foundation (with 6 other faculty members, Stuart Russell as PI) ($5,500,000).

2016-2021 “CPS: Frontier: Collaborative Research: VeHICaL: Verified Human Interfaces, Control, and Learning for Semi-Autonomous Systems,” National Science Foundation (with 7 other faculty members, Sanjit Seshia as PI) ($3,590,000).

2016-2020 “Culture-on-a-chip Computing: Crowdsourced Simulations of Culture, Group Formation, and Collective Identity,” DARPA (with 3 other faculty members, Thomas Griffiths as PI) ($4,786,471).

2016 “Evaluating semantic representations from neural networks against human behavior,” Google Faculty Research Award ($71,340).

2015-2016 “Value alignment and moral metareasoning,” Future of Life Institute ($110,883).

2015-2017 “Testing evolutionary hypotheses through large-scale behavioral simulations,” National Science Foundation, BCS-1456709 ($474,697).

2014-2017 “Diagnosing misconceptions about algebra using Bayesian inverse reinforcement learning,” National Science Foundation, DRL-1420732 ($443,248).

2013-2018 “Data on the mind: Center for data-intensive psychological science,” National Science Foundation, SMA-1228541 (with Alison Gopnik and Dacher Keltner) ($531,482).

2013-2017 “Rational randomness: Search, sampling and exploration in children’s causal learning,” National Science Foundation, BCS-1331620 (with Alison Gopnik) ($446,815).

2013-2017 “Embedded humans: Provably correct decision making for networks of human and unmanned systems,” Office of Naval Research, N00014-13-1-0341 (with 11 other faculty members, Shankar Sastry as PI) ($7,500,000).

2013-2017 “Inductive inference by humans and machines,” Air Force Office of Scientific Research, FA9550-13-1-0170 ($694,343).

2012-2017 “CRCNS: Cortical representation of phonetic, syntactic and semantic information during speech perception and language comprehension”, National Science Foundation, IIS-1208203 (with Jack Gallant and Frederic Theunissen) ($423,718).

2011-2012 “Perceptual grounding of language using probabilistic models”, DARPA, BOLT-E (with five other faculty, Trevor Darrell as PI) ($1,093,768).

2010-2013 “Probabilistic models for reconstructing ancient languages”, National Science Foundation, IIS-1018733 (with Dan Klein) ($460,143).

2010-2013 “Causal learning as sampling”, National Science Foundation, BCS-1023875 (with Alison Gopnik) ($323,030).

2010-2012 Research Fellowship in Computer Science, Sloan Foundation ($50,000).

2010-2013 “Fast, flexible, rational inductive inference”, Air Force Office of Scientific Research, FA-9550-10-1-0232 ($358,028).

2009-2013 “CAREER: Connecting human and machine learning through probabilistic models of cognition”, National Science Foundation, IIS-0845410 ($546,841).

2008-2009 “Workshop: Probabilistic models of cognitive development”, National Science Foundation, DLR-0838595 ($56,982).

2008 “Nonparametric Bayesian models for relational data” (with Michael Jordan, University of California, Berkeley), Lawrence Livermore National Laboratory ($70,000).

2006-2008 “Topic modeling and identification” DARPA/SRI Cognitive Agent that Learns and Organizes (CALO) project ($150,000).

2006-2009 “Collaborative research: Knowledge transmission through iterated learning” (with Michael Kalish, University of Louisiana at Lafayette), National Science Foundation, BCS-0704034 ($314,234 total, with $114,234 to Berkeley).

2006-2009 “Collaborative research: Bayesian methods for learning and analyzing natural language” (with Mark Johnson, Brown University), National Science Foundation, SES-0631518 ($320,000 total, with $160,000 to Berkeley).

2007-2009 “Theory-based Bayesian models of inductive inference”, Air Force Office of Scientific Research, FA9550-07-1-0351 ($325,414).

Internal

2021 “Aligning human and machine representations of language,” (with Yohei Oseki), Tokyo-Princeton University Collaboration grant ($9,998).

2019-2020 “Society-Scale Behavioral Simulations through Crowdsourcing” (with 4 other faculty members), Center for Statistics and Machine Learning DataX grant ($123,928).

2006-2007 “Computational and statistical foundations of human inductive inference” (with Stuart Russell and Michael Jordan), Chancellor’s Faculty Partnership Fund ($78,985).

2006-2009 Berkeley Committee on Research Junior Faculty Research Grants ($22,000 total).

PUBLICATION LIST

84,000+ citations, *h* index of 126 via Google Scholar:
`https://scholar.google.com/citations?hl=en&user=UAwKvEsAAAAJ`

Books

1. Christian, B., & **Griffiths, T.** (2016). *Algorithms to live by.* New York: Holt. (Named as one of the Amazon.com "Best Science Books of 2016," *Forbes* "Must-read brain books of 2016," and *MIT Technology Review* "Best books of 2016.")

2. **Griffiths, T.L.**, Chater, N., & Tenenbaum, J.B. (2024). *Bayesian models of cognition: Reverse engineering the mind.* MIT Press.

Journal articles

3. Lewandowsky, S., Kalish, M., & **Griffiths, T.L.** (2000). Competing strategies in categorization: Expediency and resistance to knowledge restructuring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1666-1684.

4. Tenenbaum, J.B., & **Griffiths, T.L.** (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*, 629-641. (target article)

5. **Griffiths, T.L.**, & Kalish, M.L. (2002). A multidimensional scaling approach to mental multiplication. *Memory and Cognition, 30*, 97-106.

6. **Griffiths, T.L.**, & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*, 5228-5235.

7. **Griffiths, T.L.**, & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51,* 354-384.

8. Navarro, D.J., **Griffiths, T.L.**, Steyvers, M., & Lee, M.I. (2006). Modeling individual differences with Dirichlet processes. *Journal of Mathematical Psychology, 50*, 101-122.

9. Steyvers, M., **Griffiths, T.L.**, & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences, 10*, 327-334.

10. Tenenbaum, J.B., **Griffiths, T.L.**, & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*, 309-318.

11. **Griffiths, T.L.**, & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science, 17,* 767-773.

12. **Griffiths, T.L.**, & Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition, 103,* 180-226.

13. Kirby, S., Dowman, M., & **Griffiths, T.L.** (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences, 104,* 5241-5245.

14. **Griffiths, T.L.**, & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science, 31,* 441-480.

15. **Griffiths, T.L.**, Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114,* 211-244.

16. Iwata, T., Saito, K., Ueda, N., Stromsten, S., **Griffiths, T.L.**, and Tenenbaum, J. B. (2007). Parametric embedding for class visualization. *Neural Computation, 19,* 2536-2556.

17. Kalish, M.L., **Griffiths, T.L.**, & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin and Review, 14,* 288-294.

18. Schulz, L., Bonawitz, E. B., & **Griffiths, T.L.** (2007). Can being scared make your tummy ache? Naive theories, ambiguous evidence, and preschoolers' causal inferences. *Developmental Psychology, 43,* 1124-1139.

19. **Griffiths, T.L.**, Steyvers, M., & Firl, A. (2007). Google and the mind: Predicting fluency with PageRank. *Psychological Science, 18,* 1069-1076.

20. **Griffiths, T.L.**, Christian, B.R., & Kalish, M.L. (2008). Using category structures to test iterated learning as a method for revealing inductive biases. *Cognitive Science, 32,* 68-107.

21. Goodman, N.D., Tenenbaum, J.B., Feldman, J., & **Griffiths, T.L.** (2008). A rational analysis of rule-based concept learning. *Cognitive Science, 32,* 108-154.

22. Navarro, D.J. & **Griffiths, T.L.** (2008). Latent features in similarity judgment: A nonparametric Bayesian approach. *Neural Computation, 20,* 2597-2628.

23. Dowman, M., Savova, V., **Griffiths, T.L.**, Körding, K., Tenenbaum, J. B., & Purver, M. (2008). A probabilistic model of meetings that combines words and discourse features. *IEEE Transactions on Audio, Speech, and Language Processing, 16,* 1238-1248.

24. **Griffiths, T.L.**, Kalish, M., & Lewandowsky, S. (2008). Theoretical and experimental evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society, 363,* 3503-3514.

25. Reali, F. & **Griffiths, T.L.** (2009). The evolution of linguistic frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition, 111*, 317-328.

26. Goldwater, S., **Griffiths, T.L.** & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition, 112*, 21-54.

27. **Griffiths, T.L.**, & Tenenbaum, J.B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661-716.

28. Feldman, N.H., **Griffiths, T.L.**, & Morgan, J.L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116*, 752-782.

29. Lewandowsky, S., **Griffiths, T.L.**, & Kalish, M.L. (2009). The wisdom of individuals: Exploring peoples knowledge about everyday events using iterated learning. *Cognitive Science*, *33*, 969-998.

30. Xu, J., & **Griffiths, T.L.** (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology, 60,* 107-126.

31. Sanborn, A.N., **Griffiths, T.L.,** & Shiffrin, R. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology, 60,* 63-106.

32. Kemp, C., Tenenbaum, J.B., Niyogi, S., & **Griffiths, T.L.** (2010). A probabilistic model of theory formation. *Cognition, 114,* 165-196.

33. Lucas, C.G., & **Griffiths, T.L.** (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science, 34,* 113-147.

34. Blei, D.M., **Griffiths, T.L.**, & Jordan, M.I. (2010). The nested Chinese restaurant process and Bayesian inference of topic hierarchies. *Journal of the ACM, 57,* 130.

35. Reali, F., & **Griffiths, T.L.** (2010). Words as alleles: Connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society, Series B, 277,* 429-436.

36. Rosen-Zvi, M., Chemudugunta, C., **Griffiths, T.**, Smyth, P., & Steyvers, M. (2010). Learning author-topic models from text corpora. *ACM Transactions on Information Systems, 28,* 1-38.

37. Shi, L., **Griffiths, T.L.,** Feldman, N.H, & Sanborn, A.N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review, 17,* 443-464. (named Best Paper Published in *Psychonomic Bulletin & Review* in 2010)

38. Hsu, A.S., **Griffiths, T.L.,** & Schreiber, E. (2010). Subjective randomness and natural scene statistics. *Psychonomic Bulletin & Review, 17,* 624-629.

39. Sanborn, A.N., **Griffiths, T.L.,** & Navarro, D.J. (2010). Rational approximations to rational models:

Alternative algorithms for category learning. *Psychological Review, 117,* 1144-1167.

40. **Griffiths, T.L.,**, Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J.B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences, 14,* 357-364.

41. Frank, M., Goldwater, S., **Griffiths, T.L.**, & Tenenbaum, J.B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117*, 107-125

42. **Griffiths, T.L.,** & Ghahramani, Z. (2011). The Indian buffet process: An introduction and review. *Journal of Machine Learning Research, 12,* 1185-1224.

43. Tenenbaum, J.B., Kemp, C., **Griffiths, T.L.**, & Goodman, N.D. (2011) How to grow a mind: Statistics, structure, and abstraction. *Science, 331,* 1279-1285.

44. Goldwater, S., **Griffiths, T.L.**, & Johnson, M. (2011). Producing power-law distributions and damping word frequencies with two-stage language models. *Journal of Machine Learning Research, 12,* 2335-2382.

45. Austerweil, J.L., & **Griffiths, T.L.** (2011). Seeking confirmation is rational for deterministic hypotheses. *Cognitive Science, 35,* 499-526.

46. Perfors, A., Tenenbaum, J.B., **Griffiths, T.L.**, & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition, 120,* 302-321.

47. Buchsbaum, D., Gopnik, A., **Griffiths, T.L.**, & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition, 120,* 331-340.

48. **Griffiths, T.L.**, Sobel, D., Tenenbaum, J.B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science, 35,* 1407-1455.

49. **Griffiths, T.L.,** & Tenenbaum, J.B. (2011). Predicting the future as Bayesian inference: People combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General, 140,* 725-743.

50. Austerweil, J.L. & **Griffiths, T.L.** (2011). A rational model of the effects of distributional information on feature learning. *Cognitive Psychology, 63,* 173-209.

51. Martin, J.B., **Griffiths, T.L.**, & Sanborn, A.N. (2012). Testing the efficiency of Markov chain Monte Carlo with people using facial affect categories. *Cognitive Science, 36,* 150-162.

52. **Griffiths, T.L.**, Vul, E., & Sanborn, A.N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science, 21,* 263-268.

53. **Griffiths, T.L.**, & Austerweil, J.L. (2012). Bayesian generalization with circular consequential regions. *Journal of Mathematical Psychology, 56,* 281-285.

54. **Griffiths, T.L.**, Lewandowsky, S., & Kalish, M.L. (2013). The effects of cultural transmission are modulated by the amount of information transmitted. *Cognitive Science, 37,* 953-967.

55. Rafferty, A.N., **Griffiths, T.L.**, & Ettlinger, M. (2013). Greater learnability is not sufficient to produce cultural universals. *Cognition, 129,* 70-87.

56. Denison, S., Bonawitz, E., Gopnik, A., & **Griffiths, T.L.** (2013). Rational variability in children's causal inferences: The sampling hypothesis. *Cognition, 126,* 285-300.

57. Schlerf, J., Xu, J., Klemfuss, N., **Griffiths, T.L.**, & Ivry, R.B. (2013). Individuals with cerebellar degeneration show similar adaptation deficits with large and small visuomotor errors. *Journal of Neurophysiology, 109,* 1164-1173.

58. Bouchard-Côté, A., Hall, D., **Griffiths, T.L.**, & Klein, D. (2013). Automated reconstruction of ancient languages using probabilistic models of sound change. *Proceedings of the National Academy of Sciences, 110,* 4224-4229.

59. Sanborn, A.N., Mansinghka, V.K., & **Griffiths, T.L.** (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review, 120,* 411-437.

60. Feldman, N.H., Myers, E.B., White, K.S., **Griffiths, T.L.**, & Morgan, J.L. (2013). Word-level information influences phonetic learning in adults and infants. *Cognition, 127,* 427-438.

61. Williams, J.J., & **Griffiths, T.L.** (2013). Why are people bad at detecting randomness? A statistical analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1473-1490.

62. Xu, J., Dowman, M., & **Griffiths, T.L.** (2013). Cultural transmission results in convergence toward colour term universals. *Proceedings of the Royal Society B, 280,* 20123073.

63. Austerweil, J., & **Griffiths, T.L.** (2013). A nonparametric Bayesian framework for constructing flexible feature representations. *Psychological Review, 120,* 817-851.

64. Feldman, N.H., **Griffiths, T.L.**, Goldwater, S., & Morgan, J. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review, 120,* 751-778.

65. Vul, E., Goodman, N.D., Tenenbaum, J.B., & **Griffiths, T.L.** (2014). One and done? Optimal decisions from very few samples. *Cognitive Science, 38,* 599-637.

66. Canini, K.R., **Griffiths, T.L.**, Vanpaemel, W., & Kalish, M.L. (2014). Revealing inductive biases for category learning by simulating cultural transmission. *Psychonomic Bulletin & Review, 21,* 785-793.

67. Lucas, C.G., Bridgers, S., **Griffiths, T.L.**, & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition, 131,* 284-299.

68. Shafto, P., Goodman, N.D., & **Griffiths, T.L.** (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology, 71,* 55-89.

69. Lucas, C.G., **Griffiths, T.L.**, Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., Markson, L., & Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PLoS One, 9(3),* e92160.

70. Rafferty, A.N., Zaharia, M., & **Griffiths, T.L.** (2014). Optimally designing games for behavioural research. *Proceedings of the Royal Society A, 470,* 20130828.

71. Bonawitz, E., Denison, S., Gopnik, A., & **Griffiths, T.L.** (2014). Win-stay, lose-sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology, 74,* 35-65.

72. Rafferty, A.N., **Griffiths, T.L.**, & Klein, D. (2014). Analyzing the rate at which languages lose the influence of a common ancestor. *Cognitive Science, 38,* 1406-1431.

73. Bonawitz, E., Denison, S., Gopnik, A., & **Griffiths, T.L.** (2014). Probabilistic models, learning algorithms, response variability: Sampling in cognitive development. *Trends in Cognitive Sciences, 18,* 497-500.

74. Kirby, S., **Griffiths, T.L.**, & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology, 28,* 108-114.

75. Maurits, L., & **Griffiths, T.L.** (2014). Tracing the roots of syntax with Bayesian phylogenetics. *Proceedings of the National Academy of Sciences, 111,* 13576-13581.

76. Rafferty, A.N., Lamar, M.M., & **Griffiths, T.L.** (2015). Inferring learners' knowledge from their actions. *Cognitive Science, 39,* 584-618.

77. **Griffiths, T.L.**, Lieder, F., & Goodman, N.D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science, 7,* 217-229.

78. Buchsbaum, D., **Griffiths, T.L.**, Plunkett, D., Gopnik, A., & Baldwin, D. (2015). Inferring action structure and causal relationships in continuous sequences of human action. *Cognitive Psychology, 76,* 30-77.

79. **Griffiths, T.L.** (2015). Revealing ontological commitments by magic. *Cognition, 136,* 43-48. (*Science* Editors' Choice)

80. Yeung, S., & Griffiths **T.L.** (2015). Identifying expectations about the strength of causal relationships.

Cognitive Psychology, 76, 1-29.

81. Gopnik, A., **Griffiths, T.L.**, & Lucas, C.G. (2015). When younger learners can be better (or at least more open-minded) than older ones. *Current Directions in Psychological Science, 24*, 87-92.

82. Abbott, J.T., Austerweil, J.L., & **Griffiths, T.L.** (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review, 122,* 558-569.

83. Lucas, C.G., **Griffiths, T.L.**, Williams, J.J., & Kalish, M.L. (2015). A rational model of function learning. *Psychonomic Bulletin & Review, 22,* 1193-1215.

84. Bridgers, S., Buchsbaum, D., Seiver, E., **Griffiths, T.L.**, & Gopnik, A. (2015). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology, 52,* 9-18.

85. Hu, J., Lucas, C.G., **Griffiths, T.L.**, & Xu, F. (2015). Preschoolers' understanding of graded preferences. *Cognitive Development, 36,* 93-102.

86. Huth, A.G., de Heer, W.A., **Griffiths, T.L.**, Theunissen, F.E., & Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*, 453-458.

87. **Griffiths, T.L.**, Abbott, J.T., & Hsu, A.S. (2016). Exploring human cognition using large image databases. *Topics in Cognitive Science, 8,* 569-588.

88. Cibelli, E., Xu, Y., Austerweil, J. L., **Griffiths, T.L.**, & Regier, T. (2016). The Sapir-Whorf Hypothesis and probabilistic inference: Evidence from the domain of color. *PLOS One, 11,* 7.

89. Rafferty, A.N., Brunswick, E., **Griffiths, T.L.**, & Shafto, P. (2018). Faster teaching via POMDP planning. *Cognitive Science.*

90. Hsu, A.S., Horng, A., **Griffiths, T.L.**, & Chater, N. (in press). When absence of evidence is evidence of absence: Rational inferences from absent data. *Cognitive Science.*

91. Eaves, B., Feldman, N., **Griffiths, T.L.**, & Shafto, P. (2016) Infant-directed speech is consistent with teaching. *Psychological Review, 123,* 758-771.

92. Abbott, J.T., **Griffiths, T.L.**, Regier, T. (2016). Focal colors and representativeness: Reconciling universals and variation. *Proceedings of the National Academy of Sciences, 113,* 11178-11183.

93. Hamrick, J.B., Battaglia, P.W., **Griffiths, T.L.**, & Tenenbaum, J.B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition, 157,* 61-76.

94. Ruggeri, A., Lombrozo, T., **Griffiths, T.L.**, & Xu, F. (2016). Sources of developmental change in the efficiency of information search. *Developmental Psychology, 52,* 2159-2173.

95. Whalen, A., & **Griffiths, T.L** (2017). Adding population structure to models of language evolution by iterated learning. *Journal of Mathematical Psychology, 76,* 1-6.

96. Austerweil, J.L., **Griffiths, T.L.**, & Palmer, S.E. (2017). Learning to be (in) variant: Combining prior knowledge and experience to infer orientation invariance in object recognition. *Cognitive Science, 41,* 1183-1201.

97. Bramley, N.R., Dayan, P., **Griffiths, T.L.**, & Lagnado, D.A. (2017). Formalizing Neuraths Ship: Approximate algorithms for online causal learning. *Psychological Review, 124,* 301-338.

98. Gopnik, A., O'Grady, S., Lucas, C.G., **Griffiths, T.L.**, Wente, A., Bridgers, S., Aboody, R., Fung, H., & Dahl, R.E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of Sciences, 114,* 7892-7899.

99. Suchow, J. W., Bourgin, D. D, & **Griffiths, T.L** (2017). Evolution in mind: Evolutionary dynamics, cognitive processes, and Bayesian inference. *Trends in Cognitive Sciences, 21,* 522-530.

100. de Heer, W. A., Huth, A. G., **Griffiths, T.L.**, Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *Journal of Neuroscience, 37*(27), 6539-6557.

101. Shenhav, A., Musslick, S., Lieder, F., Kool, W., **Griffiths, T.L.**, Cohen, J. D., & Botvinick, M. M.

(2018). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience.*

102. Lieder, F., **Griffiths, T.L.**, Huys, Q. J. M., & Goodman, N. D. (2018). Empirical evidence for resource-rational anchoring and adjustment. *Psychonomic Bulletin & Review, 25,* 775-784.

103. Lieder, F., **Griffiths, T.L.**, Huys, Q. J. M., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review, 25,* 322-349.

104. Lieder, F., & **Griffiths, T.L.** (2018). Strategy selection as rational metareasoning. *Psychological Review, 124,* 762-794.

105. Lieder, F., **Griffiths, T.L.**, & Hsu, M. (2018). Over-representation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review, 125,* 1-32.

106. Paxton, A., & **Griffiths, T.L.** (2018). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods, 49,* 1630-1638.

107. Whalen, A., **Griffiths, T.L.**, & Buchsbaum, D. (2018). Sensitivity to shared information in social learning. *Cognitive Science, 42,* 168-187.

108. **Griffiths, T.L.**, Daniels, D., Austerweil, J.L., & Tenenbaum, J.B. (2018). Subjective randomness as statistical inference. *Cognitive Psychology, 103,* 85-109.

109. Peterson, J. C., Abbott, J.T., & **Griffiths, T.L.** (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. *Cognitive Science, 42,* 2648-2669.

110. Lieder, F., Shenhav, A., Musslick, S., & **Griffiths, T.L.** (2018).Rational metareasoning and the plasticity of cognitive control. *PLoS Computational Biology, 14*, e1006043.

111. **Griffiths, T.L.**, Callaway, F., Chang, M. B., Grant, E., Krueger, P. M., & Lieder, F. (2019). Doing more with less: meta-reasoning and meta-learning in humans and machines. *Current Opinion in Behavioral Sciences, 29, 24-30.*

112. Lieder, F., Chen, O. X., Krueger, P. M., & **Griffiths, T.L.** (2019). Cognitive prostheses for goal achievement. *Nature human behaviour, 3(10)*, 1096-1106.

113. Austerweil, J. L., Sanborn, S., & **Griffiths, T.L.** (2019). Learning how to generalize. *Cognitive Science, 43,* e12777.

114. Ho, M. K., Abel, D., **Griffiths, T.L.**, & Littman, M. L. (2019). The value of abstraction. *Current Opinion in Behavioral Sciences, 29,* 111116.

115. Hsu, A. S., Martin, J. B., Sanborn, A. N., & **Griffiths, T.L.** (2019). Identifying category representations for complex stimuli using discrete Markov chain Monte Carlo with people. *Behavior Research Methods, 51,* 1706-1716.

116. Jupyter, P., Blank, D., Bourgin, D., Brown, A., Bussonnier, M., Frederic, J., Granger, B., **Griffiths, T.L.**, Hamrick, J., Kelley, K., Pacer, M., Page, L., Perez, F., Ragan-Kelley, B., Suchow, J. W., & Willing, C. (2019). nbgrader: A tool for creating and grading assignments in the Jupyter notebook. *Journal of Open Source Education, 2(11)*, 32.

117. Morgan, T.J., Suchow, J.W., & **Griffiths, T.L.** (2020). What the Baldwin Effect affects depends on the nature of plasticity. *Cognition, 197,* 104165.

118. Dubey, R., & **Griffiths, T.L.** (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review, 127,* 455476.

119. Lieder, F., & **Griffiths, T.L.** (2020). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences, 43,* e1.

120. Agrawal, M., Peterson, J.C., & **Griffiths, T.L.** (2020). Scaling up psychology via Scientific Regret Minimization. *Proceedings of the National Academy of Sciences, 117,* 8825-8835.

121. Gates, V., **Griffiths, T.L.**, & Dragan, A.D. (2020). How to be helpful to multiple people at once.

Cognitive Science, 44, e12841.

122. Morgan, T.J.H, Suchow, J.W. & **Griffiths, T.L.** (2020). Experimental evolutionary simulations of learning, memory and life history. *Philosophical Transactions of the Royal Society B, 375* 20190504.

123. Peterson, J.C., Chen, D. & **Griffiths, T.L.** (2020). Parallelograms revisited: Exploring the limitations of vector space models for simple analogies. *Cognition, 205,* 104440.

124. Battleday, R.M., Peterson, J.C., & **Griffiths, T.L.** (2020). Modeling human categorization of natural images using deep feature representations. *Nature Communications, 11*, 1-14.

125. Dubey, R., & **Griffiths, T.L.** (2020). Understanding exploration in humans and machines by formalizing the function of curiosity. *Current Opinion in Behavioral Sciences, 35,* 118-124.

126. Alon, N., Cohen, J.D., **Griffiths, T.L.**, Manurangsi, P., Reichman, D., Shinkar, I., Wagner, T., & Yu, A. (2020). Multitasking capacity: Hardness results and improved constructions. *SIAM Journal on Discrete Mathematics, 34(1),* 885-903.

127. Rafferty, A.N, Jansen, R.A., & **Griffiths, T.L.** (in press). Assessing mathematics misunderstandings via Bayesian inverse planning. *Cognitive Science.*

128. **Griffiths, T.L.** (2020). Understanding human intelligence via human limitations. *Trends in Cognitive Sciences, 24(11)*, 873-883.

129. Mormann, M., **Griffiths, T.**, Janiszewski, C., Russo, J. E., Aribarg, A., Ashby, N. J., Bagchi, R., Bhatia, S., Kovacheva, M. M., & Mrkva, K. J. (2020). Time to pay attention to attention: using attention-based process traces to better understand consumer decision-making. *Marketing Letters, 31*, 381-392.

130. Thompson, B., & **Griffiths, T.L.** (2021). Human biases limit cumulative innovation. *Proceedings of the Royal Society B, 288,* 20202752.

131. Callaway, F., Rangel, A., & **Griffiths, T.L.**(2021). Fixation patterns in simple choice reflect optimal information sampling. *PLOS Computational Biology, 17(3),* e1008863.

132. Langlois, T. A., Jacoby, N., Suchow, J. W., & **Griffiths, T.L.** (2021). Serial reproduction reveals the geometry of visuospatial representations. *Proceedings of the National Academy of Sciences, 118(13).*

133. Bourgin, D.D., Abbott, J. T., & **Griffiths, T.L.** (2021). Recommendation as generalization: Using big data to evaluate cognitive models. *Journal of Experimental Psychology: General, 150 (7),* 1398-1409.

134. Krafft, P.M., Shmueli, E., **Griffiths, T.L.**, Tenenbaum, J.B., & Pentland, A. (2021). Bayesian collective learning emerges from heuristic social learning. *Cognition, 212,* 104469.

135. Peterson, J. C., Bourgin, D., Agrawal, M., Reichman, D., & **Griffiths, T.L.** (2021). Using large-scale experiments and machine learning to discover theories of human decision-making. *Science, 372 (6547)*, 1209-1214.

136. Battleday, R. M., Peterson, J. C., & **Griffiths, T.L.** (2021). From convolutional neural networks to models of higher-level cognition (and back again). *Annals of the New York Academy of Sciences.*

137. Lewry, C., Curtis, K., Vasilyeva, N., Xu, F., & **Griffiths, T.L.** (2021). Intuitions about magic track the development of intuitive physics. *Cognition, 214,* 104762.

138. Meylan, S. C., Nair, S., & **Griffiths, T.L.** (2021). Evaluating models of robust word recognition with serial reproduction. *Cognition, 210, 104553.*

139. Jansen, R.A., Rafferty, A.N., & **Griffiths, T.L.** (2021) A rational model of the DunningKruger effect supports insensitivity to evidence in low performers. *Nature Human Behavior, 5 (6)*, 756-763.

140. Milli, S., Lieder, F., & **Griffiths, T.L.** (2021). A rational reinterpretation of dual-process theories. *Cognition, 217*, 104881.

141. Ho, M.K. & **Griffiths, T.L.** (2021). Cognitive science as a source of forward and inverse models of human decisions for robotics and control. *Annual Review of Control, Robotics, and Autonomous Systems 5.*

142. Hofman, J.M., Watts, D.J., Athey, S., Garip, F., **Griffiths, T.L.**, Kleinberg, J., Margetts, M., Mullainathan, S., Salganik, M.J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature, 595 (7866)*, 181-188.

143. Meylan, S. & **Griffiths, T.L.** (2021). The challenges of large-scale, web-based language datasets: Word length and predictability revisited. *Cognitive Science, 45,* e12983.

144. Bai, X., Fiske, S. T., & **Griffiths, T.L.** (2022). Globally inaccurate stereotypes can result from locally adaptive exploration. *Psychological Science, 33(5)*, 671684.

145. Barnett, S. A., Hawkins, R. D., & **Griffiths, T.L.** (2021). A pragmatic account of the weak evidence effect. *Open Mind, 6*, 169182.

146. Callaway, F., Jain, Y. R., van Opheusden, B., Das, P., Iwama, G., Gul, S., Krueger, P. M., Becker, F., **Griffiths, T.L.**, & Lieder, F. (2022). Leveraging artificial intelligence to improve peoples planning strategies. *Proceedings of the National Academy of Sciences, 119(12)*, e2117432119.

147. Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., **Griffiths, T.L.**, & Lieder, F. (2022). Rational use of cognitive resources in human planning. Nature Human Behaviour, 6, 1112-1125.

148. Dasgupta, I., & **Griffiths, T.L.** (2022). Clustering and the efficient use of cognitive resources. *Journal of Mathematical Psychology, 109*, 102675.

149. Dubey, R., **Griffiths, T.L.**, & Dayan, P. (2022). The pursuit of happiness: A reinforcement learning perspective on habituation and comparisons. *PLoS Computational Biology, 18(8)*, e1010316.

150. Dubey, R., **Griffiths, T.L.**, & Lombrozo, T. (2022). If its important, then Im curious: Increasing perceived usefulness stimulates curiosity. *Cognition, 226*, 105193.

151. Gates, V., Suchow, J. W., & **Griffiths, T.L.**(2022). Memory transmission in small groups and large networks: An empirical study. *Psychonomic Bulletin & Review, 29(2)*, 581-588.

152. Hardy, M. D., Krafft, P. M., Thompson, B., & **Griffiths, T.L.** (2022). Overcoming individual limitations through distributed computation: Rational information accumulation in multigenerational populations. *Topics in Cognitive Science, 14(3)*, 550573.

153. Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., **Griffiths, T.L.**, & Goodman, N. D. (2022). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review.*

154. Ho, M. K., Abel, D., Correa, C. G., Littman, M. L., Cohen, J. D., & **Griffiths, T.L.**(2022). People construct simplified mental representations to plan. *Nature, 606(7912)*, 129136.

155. Ho, M. K., & **Griffiths, T.L.** (2022). Cognitive science as a source of forward and inverse models of human decisions for robotics and control. *Annual Review of Control, Robotics, and Autonomous Systems, 5*, 33-53.

156. Jain, Y. R., Callaway, F., **Griffiths, T.L.**, Dayan, P., He, R., Krueger, P. M., & Lieder, F. (2022). A computational process-tracing method for measuring peoples planning strategies and how they change over time. *Behavior Research Methods.*

157. Morgan, T. J., Suchow, J. W., & **Griffiths, T.L.** (2022). The experimental evolution of human culture: flexibility, fidelity and environmental instability. *Proceedings of the Royal Society B, 289(1986)*, 20221614.

158. Murthy, S. K., Hawkins, R. D., & **Griffiths, T.L.** (2022). Shades of confusion: Lexical uncertainty modulates ad hoc coordination in an interactive communication task. *Cognition, 225*, 105152.

159. Peterson, J. C., Uddenberg, S., **Griffiths, T.L.**, Todorov, A., & Suchow, J. W. (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences, 119(17)*, e2115228119.

160. Thompson, B., van Opheusden, B., Sumers, T., & **Griffiths, T.L.** (2022). Complex cognitive algorithms preserved by selective social learning in experimental populations. *Science, 376(6588)*, 95-98.

161. Zhang, Q., **Griffiths, T.L.**, & Norman, K. A. (2022). Optimal policies for free recall. *Psychological*

Review, 130(4), 1104.

162. Agrawal, M., Peterson, J. C., Cohen, J. D., & **Griffiths, T.L.** (2023). Stress, intertemporal choice, and mitigation behavior during the COVID-19 pandemic. *Journal of Experimental Psychology: General, 152(9),* 26952702.

163. Brinkmann, L., Baumann, F., Bonnefon, J., Derex, M., Mller, T. F., Nussberger, A., Czaplicka, A., Acerbi, A., **Griffiths, T.L.**, Henrich, J., Leibo, J.Z., McElreath, R., Oudeyer, P., Stray, J., & Rahwan, I. (2023). Machine culture. *Nature Human Behaviour, 7(11),* 1855-1868.

164. Correa, C.G., Ho, M.K., Callaway, F., Daw, N.D., & **Griffiths, T.L.** (2023). Humans decompose tasks by trading off utility and computational cost. *PLOS Computational Biology, 19(6)*, e1011087.

165. **Griffiths, T.L.**, Kumar, S., & McCoy, R. T. (2023). On the hazards of relating representations and inductive biases. *Behavioral and Brain Sciences, 46*, e275.

166. Hardy, M. D., Thompson, B., Krafft, P. M., & **Griffiths, T.L.** (2023). Resampling reduces bias amplification in experimental social networks. *Nature Human Behavior, 7*, 2084-2098.

167. Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., **Griffiths, T.L.**, & Goodman, N. D. (2023). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review, 130(4)*, 9771016.

168. Ho, M. K., Cohen, J. D., & **Griffiths, T.L.** (2023). Rational simplification and rigidity in human planning. Psychological Science, 34(11), 1281-1292.

169. Jain, Y. R., Callaway, F., **Griffiths, T.L.**, Dayan, P., He, R., Krueger, P.M., & Lieder, F. (2023). A computational process-tracing method for measuring peoples planning strategies and how they change over time. *Behavior Research Methods, 55*, 20372079.

170. Jha, A., Peterson, J. C., & **Griffiths, T.L.** (2023). Extracting lowdimensional psychological representations from convolutional neural networks. *Cognitive Science, 47(1)*, e13226.

171. Kumar, S., Dasgupta, I., Daw, N. D., Cohen, J. D., & **Griffiths, T.L.** (2023). Disentangling abstraction from statistical pattern matching in human and machine learning. *PLoS Computational Biology 19(8).*

172. Li, M. Y., Callaway, F., Thompson, W. D., Adams, R., & **Griffiths, T.L.** (2023). Learning to learn functions. *Cognitive Science, 47(4)*, e13262.

173. Reichman, D., Lieder, F., Bourgin, D. D., Talmon, N., & **Griffiths, T.L.** (2023). The computational challenges of means selection problems: Network structure of Goal Systems predicts human performance. *Cognitive Science, 47(8)*, e13330.

174. Shin, M., Kim, J., van Opheusden, B., & **Griffiths, T.L.** (2023). Superhuman artificial intelligence can improve human decision-making by increasing novelty. *Proceedings of the National Academy of Sciences, 120(12)*, e2214840120.

175. Sumers, T. R., Ho, M. K., Hawkins, R. D., & **Griffiths, T.L.** (2023). Show or tell? Exploring when (and why) teaching with language outperforms demonstration. *Cognition, 232*, 105326.

176. Uddenberg, S., Thompson, B.D., Vlasceanu, M., **Griffiths, T.L.**, & Todorov, A. (2023). Iterated learning reveals stereotypes of facial trustworthiness that propagate in the absence of evidence. *Cognition, 237*, 105452.

177. Vélez, N., Christian, B., Hardy, M., Thompson, B. D., & **Griffiths, T.L.** (2023). How do humans overcome individual computational limitations by working together? *Cognitive Science, 47(1)*, e13232.

178. Allen, K., Brndle, F., Botvinick, M., Fan, J. E., Gershman, S. J., Gopnik, A., **Griffiths, T.L.**, Hartshorne, J. K., Hauser, T. U., Ho, M., de Leeuw, J. R., Ma, W. J., Murayama, K., Nelson, J. D., van Opheusden, B., Pouncy, T., Rafner, J., Rahwan, I., Rutledge, R. B., Sherson, J., imek, ., Spiers, H., Summerfield, C., Thalmann, M., Vélez, N., Watrous, A. J., Tenenbaum, J. B., & Schulz, E. (2024). Using games to understand the mind. *Nature Human Behaviour, 8*, 10341043.

179. Almaatouq, A., **Griffiths, T.L.**, Suchow, J. W., Whiting, M. E., Evans, J., & Watts, D. J. (2024). Beyond playing 20 questions with nature: Integrative experiment design in the social and behavioral sciences. *Behavioral and Brain Sciences, 47*, e33.

180. Bai, X., **Griffiths, T.L.**, & Fiske, S. T. (2024). Costly exploration produces stereotypes with dimensions of warmth and competence. *Journal of Experimental Psychology: General.*

181. Collins, K. M., Sucholutsky, I., Bhatt, U., Chandra, K., Wong, L., Lee, M., Zhang, C. E., Zhi-Xuan, T., Ho, M., Mansinghka, V., Weller, A., Tenenbaum, J. B., & **Griffiths, T.L.** (2024). Building machines that learn and think with people. *Nature Human Behaviour, 8(10)*, 18511863.

182. Cornell, C. A., Norman, K. A., **Griffiths, T.L.**, & Zhang, Q. (2024). Improving memory search through model-based cue selection. *Psychological Science, 35(1)*, 5571.

183. Correa, C. G., Sanborn, S., Ho, M. K., Callaway, F., Daw, N. D., & **Griffiths, T.L.** (2024). Exploring the hierarchical structure of human plans via program generation. *Cognition, 255*, 105990.

184. Devraj, A., **Griffiths, T.L.**, & Zhang, Q. (2024). Reconciling categorization and memory via environmental statistics. *Psychonomic Bulletin & Review, 31(5)*, 21182136.

185. Dubey, R., Hardy, M., **Griffiths, T.L.**, & Bhui, R. (2024). AI-generated visuals of car-free American cities help increase support for sustainable transport policies. *Nature Sustainability, 7*, 399403.

186. **Griffiths, T.L.**, Zhu, J. Q., Grant, E., & McCoy, R. T. (2024). Bayes in the age of intelligent machines. *Current Directions in Psychological Science, 33(5)*, 283291.

187. Kumar, S., Sumers, T. R., Yamakoshi, T., Goldstein, A., Hasson, U., Norman, K. A., **Griffiths, T.L.**, Hawkins, R. D., & Nastase, S. A. (2024). Shared functional specialization in transformer-based language models and the human brain. *Nature Communications, 15(1)*, 5523.

188. Lu, Q., Nguyen, T. T., Zhang, Q., Hasson, U., **Griffiths, T.L.**, & Zacks, J. M. (2024). Reconciling shared versus context-specific information in a neural network model of latent causes. *Scientific Reports, 14*(1), 16782.

189. Marjieh, R., Jacoby, N., Peterson, J. C., & **Griffiths, T.L.** (2024). The Universal Law of Generalization holds for naturalistic stimuli. *Journal of Experimental Psychology: General, 153(3)* , 573589.

190. Marjieh, R., Sucholutsky, I., van Rijn, P., Jacoby, N., & **Griffiths, T.L.** (2024). Large language models predict human sensory judgments across six modalities. *Scientific Reports, 14(1)*, 21445.

191. Meylan, S. C., & **Griffiths, T.L.** (2024). Word forms reflect tradeoffs between speaker effort and robust listener recognition. *Cognitive Science, 48(7)*, e13478.

192. McCoy, R. T., Yao, S., Friedman, D., Hardy, M. D., & **Griffiths, T.L.** (2024). Embers of autoregression show how large language models are shaped by the problem they are trained to solve. *Proceedings of the National Academy of Sciences, 121(41)*, e2322420121.

193. Oktar, K., Lombrozo, T., & **Griffiths, T.L.** (2024). Learning from aggregated opinion. *Psychological Science, 35(9)*, 10101024.

194. Reichman, D., Peterson, J. C., & **Griffiths, T.L.** (2024). Machine learning for modeling human decisions. *Decision, 11(4)*, 619632.

195. Russek, E. M., Callaway, F., & **Griffiths, T.L.** (2024). Inverting cognitive models with neural networks to infer preferences from fixations. *Cognitive Science, 48*(11), e70015.

196. Sumers, T. R., Yao, S., Narasimhan, K., & **Griffiths, T.L.** (2024). Cognitive Architectures for Language Agents. *Transactions in Machine Learning Research.* (Outstanding Certification Finalist)

197. Turner, C. R., Morgan, T. J. H., & **Griffiths, T.L.** (2024). Environmental complexity and regularity shape the evolution of cognition. *Proceedings of the Royal Society B, 291(2033)*, 20241524.

198. Bai, X., Wang, A., Sucholutsky, I., & **Griffiths, T.L.** (2025). Explicitly unbiased large language models still form biased associations. *Proceedings of the National Academy of Sciences, 122*(8), e2416228122.

199. Binz, M., Akata, E., Bethge, M., Brndle, F., Callaway, F., Coda-Forno, J., Dayan, P., Demircan, C., Eckstein, M. K., ltet, N., **Griffiths, T.L.**, Haridi, S., Jagadish, A. K., Ji-An, L., Kipnis, A., Kumar, S., Ludwig, T., Mathony, M., Mattar, M., Modirshanechi, A., Nath, S. S., Peterson, J. C., Rmus, M., Russek, E. M., Saanum, T., Scharfenberg, N., Schubert, J. A., Schulze Buschoff, L. M., Singhi, N., Sui, X., Thalmann, M., Theis, F., Truong, V., Udandarao, V., Voudouris, K., Wilson, R., Witte, K., Wu, S., Wulff, D., Xiong, H., & Schulz, E. (2025). A foundation model to predict and capture human cognition. *Nature*, *644*, 10021009.

200. Correa, C. G., Sanborn, S., Ho, M. K., Callaway, F., Daw, N. D., & **Griffiths, T.L.** (2025). Exploring the hierarchical structure of human plans via program generation. *Cognition*, *255*, 105990.

201. Elga, A., Zhu, J. Q., & **Griffiths, T.L.** (2025). People make suboptimal decision about existential risks. *Cognition*, *265*, 106216.

202. Frömer, R., Callaway, F., **Griffiths, T.L.**, & Shenhav, A. (2025). Considering what we know and what we don't know: Expectations and confidence guide value integration in value-based decision-making. *Open Mind*, *9*, 791–813.

203. Gelpí, R. A., Whalen, A., **Griffiths, T.L.**, Xu, F., & Buchsbaum, D. (2025). Can children and adults balance majority size with information quality in learning from preferences? *Journal of Experimental Psychology: General*, *54*, 1388-1406.

204. Ham, H., Zhao, B., **Griffiths, T.L.**, & Velez, N. (2025). Teaching recombinable motifs through simple examples. *Cognitive Science*, *49(8)*, e70103.

205. Liu, G., Snell, J. C., **Griffiths, T.L.**, & Dubey, R. (2025). Binary climate data visuals amplify perceived impact of climate change. *Nature Human Behaviour*, *9(7)*, 1355-1364.

206. Marjieh, R., van Rijn, P., Sucholutsky, I., Lee, H., Jacoby, N., & **Griffiths, T.L.** (2025). Characterizing the large-scale structure of multimodal semantic networks. *Cognitive Science*, *49(10)*, e70131.

207. McCoy, R. T., & **Griffiths, T.L.** (2025). Modeling rapid language learning by distilling Bayesian priors into artificial neural networks. *Nature Communications*, *16(1)*, 4676.

208. Mieczkowski, E., Turner, C., Velez, N., & **Griffiths, T.L.** (2025). People evaluate idle collaborators based on their impact on task efficiency. *Cognition, 264*, 106200.

209. Musslick, S., Bartlett, L. K., Chandramouli, S. H., Dubova, M., Gobet, F., **Griffiths, T.L.**, Hullman, J., King, R. D., Kutz, J. N., Lucas, C. G., Mahesh, S., Pestilli, F., Sloman, S. J., & Holmes, W. R. (2025). Automating the practice of science: Opportunities, challenges, and implications. *Proceedings of the National Academy of Sciences, 122*(5), e2401238121.

210. Russek, E., Acosta-Kane, D., van Opheusden, B., Mattar, M. G., & Griffiths, T. (2025). Time spent thinking in online chess reflects the value of computation. *Cognitive Science, 49(10)*, e70119.

211. Sukhov, N., Dubey, R., Duke, A., & Griffiths, T. (2025). When to keep trying and when to let go: Benchmarking optimal quitting. *Journal of Experimental Psychology: General, 154(9)*, 2599-2618.

212. Zhu, J. Q., Peterson, J. C., Enke, B., & **Griffiths, T.L.** (2025). Capturing the complexity of human strategic decision-making with machine learning. *Nature Human Behaviour*, *9*, 21142120.

213. Gong, T., Pacer, M., **Griffiths, T.L.**, & Bramley, N. R. (2025). Rational causal induction from events in time. *Psychological Review.*

214. Kuperwajs, I., Ruek, E. M., Mattar, M. G., Ma, W. J., & **Griffiths, T.L.** (2025). Looking deeper into the algorithm underlying human planning. *Trends in Cognitive Science.*

215. Zhu, J. Q., & **Griffiths, T.L.** (2025). Computation-limited Bayesian updating: A resource-rational analysis of approximate Bayesian inference. *Psychological Review.*

Peer-reviewed conference papers

216. **Griffiths, T.L.**, & Tenenbaum, J.B. (2000). Teacakes, trains, toxins, and taxicabs: A Bayesian account

of predicting the future. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society.*

217. **Griffiths, T.L.**, & Tenenbaum, J.B. (2001). Randomness and coincidences: Reconciling intuition and probability theory. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.*

218. Tenenbaum, J.B., & **Griffiths, T.L.** (2001). Structure learning in human causal induction. *Advances in Neural Information Processing Systems 13.*

219. Tenenbaum, J.B., & **Griffiths, T.L.** (2001). The rational basis of representativeness. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.*

220. **Griffiths, T.L.**, & Tenenbaum, J.B. (2002). Using vocabulary knowledge in Bayesian multinomial estimation. *Advances in Neural Information Processing Systems 14.*

221. **Griffiths, T.L.**, & Steyvers, M. (2002). A probabilistic approach to semantic representation. *Proceedings of the 24th Annual Conference of the Cognitive Science Society.*

222. **Griffiths, T.L.**, & Tenenbaum, J.B. (2003). Probability, algorithmic complexity, and subjective randomness. *Proceedings of the 25th Annual Conference of the Cognitive Science Society.*

223. Danks, D., **Griffiths, T.L.**, & Tenenbaum, J.B. (2003). Dynamical causal learning. *Advances in Neural Information Processing Systems 15.*

224. **Griffiths, T.L.**, & Steyvers, M. (2003). Prediction and semantic association. *Advances in Neural Information Processing Systems 15.*

225. Tenenbaum, J.B., & **Griffiths, T.L.** (2003). Theory-based causal inference. *Advances in Neural Information Processing Systems 15.*

226. **Griffiths, T.L.**, & Tenenbaum, J.B. (2004). From algorithmic to subjective randomness. *Advances in Neural Information Processing Systems 16.* (winner of best student paper prize – natural systems)

227. Blei, D.M., **Griffiths, T.L.**, Jordan, M.I., & Tenenbaum, J.B. (2004). Hierarchical topic models and the nested Chinese restaurant process. *Advances in Neural Information Processing Systems 16.* (winner of best student paper prize – synthetic systems)

228. Kemp, C. S., **Griffiths, T.L.**, Stromsten, S., & Tenenbaum, J.B. (2004) Semi-supervised learning with trees. *Advances in Neural Information Processing Systems 16.*

229. Steyvers, M., Smyth, P., Rosen-Zvi, M., & **Griffiths, T.** (2004). Probabilistic Author-Topic models for information discovery. *The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*

230. Rosen-Zvi, M., **Griffiths, T.**, Steyvers, M., & Smyth, P. (2004). The Author-Topic model for authors and documents. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI).*

231. **Griffiths, T.L.**, Baraff, E.R., & Tenenbaum, J.B. (2004). Using physical theories to infer hidden causes. *Proceedings of the 26th Annual Conference of the Cognitive Science Society.* (honorable mention for Marr prize for best student paper)

232. **Griffiths, T.L.**, Steyvers, M., Blei, D.M., & Tenenbaum, J.B. (2005). Integrating topics and syntax. *Advances in Neural Information Processing Systems 17.*

233. Iwata, T., Saito, K., Ueda, N., Stromsten, S., **Griffiths, T.**, & Tenenbaum, J. (2005). Parametric embedding for class visualization. *Advances in Neural Information Processing Systems 17.*

234. **Griffiths, T.L.** & Kalish, M.L. (2005). A Bayesian view of language evolution by iterated learning. *Proceedings of the 27th Annual Conference of the Cognitive Science Society.*

235. Navarro, D.J., **Griffiths, T.L.**, Steyvers, M., & Lee, M.I. (2005). Modeling individual differences with Dirichlet processes. *Proceedings of the 27th Annual Conference of the Cognitive Science Society.*

236. Goldwater, S., **Griffiths, T.L.**, & Johnson, M. (2006). Interpolating between types and tokens by estimating power law generators. *Advances in Neural Information Processing Systems 18.*

237. **Griffiths, T.L.**, & Ghahramani, Z. (2006). Infinite latent feature models and the Indian buffet process. *Advances in Neural Information Processing Systems 18.*

238. Dowman, M., Kirby, S., & **Griffiths, T.L.** (2006). Innateness and culture in the evolution of language. In A. Cangelosi, A. D. M. Smith, & K. Smith (Eds.) *The evolution of language: Proceedings of the 6th international conference on language evolution (EVOLANG6)* (pp. 83-90). Hackensack, NJ: World Scientific.

239. Purver, M., Kording, K.P., **Griffiths, T.L.**, & Tenenbaum, J. B. (2006). Unsupervised topic modelling for multi-party spoken discourse. *Proceedings of COLING/ACL 2006.*

240. Goldwater, S., **Griffiths, T.L.**, & Johnson, M. (2006). Contextual dependencies in unsupervised word segmentation. *Proceedings of COLING/ACL 2006.*

241. Mansinghka, V.K., Kemp, C., Tenenbaum, J.B., & **Griffiths, T.L.** (2006). Structured priors for structure learning. *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI 2006).*

242. Wood, F., **Griffiths, T.L.**, & Ghahramani, Z. (2006). A non-parametric Bayesian method for inferring hidden causes. *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence (UAI 2006).*

243. Kemp, C., Tenenbaum, J. B., **Griffiths, T.L.**, Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI '06).*

244. **Griffiths, T.L.**, Christian, B.R., & Kalish, M.L. (2006). Revealing priors on category structures through iterated learning. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.*

245. Bonawitz, E.B., **Griffiths, T.L.**, & Schulz, L. (2006). Modeling cross-domain causal learning in preschoolers as Bayesian inference. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.* (winner of Marr prize for best student paper)

246. Sanborn, A.N., **Griffiths, T.L.**, & Navarro, D.J. (2006). A more rational model of categorization. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.*

247. Goldwater, S., **Griffiths, T.L.**, & Johnson, M. (2007). Distributional cues to word segmentation: Context is important. *Proceedings of the 31st Boston University Conference on Language Development.*

248. Johnson, M., **Griffiths, T.L.**, & Goldwater, S. (2007). Bayesian inference for PCFGs via Markov chain Monte Carlo. *Proceedings of the North American Conference on Computational Linguistics (NAACL'07).*

249. Goldwater, S., & **Griffiths, T.L.** (2007). A fully Bayesian approach to unsupervised part-of-speech tagging. *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics.*

250. Bouchard-Côté, A., Liang, P., **Griffiths, T.L.**, & Klein, D. (2007). A probabilistic approach to diachronic phonology. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL).*

251. Wood, F., & **Griffiths, T.L.** (2007). Particle filtering for nonparametric Bayesian matrix factorization. *Advances in Neural Information Processing Systems 19.*

252. Johnson, M., **Griffiths, T.L.**, & Goldwater, S. (2007). Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models. *Advances in Neural Information Processing Systems 19.*

253. Navarro, D. J., & **Griffiths, T.L.** (2007). A nonparametric Bayesian method for inferring features from similarity judgments. *Advances in Neural Information Processing Systems 19.*

254. Schreiber, E., & **Griffiths, T.L.** (2007). Subjective randomness and natural scene statistics. *Proceedings of the 29th Annual Conference of the Cognitive Science Society.*

255. Feldman, N., & **Griffiths, T.L.** (2007). A rational account of the perceptual magnet effect. *Proceedings of the 29th Annual Conference of the Cognitive Science Society.*

256. **Griffiths, T.L.**, Canini, K. R., Sanborn A. N., & Navarro, D. J. (2007). Unifying rational models of categorization via the hierarchical Dirichlet process. *Proceedings of the 29th Annual Conference of the*

Cognitive Science Society.

257. Frank, M., Goldwater, S., **Griffiths, T.L.**, & Tenenbaum, J. B. (2007). Modeling human performance in statistical word segmentation. *Proceedings of the 29th Annual Conference of the Cognitive Science Society.*

258. Goodman, N., **Griffiths, T.L.**, Feldman, J., & Tenenbaum, J. B. (2007). A rational analysis of rule-based concept learning. *Proceedings of the 29th Annual Conference of the Cognitive Science Society.*

259. Sanborn, A. N., & **Griffiths, T.L.** (2008). Markov chain Monte Carlo with people. *Advances in Neural Information Processing Systems 20.* (winner of the Outstanding Student Paper prize)

260. Bouchard-Côté, A., Liang, P., **Griffiths, T.L.**, & Klein, D. (2008). A probabilistic approach to language change. *Advances in Neural Information Processing Systems 20.*

261. Reali, F., & **Griffiths, T.L.** (2008). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

262. Xu, J., Reali, F., & **Griffiths, T.L.** (2008). A formal analysis of cultural evolution by replacement. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

263. Austerweil, J., & **Griffiths, T.L.** (2008). A rational analysis of confirmation with deterministic hypotheses. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

264. Williams, J.J., & **Griffiths, T.L.** (2008). Why are people bad at detecting randomness? Because it is hard. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

265. Shi, L., Feldman, N.H., & **Griffiths, T.L.** (2008). Performing Bayesian inference with exemplar models. *Proceedings of the 30th Annual Conference of the Cognitive Science Society.*

266. Miller, K. T., **Griffiths, T.L.**, & Jordan, M. I. (2008). The phylogenetic Indian buffet process: A non-exchangeable nonparametric prior for latent features. *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI 2008).*

267. **Griffiths, T.L.**, Lucas, C., Williams, J.J., & Kalish, M.L. (2009). Modeling human function learning with Gaussian processes. *Advances in Neural Information Processing Systems 21.*

268. Levy, R., Reali, F., & **Griffiths, T.L.** (2009). Modeling the effects of memory on human online sentence processing with particle filters. *Advances in Neural Information Processing Systems 21.*

269. Xu, J. & **Griffiths, T.L.** (2009). How memory biases affect information transmission: A rational analysis of serial reproduction. *Advances in Neural Information Processing Systems 21.*

270. Lucas, C., **Griffiths, T.L.**, Xu, F., & Fawcett, C. (2009). A rational model of preference learning and choice prediction by children. *Advances in Neural Information Processing Systems 21.*

271. Austerweil, J. & **Griffiths, T.L.** (2009). Analyzing human feature learning as nonparametric Bayesian inference. *Advances in Neural Information Processing Systems 21.*

272. Bouchard-Côté, A., **Griffiths, T.L.**, & Klein, D. (2009). Improved reconstruction of protolanguage word forms.*Proceedings of the North American Conference on Computational Linguistics (NAACL'09).*

273. Canini, K. R., Shi, L., & **Griffiths, T.L.** (2009). Online inference of topics with Latent Dirichlet Allocation. *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics (AISTATS) 2009.*

274. Vul, E., Goodman, N. D., **Griffiths, T.L.**, & Tenenbaum, J. B. (2009). One and done? Optimal decisions from very few samples. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

275. Austerweil, J. L., & **Griffiths, T.L.** (2009). The effect of distributional information on feature learning. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

276. Beppu, A., & **Griffiths, T.L.** (2009). Iterated learning and the cultural ratchet. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

277. Buchsbaum, D., **Griffiths, T.L.**, Gopnik, A., & Baldwin, D. (2009). Learning from actions and their consequences: Inferring causal variables from continuous sequences of human action. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

278. Feldman, N. H., **Griffiths, T.L.**, & Morgan, J. L. (2009). Learning phonetic categories by learning a lexicon. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

279. Rafferty, A., **Griffiths, T.L.**, & Klein, D. (2009). Convergence bounds for language evolution by iterated learning. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

280. Sanborn, A. N., Mansinghka, V. K., & **Griffiths, T.L.** (2009). A Bayesian framework for modeling intuitive dynamics. *Proceedings of the 31st Annual Conference of the Cognitive Science Society.*

281. Hsu, A., & **Griffiths, T.L.** (2009). Differential use of implicit negative evidence in generative and discriminative language learning. *Advances in Neural Information Processing Systems 22.*

282. Miller, K. T., **Griffiths, T.L.**, & Jordan, M. I. (2009). Nonparametric latent feature models for link prediction. *Advances in Neural Information Processing Systems 22.*

283. Shi, L., & **Griffiths, T.L.** (2009). Neural implementation of hierarchical Bayesian inference by importance sampling. *Advances in Neural Information Processing Systems 22.*

284. Burkett, D., & **Griffiths, T.L.** (2010). Iterated learning of multiple languages from multiple teachers. *Evolang 8.*

285. Canini, K.R., Shashkov, M.M., & **Griffiths, T.L.** (2010). Modeling transfer learning in human categorization with the hierarchical Dirichlet process. *Proceedings of the 27th International Conference on Machine Learning.* (winner of Best Application Paper award)

286. Hsu, A.S. & **Griffiths, T.L.** (2010). Effects of generative and discriminative learning on use of category variability. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

287. Rafferty, A.N., & **Griffiths, T.L.** (2010). Optimal language learning: The importance of starting representative. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

288. Buchsbaum, D., Gopnik, A., & **Griffiths, T.L.** (2010). Children's imitation of action sequences is influenced by statistical evidence and inferred causal structure. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

289. Bonawitz, E.B., & **Griffiths, T.L.** (2010). Deconfounding hypothesis generation and evaluation in Bayesian models. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

290. Denison, S., Bonawitz, E.B., Gopnik, A., & **Griffiths, T.L.** (2010). Preschoolers sample from probability distributions. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

291. Austerweil, J.L., & **Griffiths, T.L.** (2010). Learning hypothesis spaces and dimensions through concept learning. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

292. Lucas, C.G., Gopnik, A., & **Griffiths, T.L.** (2010) Developmental differences in learning the forms of causal relationships. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

293. Xu, J., **Griffiths, T.L.,** & Dowman, M. (2010). Replicating color term universals through human iterated learning. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society.*

294. Austerweil, J.L. & **Griffiths, T.L.** (2010). Learning invariant features using the transformed Indian buffet process. *Advances in Neural Information Processing Systems 23.*

295. Feldman, N., Myers, E., White, K., **Griffiths, T.**, & Morgan, J. (2011). Learners use word-level statistics in phonetic category acquisition. *Proceedings of the 35th Boston University Conference on Language Development.*

296. Rafferty, A. N., **Griffiths, T.L.**, & Ettlinger, M. (2011). Exploring the relationship between learnability and linguistic universals. *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics at ACL 2011.*

297. Rafferty, A. N., Brunswick, E., **Griffiths, T.L.,** & Shafto, P. (2011). Faster teaching by POMDP planning. *Proceedings of the 16th International Conference on Artificial Intelligence in Education, (AIED11).*

298. Yeung, S. & **Griffiths, T.L.** (2011). Estimating human priors on causal strength. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

299. Canini, K.R., Vanpaemel, W., **Griffiths, T.L.,** & Kalish, M.L. (2011). Discovering inductive biases in categorization through iterated learning. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

300. Waisman, A., Lucas, C.G., **Griffiths, T.L.,** & Jacobs, L.F. (2011). A Bayesian model of navigation in squirrels. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

301. Buchsbaum, D., Canini, K.R., & **Griffiths, T.L.** (2011). Segmenting and recognizing human action using low-level video features. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

302. Abbott, J. & **Griffiths, T.L.** (2011). Exploring the influence of particle filter parameters on order effects in causal learning. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society.*

303. Canini, K.R., & **Griffiths, T.L.** (2011). A nonparametric Bayesian model of multi-level category learning. *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence.*

304. Austerweil, J.L., Friesen, A., & **Griffiths, T.L.** (2011). An ideal observer model for identifying the reference frame of objects. *Advances in Neural Information Processing Systems 24.*

305. Pacer, M., & **Griffiths, T.L.** (2011). A rational model of causal inference with continuous causes. *Advances in Neural Information Processing Systems 24.*

306. Abbott, J., Heller, K., **Griffiths, T.L.**, & Ghahramani, Z. (2011). Testing a Bayesian measure of representativeness using a large image database. *Advances in Neural Information Processing Systems 24.*

307. Pacer, M., & **Griffiths, T.L.** (2012). Elements of a rational framework for continuous-time causal induction. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

308. Abbott, T.J., Regier, T., & **Griffiths, T.L.** (2012). Predicting focal colors with a rational model of representativeness. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

309. Hsu, A.S., Martin, J.B., Sanborn, A.N., & **Griffiths, T.L.** (2012). Identifying representations of categories of discrete items using Markov chain Monte Carlo with people. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

310. Buchsbaum, D., Bridgers, S., Whalen, A., Seiver, E., **Griffiths, T.L.**, & Gopnik, A. (2012). Do I know that you know what you know? Modeling testimony in causal inference. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

311. Blundell, C., Sanborn, A., & **Griffiths, T.L.** (2012). Look-ahead Monte Carlo with people. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

312. Abbott, J., Austerweil, J.L., & **Griffiths, T.L.** (2012). Constructing a hypothesis space from the Web for large-scale Bayesian word learning. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

313. **Griffiths, T.L.**, Austerweil, J.L., & Berthiaume, V. (2012). Comparing the inductive biases of simple neural networks and Bayesian models. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

314. Rafferty, A.N., Zaharia, M., & **Griffiths, T.L.** (2012). Optimally designing games for cognitive science research. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

315. Little, D., Lewandowsky, S., & **Griffiths, T.L.** (2012). A Bayesian model of rule induction in Raven's progressive matrices. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*

316. Lieder, F., Goodman, N.D., & **Griffiths, T.L.** (2013). Burn-in, bias, and the rationality of anchoring. *Advances in Neural Information Processing Systems 25.*

317. Abbott, J., Austerweil, J.L., & **Griffiths, T.L.** (2013). Human memory search as a random walk in a semantic network. *Advances in Neural Information Processing Systems 25.*

318. Abbott, J. T., Hamrick, J. B., & **Griffiths, T.L.** (2013). Approximating Bayesian inference with a sparse distributed memory system. *Proceedings of the 35th Annual Conference of the Cognitive Science Society.*

319. Hu, J. C., Buchsbaum, D., **Griffiths, T.L.**, & Xu, F. (2013). When does the majority rule? Preschoolers' trust in majority informants varies by task domain. *Proceedings of the 35th Annual Conference of the Cognitive Science Society.*

320. Whalen, A., Buchsbaum, D., & **Griffiths, T.L.** (2013). How do you know that? Sensitivity to statistical dependency in social learning. *Proceedings of the 35th Annual Conference of the Cognitive Science Society.*

321. Pacer, M., Williams, J., Chen, X., Lombrozo, T., & **Griffiths, T.L.** (2013). Evaluating computational models of explanation using human judgments. *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence (UAI 2013).*

322. Jia, Y., Abbott, J., Austerweil, J.A., **Griffiths, T.L.**, & Darrell, T. (2013). Visual concept learning: Combining machine vision and Bayesian generalization on concept hierarchies. *Advances in Neural Information Processing Systems 26.*

323. Bertolero, M. A., & **Griffiths, T.L.** (2014). Is holism a problem for inductive inference? A computational analysis. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

324. Bourgin, D. D., Abbott, J. T., Griffiths, T.L., Smith, K. A., & Vul, E. (2014). Empirical evidence for Markov chain Monte Carlo in memory search. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

325. Hamrick, J., & **Griffiths, T.L.** (2014). What to simulate? Inferring the right direction for mental rotation. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

326. Lieder, F., Hsu, M., & **Griffiths, T.L.** (2014). The high availability of extreme events serves resource-rational decision-making. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

327. Neumann, R., Rafferty, A. N., & **Griffiths, T.L.** (2014). A bounded rationality account of wishful thinking. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

328. Press, A., Pacer, M., **Griffiths, T.L.**, & Christian, B. (2014). Caching algorithms and rational models of memory. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

329. Whalen, A., Maurits, L., Pacer, M., & **Griffiths, T.L.** (2014). Cultural evolution with sparse testimony: When does the cultural ratchet slip? *Proceedings of the 36th Annual Conference of the Cognitive Science Society.*

330. Rafferty, A. N., & **Griffiths, T.L.** (2015). Interpreting freeform equation solving. *Proceedings of the 17th International Conference on Artificial Intelligence in Education.*

331. Hamrick, J. B., Smith, K. A., **Griffiths, T.L.**, & Vul, E. (2015). Think again? The amount of mental simulation tracks uncertainty in the outcome. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

332. Hu, J., Whalen, A., Buchsbaum, D., **Griffiths, T.L.**, & Xu, F. (2015). Can children balance the size of a majority with the quality of their information? *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

333. Lieder, F., & **Griffiths, T.L.** (2015). When to use which heuristic: A rational solution to the strategy selection problem. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

334. Lieder, F., Sim, Z., Hu, J. C., & **Griffiths, T.L.** (2015). Children and adults differ in their strategies for social learning. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

335. Meylan, S. C., & **Griffiths, T.L.** (2015). A Bayesian framework for learning words from multiword utterances. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

336. Morgan, T. J. H., & **Griffiths, T.L.** (2015). What the Baldwin Effect affects. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

337. Pacer, M. D., & **Griffiths, T.L.** (2015). Upsetting the contingency table: Causal induction over sequences of point events. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

338. Ruggeri, A., Lombrozo, T., **Griffiths, T.L.**, & Xu, F. (2015). Children search for information as efficiently as adults, but seek additional confirmatory evidence. *Proceedings of the 37th Annual Conference of the Cognitive Science Society.*

339. Liu, C., Hamrick, J. B., Fisac, J. F., Dragan, A. D., Hedrick, J. K., Sastry, S. S., & **Griffiths, T.L.** (2016). Goal Inference Improves Objective and Perceived Performance in Human-Robot Collaboration. *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016).*

340. Suchow, J. W., Pacer, M. D., **Griffiths, T.L.** (2016). Design from zeroth principles. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

341. Suchow, J. W., **Griffiths, T.L.** (2016). Deciding to remember: Memory maintenance as a Markov Decision Process. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

342. Peterson, J., Abbott, J. T., **Griffiths, T.L.** (2016). Adapting deep network features to capture psychological representations. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

343. O'Grady, S., **Griffiths, T.L.**, Xu, F. (2016). Do simple probability judgments rely on integer approximation? *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

344. Foushee, R, **Griffiths, T.L.**, & Srinivasan, M. (2016). Lexical complexity of child-directed and overheard speech: Implications for learning. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

345. Milli, S., Lieder, F., & **Griffiths, T.L.** (2017). When does bounded-optimal metareasoning favor few cognitive systems?. *AAAI.*

346. Meng, Y., **Griffiths, T.L.**, & Xu, F. (2017). Inferring intentional agents from violation of randomness. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

347. Langlois, T. A., Jacoby, N., Suchow, J.W., & **Griffiths, T.L.** (2017). Uncovering visual priors in spatial memory using serial reproduction. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

348. Bourgin, D. D., Lieder, F., Reichman, D., Talmon, N., & **Griffiths, T.L.** (2017). The structure of goal systems predicts human performance. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

349. Dubey, R., & **Griffiths, T.L.** (2017). A rational analysis of curiosity. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

350. Grant, E., Nematzadeh, A., & **Griffiths, T.L.** (2017). How can memory-augmented neural networks pass a false-belief task? *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

351. Nematzadeh, A., Meylan, S.C., & **Griffiths, T.L.** (2017). Evaluating vector-space models of word representation, or the unreasonable effectiveness of counting words near other words. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

352. Jansen, R. A., Rafferty, A. N., & **Griffiths, T.L.** (2017). Algebra is not like trivia: Evaluating self-assessment in an online math tutor. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

353. Krueger, P.M., Lieder, F., & **Griffiths, T.L.** (2017). Enhancing metacognitive reinforcement learning using reward structures and feedback. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

354. Lieder, F., Krueger, P. M., & **Griffiths, T.L.** (2017). An automatic method for discovering rational

heuristics for risky choice. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

355. Callaway, F., Hamrick, J. B., & **Griffiths, T.L.** (2017). Discovering simple heuristics from mental simulation. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

356. Chen, D., Peterson, J. C., & **Griffiths, T.L.** (2017). Evaluating vector-space models of analogy. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

357. Gates, M. A., Suchow, J. W., & **Griffiths, T.L.** (2017). Empirical tests of large-scale collaborative recall. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

358. Peterson, J. C., & **Griffiths, T.L.** (2017). Evidence for the size principle in semantic and perceptual domains. *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*

359. Fisac, J. F., Gates, M. A., Hamrick, J. B., Liu, C., Hadfield-Menell, D., Palaniappan, M., Malik, D., Sastry, S. S., **Griffiths, T.L.**, & Dragan, A. D. (2017). Pragmatic-pedagogic value alignment. *International Symposium on Robotics Research.*

360. Grant, E., Finn, C., Levine, S., Darrell, T., & **Griffiths, T.L.** (2018). Recasting gradient-based meta-learning as hierarchical Bayes. *Proceedings of the 6th International Conference on Learning Representations (ICLR).*

361. Krueger, P. M., & **Griffiths, T.L.** (2018). Shaping model-free habits with model-based goals. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

362. Callaway, F., Lieder, F., Das, P., Gul, S., Krueger, P.M., & **Griffiths, T.L.** (2018). A resource-rational analysis of human planning. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

363. Bourgin, D.D., Abbott, J.T., & **Griffiths, T.L.** (2018). Recommendation as generalization: Evaluating cognitive models in the wild. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

364. Krafft, P.M., & **Griffiths, T.L.** (2018). Levels of analysis in computational social science. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

365. Sanborn, S., Bourgin, D.D., Chang, M., & **Griffiths, T.L.** (2018). Representational efficiency outweighs action efficiency in human program induction. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

366. Peterson, J.C., Suchow, J.W., Aghi, K., Ku, A.Y., & **Griffiths, T.L.** (2018). Capturing human category representations by sampling in deep feature spaces. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

367. Suchow, J.W., Peterson, J. C., & **Griffiths, T.L.** (2018). Learning a face space for experiments on human identity. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

368. Peterson, J.C., Soulos, P., Nematzadeh, A., & **Griffiths, T.L.** (2018). Learning hierarchical visual representations in deep neural networks using hierarchical linguistic labels. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

369. Jansen, R.A., Rafferty, A.N., & **Griffiths, T.L.** (2018). Modeling the Dunning-Kruger Effect: A rational account of inaccurate self assessment. *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

370. Dubey, R., Agrawal, P., Pathak, D., **Griffiths, T.L.**, & Efros, A.A. (2018). Investigating human priors for playing video games. *Proceedings of the International Conference on Machine Learning (ICML).*

371. Nematzadeh, A., Burns, K., Grant, E., Gopnik, A., & **Griffiths, T.L.** (2018). Evaluating theory of mind in question answering. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*

372. Callaway, F., Gul, S., Krueger, P. M., **Griffiths, T.L.**, & Lieder, F. (2018). Learning to select computations. *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI 2018).*

373. Grant, E., Peterson, J., & **Griffiths, T.L.** (2019). Learning deep taxonomic priors for concept learning

from few positive examples. *Proceedings of the 41st Annual Conference of the Cognitive Science Society.*

374. Peterson, J., Battleday, R., **Griffiths, T.L.**, & Russakovsky, O. (2019). Human uncertainty makes classification more robust. *Proceedings of the IEEE International Conference on Computer Vision.*

375. Dubey, R., **Griffiths, T.L.**, & Lombrozo, T. (2019). If its important, then I am curious: A value intervention to induce curiosity. *Proceedings of the 41st Annual Conference of the Cognitive Science Society.*

376. Chang, M. B., Gupta, A., Levine, S., & **Griffiths, T.L.** (2019). Automatically composing representation transformations as a means for generalization. *Proceedings of the 7th International Conference on Learning Representations (ICLR).*

377. Thompson, B., & **Griffiths, T.L.** (2019). Inductive biases constrain cumulative cultural evolution. *Proceedings of the 41st Annual Conference of the Cognitive Science Society.*

378. Ho, M. K., Korman, J., & **Griffiths, T.L.** (2019). The computational structure of unintentional meaning. *Proceedings of the 41st Annual Conference of the Cognitive Science Society.*

379. Agrawal, M., Peterson, J.C., & **Griffiths, T.L.** (2019). Using machine learning to guide cognitive modeling: a case study in moral reasoning. *Proceedings of the 41st Annual Conference of the Cognitive Science Society.*

380. Bourgin, D., Peterson, J., Reichman, D., Russell, S., & **Griffiths, T.L.** (2019). Cognitive model priors for predicting human decisions. *Proceedings of the 36th International Conference on Machine Learning (ICML).*

381. Jerfel, G., Grant, E. L., **Griffiths, T.L.**, & Heller, K. (2019). Reconciling meta-learning and continual learning with online mixtures of tasks. *Advances in Neural Information Processing Systems.*

382. Carroll, M., Shah, R., Ho, M. K., **Griffiths, T.**, Seshia, S., Abbeel, P., & Dragan, A. (2019). On the Utility of Learning about Humans for Human-AI Coordination. *Advances in Neural Information Processing Systems.*

383. Ho, M.K., Abel, D., Cohen, J.D., Littman, M.L., & **Griffiths, T.L.** (2020). The efficiency of human cognition reflects planned information processing. *Proceedings of the 34th AAAI Conference on Artificial Intelligence.*

384. Singh, P., Peterson, J. C., Battleday, R. M., & **Griffiths, T.L.** (2020). End-to-end deep prototype and exemplar models for predicting human behavior. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

385. Sumers, T. R., Ho, M. K., & **Griffiths, T.L.** (2020). Show or tell? Demonstration is more robust to changes in shared perception than explanation. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

386. McCoy, R. T., Grant, E., Smolensky, P., **Griffiths, T.L.**, & Linzen, T. (2020). Universal linguistic inductive biases via meta-learning. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

387. Jansen, R. A., Rafferty, A. N., & **Griffiths, T.L.** (2020). A rational model of sequential self-assessment. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

388. Jha, A., Peterson, J., & **Griffiths, T.L.** (2020). Extracting low-dimensional psychological representations from convolutional neural networks. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

389. Hawkins, R. D., Goodman, N. D., Goldberg, A. E., & **Griffiths, T.L.** (2020). Generalizing meanings from partners to populations: Hierarchical inference supports convention formation on networks. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

390. Correa, C. G., Ho, M.K., Callaway, F., & **Griffiths, T.L.** (2020). Resource-rational task decomposition to minimize planning costs. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society.*

391. Sumers, T. R., Ho, M. K., Hawkins, R.D., Narasimhan, K., & **Griffiths, T.L.** (2021). Learning rewards

from linguistic feedback. *Proceedings of the 35th AAAI Conference on Artificial Intelligence.*

392. Kumar, S., Dasgupta, I., Cohen, J. D., Daw, N. D., & **Griffiths, T.L.** (2021). Meta-learning of structured task distributions in humans and machines. *Proceedings of the 9th International Conference on Learning Representations (ICLR).*

393. Devraj, A., Zhang, Q., & **Griffiths, T.L.** (2021). The dynamics of exemplar and prototype representations depend on environmental statistics. *Proceedings of the 43rd Annual Conference of the Cognitive Science Society.*

394. Wilson, S., Arora, S., Zhang, Q., & **Griffiths, T.L.** (2021). A rational account of anchor effects in hindsight bias. *Proceedings of the 43rd Annual Conference of the Cognitive Science Society.*

395. Sumers, T. R., Hawkins, R.D., Ho, M. K., & **Griffiths, T.L.** (2021). Extending rational models of communication from beliefs to actions. *Proceedings of the 43rd Annual Conference of the Cognitive Science Society.*

396. Tuli, S., Dasgupta, I., Grant, E., & **Griffiths, T.L.** (2021). Are convolutional neural networks or transformers more like human vision? *Proceedings of the 43rd Annual Conference of the Cognitive Science Society.*

397. Langlois, T.A., Zhao, H.C., Grant, E., Dasgupta, I **Griffiths, T.L.**, & Jacoby, N. (2021). Passive attention in artificial neural networks predicts human visual selectivity. *Advances in Neural Information Processing Systems 35.*

398. Chang, M., **Griffiths, T.L.**, & Levine, S. (2022). Object representations as fixed points: Training iterative refinement algorithms with implicit differentiation. *Advances in Neural Information Processing Systems 36.*

399. Dasgupta, I., Grant, E., & **Griffiths, T.L.** (2022). Distinguishing rule- and exemplar-based generalization in learning systems. *Proceedings of the International Conference on Machine Learning.*

400. Kumar, S., Correa, C. G., Dasgupta, I., Marjieh, R., Hu, M. Y., Hawkins, R.D., Daw, N. D., Cohen, J. D., Narasimhan, K. R., & **Griffiths, T.L.** (2022). Using natural language and program abstractions to instill human inductive biases in machines. *Advances in Neural Information Processing Systems 36.* (Outstanding Paper Award recipient.)

401. Malaviya, M., Sucholutsky, I., Oktar, K., & **Griffiths, T.L.** (2022). Can humans do less-than-one-shot learning? *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

402. Marjieh, R., Sucholutsky, I., Sumers, T. R., Jacoby, N., & **Griffiths, T.L.** (2022). Predicting Human Similarity Judgments Using Large Language Models. *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

403. Sumers, T. R., Hawkins, R. D., Ho, M. K., **Griffiths, T.L.**, & Hadfield-Menell, D. (2022). How to talk so your robot will learn: Instructions, descriptions, and pragmatics. *Advances in Neural Information Processing Systems 36.*

404. Yamakoshi, T., **Griffiths, T.L.**, & Hawkins, R.D. (2022). Probing BERT's priors with serial reproduction chains. *Findings of the Association for Computational Linguistics (ACL).*

405. Chang, M., Dayan, A.L., Meier, F., **Griffiths, T.L.**, Levine, S., & Zhang, A. (2023). Neural Constraint Satisfaction: Hierarchical abstraction for combinatorial generalization in object rearrangement. *Proceedings of the 11th International Conference on Learning Representations.*

406. Zhu, J. Q., Sanborn, A., Chater, N., & **Griffiths, T.** (2023). Computation-Limited Bayesian updating. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society.*

407. Yao, S., Yu, D., Zhao, J., Shafran, I., **Griffiths, T.L.**, Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems 37.*

408. Wang, Z., Ku, A., Baldridge, J., **Griffiths, T.L.**, & Kim, B. (2023). Gaussian Process Probes (GPP) for

uncertainty-aware probing. *Advances in Neural Information Processing Systems 37.*

409. Xia, F., Zhu, J., & **Griffiths, T.** (2023). Comparing human predictions from expert advice to on-line optimization algorithms. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society.*

410. Sucholutsky, I., & *Griffiths, T.L.* (2023). Alignment with human representations supports robust few-shot learning. *Advances in Neural Information Processing Systems 37.*

411. Dedhia, B., Chang, M., Snell, J.C., **Griffiths, T.L.**, & Jha, N. K. (2023). Im-Promptu: In-context composition from image prompts. *Advances in Neural Information Processing Systems 37.*

412. Sucholutsky, I., Battleday, R., Collins, K., Marjieh, R., Peterson, J.C., Singh, P., Bhatt, U., Jacoby, N., Weller, A., & **Griffiths, T.L.** (2023). On the informativeness of supervision signals. *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence.*

413. Li, M. Y., Grant, E., & **Griffiths, T.L.** (2023). Gaussian process surrogate models for neural networks. *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence.*

414. Rane, S., Nencheva, M.L., Wang, Z., Lew-Williams, C., Russakovsky, O., & **Griffiths, T.L.** (2023). Predicting word learning in children from the performance of computer vision systems. Proceedings of the 45th Annual Meeting of the Cognitive Science Society.

415. Marjieh, R., Sucholutsky, I., van Rijn, P., Jacoby, N., & **Griffiths, T.L.** (2023). What language reveals about perception: Distilling psychophysical knowledge from large language models. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society.*

416. Peterson, J., Mancoridis, M., & **Griffiths, T.** (2023). To each their own theory: Exploring the limits of individual differences in decisions under risk. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society.*

417. Barretto, D., Marjieh, R., & **Griffiths, T.L.** (2024). Reaching Consensus through Theory of Mind in Social Networks with Locally Distributed Interactions. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

418. Bencomo, G. M., Snell, J. C., & **Griffiths, T.L.** (2024). Implicit Maximum a Posteriori Filtering via adaptive optimization. *Proceedings of the International Conference on Learning Representations.*

419. Campbell, D., Kumar, S., Giallanza, T., **Griffiths, T.L.**, & Cohen, J. D. (2024). Human-Like Geometric Abstraction in Large Pre-trained Neural Networks. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

420. Campbell, D., Rane, S., Giallanza, T., De Sabbata, N., Ghods, K., Joshi, A., Ku, A., Frankland, S. M., **Griffiths, T.L.**, Cohen, J. D., & Webb, T. W. (2024). Understanding the limits of vision language models through the lens of the binding problem. *Advances in Neural Information Processing Systems, 38.*

421. Chen, A., Sucholutsky, I., Russakovsky, O., & **Griffiths, T.L.** (2024). Analyzing the Roles of Language and Vision in Learning from Limited Data. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

422. Correa, C. G., **Griffiths, T.L.**, & Daw, N. D. (2024). Program-Based Strategy Induction for Reinforcement Learning. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

423. Harootonian, S. K., Niv, Y., **Griffiths, T.L.**, & Ho, M. K. (2024). Modeling Cognitive Strategies in Teaching: Integrating Theory of Mind and Heuristics. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

424. Kumar, S., Marjieh, R., Zhang, B., Campbell, D., Hu, M. Y., Bhatt, U., Lake, B. M., & **Griffiths, T.L.** (2024). Comparing Abstraction in Humans and Large Language Models Using Multimodal Serial Reproduction. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

425. Liu, R., Sumers, T. R., Dasgupta, I., & **Griffiths, T.L.** (2024). How do Large Language Models Navigate Conflicts between Honesty and Helpfulness? *Proceedings of the 41st International Conference on Machine Learning (ICML).*

426. Malaviya, M., Sucholutsky, I., & **Griffiths, T.L.** (2024). Pushing the Limits of Learning from Limited Data. *Proceedings of the AAAI Symposium Series, 3*(1), 559561.

427. Mancoridis, M., Sumers, T., & **Griffiths, T.L.** (2024). Publish or Perish: Simulating the Impact of Publication Policies on Science. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

428. Marinescu, I. R., McCoy, R. T., & **Griffiths, T.L.** (2024). Distilling Symbolic Priors for Concept Learning into Neural Networks. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

429. Marjieh, R., Gokhale, A., Bullo, F., & **Griffiths, T.L.** (2024). Task Allocation in Teams as a Multi-Armed Bandit. *Proceedings of Collective Intelligence 2024.*

430. Marjieh, R., van Rijn, P., Sucholutsky, I., Lee, H., **Griffiths, T.L.**, & Jacoby, N. (2024). A Rational Analysis of the Speech-to-Song Illusion. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

431. Mieczkowski, E., Turner, C. R., Vlez, N., & **Griffiths, T.L.** (2024). Many Hands Don't Always Make Light Work: Explaining Social Loafing via Multiprocessing Efficiency. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

432. Niedermann, J. P., Sucholutsky, I., Marjieh, R., elen, E., **Griffiths, T.L.**, Jacoby, N., & van Rijn, P. (2024). Studying the Effect of Globalization on Color Perception using Multilingual Online Recruitment and Large Language Models. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

433. Oktar, K., Sumers, T., & **Griffiths, T.L.** (2024). A Rational Model of Vigilance in Motivated Communication. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

434. Peng, A., Bobu, A., Li, B. Z., Sumers, T. R., Sucholutsky, I., Kumar, N., & **Griffiths, T.L.** (2024). Preference-Conditioned Language-Guided Abstraction. *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction.*

435. Prabhakar, A., **Griffiths, T.L.**, & McCoy, R. T. (2024). Deciphering the factors influencing the efficacy of chain-of-thought: Probability, memorization, and noisy reasoning. *Findings of the Association for Computational Linguistics: EMNLP.*

436. Rane, S., Ho, M., Sucholutsky, I., & **Griffiths, T.L.** (2024). Concept Alignment as a Prerequisite for Value Alignment. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

437. Rane, S., Ku, A., Baldridge, J., Tenney, I., **Griffiths, T.L.**, & Kim, B. (2024). Can Generative Multimodal Models Count to Ten? *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

438. Russek, E. M., Turner, C. R., McEwen, E., Miscov, A. M., Seed, A., & **Griffiths, T.L.** (2024). Modeling the Contributions of Capacity and Control to Working Memory Development. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

439. Snell, J. C., Bencomo, G. M., & **Griffiths, T.L.** (2024). A metalearned neural circuit for nonparametric Bayesian inference. *Advances in Neural Information Processing Systems 38.*

440. Sucholutsky, I., & **Griffiths, T.L.** (2024). Why should we care if machines learn human-like representations? *AAAI-24 Spring Symposium on Human-Like Learning.*

441. Sucholutsky, I., Zhao, B., & **Griffiths, T.L.** (2024). Using Compositionality to Learn Many Categories from Few Examples. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

442. Tian, Y., Ravichander, A., Qin, L., Bras, R. L., Marjieh, R., Peng, N., Choi, Y., **Griffiths, T.L.**, & Brahman, F. (2024). MacGyver: Are large language models creative problem solvers? *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics.*

443. Urano, Y., Marjieh, R., **Griffiths, T.L.**, & Jacoby, N. (2024). The Influence of Social Information and Presentation Interface on Aesthetic Evaluations. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

444. Wynn, A. H., Sucholutsky, I., & **Griffiths, T.L.** (2024). Learning human-like representations to enable

learning human values. *Advances in Neural Information Processing Systems 38.*

445. Zhang, L., Nelson, L., & **Griffiths, T.L.** (2024). Analyzing the Benefits of Prototypes for Semi-Supervised Category Learning. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

446. Zhao, B., Vlez, N., & **Griffiths, T.L.** (2024). A Rational Model of Innovation by Recombination. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

447. Zhu, J. Q., & **Griffiths, T.L.** (2024). Incoherent Probability Judgments in Large Language Models. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

448. Zhu, J. Q., Yan, H., & **Griffiths, T.L.** (2024). Recovering Mental Representations from Large Language Models with Markov Chain Monte Carlo. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society.*

449. Arumugam, D., & **Griffiths, T.L.** (2025). On Temporal Credit Assignment and Data-Efficient Reinforcement Learning. *Finding the Frame Workshop at RLC.*

450. Bencomo, G., Gupta, M., Marinescu, I., McCoy, R. T., & **Griffiths, T.L.** (2025). Teasing Apart Architecture and Initial Weights as Sources of Inductive Bias in Neural Networks. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society.*

451. Collins, K. M, Todd, G., Zhang, C. E, Weller, A., Togelius, J., Chu, J., Wong, L., **Griffiths, T.L.**, & Tenenbaum, J. B. (2025). Generation and Evaluation in the Human Invention Process through the Lens of Game Design. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society.*

452. Gupta, M., Rane, S., McCoy, R. T., & **Griffiths, T.L.** (2025). Convolutional neural networks can (meta-) learn the same-different relation. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society.*

453. Kuperwajs, I., Russek, E., Schut, L., Sagiv, Y., Mattar, M. G., Ma, W. J., & **Griffiths, T.** (2025). Exploring resource-rational planning under time pressure in online chess. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society.*

454. Liu, R., Geng, J., Peterson, J. C., Sucholutsky, I., & **Griffiths, T.L.** (2025). Large language models assume people are more rational than we really are. *Proceedings of the 13th International Conference on Learning Representations (ICLR).*

455. Liu, R., Geng, J., Wu, A. J., Sucholutsky, I., Lombrozo, T., & **Griffiths, T.L.** (2025). Mind your step (by step): Chain-of-thought can reduce performance on tasks where thinking makes humans worse. *Proceedings of the 42nd International Conference on Machine Learning (ICML).*

456. Marjieh, R., Anglada-Tort, M., **Griffiths, T.L.**, & Jacoby, N. (2025). Characterizing the interaction of cultural evolution mechanisms in experimental social networks. *Proceedings of the 47th Annual Conference of the Cognitive Science Society.*

457. Mon-Williams, R., Taylor-Davies, M., Mieczkowski, E., Velez, N., Bramley, N. R., Wang, Y., **Griffiths, T.L.**, & Lucas, C. G. (2025). Partner modelling emerges in recurrent agents (but only when it matters). *Advances in Neural Information Processing Systems, 39.*

458. Nam, A., Conklin, H., Yang, Y., **Griffiths, T.L.**, Cohen, J., & Leslie, S.-J. (2025). Causal head gating: A framework for interpreting roles of attention heads in transformers. *Advances in Neural Information Processing Systems 39.*

459. Snell, J. C., & **Griffiths, T.L.** (2025). Conformal prediction as Bayesian quadrature. *Proceedings of the 42nd International Conference on Machine Learning (ICML).*

460. Sucholutsky, I., Collins, K. M., Malaviya, M., Jacoby, N., Liu, W., Sumers, T. R., Korakakis, M., Bhatt, U., Ho, M., Tenenbaum, J. B., Love, B., Pardos, Z. A., Weller, A., & **Griffiths, T.L.** (2025). Representational alignment supports effective machine teaching. *ICLR 2025 Workshop on Bidirectional Human-AI Alignment.*

461. Turner, C. R., Arumugam, D., Nelson, L., & **Griffiths, T.L.** (2025). Trade-offs between tasks induced by capacity constraints bound the scope of intelligence. *Proceedings of the 47th Annual Meeting of the Cognitive*

Science Society.

462. Veselovsky, V., Stroebl, B., Bencomo, G., Arumugam, D., Schut, L., Narayanan, A., & **Griffiths, T.L.** (2025). Hindsight Merging: Diverse Data Generation with Language Models. *Proceedings of the 41st Conference on Uncertainty in Artificial Intelligence.*

463. Wu, A. J., Liu, R., Oktar, K., Sumers, T. R., & **Griffiths, T.L.** (2025). Are Large Language Models Sensitive to the Motives Behind Communication? *Advances in Neural Information Processing Systems 39.*

464. Yamakoshi, T., **Griffiths, T.L.**, McCoy, R. T., & Hawkins, R. D. (2025). Evaluating distillation methods for data-efficient syntax learning. *Findings of the Association for Computational Linguistics: EMNLP 2025.*

465. Zhang, L., Snell, J. C., & **Griffiths, T.L.** (2025). Amoritzed Bayesian Meta-Learning for Low-Rank Adaptation of Large Language Models. *Proceedings of the 2nd Workshop on Uncertainty-Aware NLP.*

466. Zhao, B., Mieczkowski, E., Arumugam, D., Velez, N., & **Griffiths, T.L.** (2025). Discovering Hidden Laws in Innovation by Recombination. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society.*

467. Zuo, Y., Kayan, K., Wang, M., Jeon, K., Deng, J., & **Griffiths, T.L.** (2025). Towards Foundation Models for 3D Vision: How Close are We? *International Conference on 3D Vision (3DV).*

Book chapters

468. Steyvers, M., & **Griffiths, T.L.** (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis.* Hillsdale, NJ: Erlbaum.

469. Tenenbaum, J.B., **Griffiths, T.L.**, & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. In Gopnik, A., & Schulz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation.* Oxford: Oxford University Press.

470. **Griffiths, T.L.**, & Tenenbaum, J.B. (2007). Two proposals for causal grammars. In Gopnik, A., & Schulz, L. (Eds.), *Causal learning: Psychology, philosophy, and computation.* Oxford: Oxford University Press.

471. Ghahramani, Z., **Griffiths, T.L.**, & Sollich, P. (2007). Bayesian nonparametric latent feature models (with discussion and rejoinder). In Bernardo, J. M., Bayarri, M. J, Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (Eds.) *Bayesian statistics 8.* Oxford: Oxford University Press.

472. **Griffiths, T.L.**, Sanborn, A. N., Canini, K. R., & Navarro, D. J. (2008). Categorization as nonparametric Bayesian density estimation. To appear in M. Oaksford and N. Chater (Eds.). *The probabilistic mind: Prospects for rational models of cognition.* Oxford: Oxford University Press.

473. Goodman, N. D., Tenenbaum, J. B., **Griffiths, T.L.**, & Feldman, J. (2008). Compositionality in rational analysis: Grammar-based induction for concept learning. To appear in M. Oaksford and N. Chater (Eds.). *The probabilistic mind: Prospects for rational models of cognition.* Oxford: Oxford University Press.

474. Steyvers, M., & **Griffiths, T.L.** (2008). Rational analysis as a link between human memory and information retrieval. To appear in M. Oaksford and N. Chater (Eds.). *The probabilistic mind: Prospects for rational models of cognition.* Oxford: Oxford University Press.

475. **Griffiths, T.L.**, & Yuille, A. (2008). A primer on probabilistic inference. To appear in M. Oaksford and N. Chater (Eds.).*The probabilistic mind: Prospects for rational models of cognition.* Oxford: Oxford University Press.

476. **Griffiths, T.L.**, Kemp, C., & Tenenbaum, J.B. (2008). Bayesian models of cognition. In R. Sun (ed.), *Cambridge handbook of computational psychology.* Cambridge, UK: Cambridge University Press.

477. Jaeger, H., Baronchelli, A., Briscoe, T., Christiansen, M. H., **Griffiths, T.**, Jäger, G., Kirby, S., Komarova, N. L., Richerson, P. J., Steels, L., & Triesch, J. (2009). What can mathematical, computational and robotic models tell us about the origins of syntax? In D. Bickerton & E. Szathmáry (Eds.) *Biological foundations and origins of syntax.* Cambridge, MA: MIT Press.

478. **Griffiths, T.L.** (2010). Bayesian models as tools for exploring inductive biases. In M. Banich & D. Caccamise (Eds.) *Generalization of knowledge: Multidisciplinary perspectives.* New York: Psychology Press.

479. **Griffiths, T.L.,** Sanborn, A.N., Canini, K.R., Navarro, D.J., & Tenenbaum, J.B. (2011). Nonparametric Bayesian models of category learning. In E. M. Pothos & A. J. Wills (Eds.) *Formal approaches in categorization.* Cambridge, UK: Cambridge University Press.

480. Austerweil, J.L., & **Griffiths, T.L.** (2012). Human feature learning. In N.M. Seel (Ed.) *Encyclopedia of the Sciences of Learning.* New York: Springer.

481. **Griffiths, T.L.,** Tenenbaum, J.B., & Kemp, C. (2012). Bayesian inference. In *Oxford Handbook of Thinking and Reasoning.* Oxford: Oxford University Press.

482. Bonawitz, E., Gopnik, A., Denison, S., & **Griffiths, T.L.** (2012). Rational randomness: The role of sampling in an algorithmic account of preschoolers' causal learning. In F. Xu (Ed.) *Rational constructivism in cognitive development.* Waltham, MA: Academic Press.

483. Bugnyar, T., Boyd, R., Bossan, B., Gächter, S., **Griffiths, T.**, Hammerstein, P., Jensen, K., Mussweiler, T., Nagel, R., & Warneken, F. (2012). Evolutionary perspectives on social cognition. In *Evolving the Mechanisms of Decision Making: Toward a Darwinian Decision Theory.* Cambridge, MA: MIT Press.

484. Sanborn, A.N., & **Griffiths, T.L.** (2015). Exploring the structure of mental representations by implementing computer algorithms with people. In Raaijmakers, J.G.W., Criss, A.H., Goldstone, R. L., Nosofsky, R. M., & Steyvers, M. (2015). (Eds.). *Cognitive Modeling in Perception and Memory: A Festschrift for Richard M. Shiffrin.* New York: Psychology Press.

485. Austerweil, J.L., Gershman, S.J., Tenenbaum, J.B., & **Griffiths, T.L.** (2015). Structure and flexibility in Bayesian models of cognition. In J.R. Busemeyer, J.T. Townsend, Z. Wang, & A. Eidels, Eds, *Oxford Handbook of Computational and Mathematical Psychology.* Oxford: Oxford University Press.

486. **Griffiths, T.L.** (2017). Formalizing prior knowledge in causal induction. In Waldmann (Ed.) *Oxford handbook of causal reasoning.* Oxford: Oxford University Press.

487. Tenenbaum, J. B., **Griffiths, T.L.**, & Chater, N. (2024). Introducing the Bayesian approach to cognitive science. In T.L. Griffiths, N. Chater, & J. B. Tenenbaum, (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

488. Chater, N., **Griffiths, T.L.**, & Tenenbaum, J. B. (2024). Probabilistic models of cognition in historical context. In T.L. Griffiths, N. Chater, & J. B. Tenenbaum, (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

489. **Griffiths, T.L.**, & Tenenbaum, J. B. (2024). Bayesian inference. In T.L. Griffiths, N. Chater, & J. B. Tenenbaum, (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

490. **Griffiths, T.L.**, & Yuille, A. (2024). Graphical models. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

491. **Griffiths, T.L.**, & Yuille, A. (2024). Building complex generative models. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

492. **Griffiths, T.L.**, & Sanborn, A. N. (2024). Approximate probabilistic inference. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

493. Chater, N., **Griffiths, T.L.**, & Ho, M. K. (2024). From probabilities to actions. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

494. Austerweil, J., Sanborn, A. N., Lucas, C., & **Griffiths, T.L.** (2024). Capturing the growth of knowledge with nonparametric Bayesian models. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

495. **Griffiths, T.L.**, Sanborn, A. N., Marjieh, R., Langlois, T., Xu, J., & Jacoby, N. (2024). Estimating

subjective probability distributions. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

496. **Griffiths, T.L.**, Vul, E., Sanborn, A. N., & Chater, N. (2024). Sampling as a bridge across levels of analysis. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

497. **Griffiths, T.L.**, Dasgupta, I., & Grant, E. (2024). Bayesian models and neural networks. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

498. Lieder, F., Callaway, F., & **Griffiths, T.L.** (2024). Resource-rational analysis. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

499. Kemp, C., Goodman, N. D., & **Griffiths, T.L.** (2024). Bayesian inference over logical representations. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

500. Chater, N., **Griffiths, T.L.**, & Tenenbaum, J. B. (2024). A Bayesian conversation. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian Models of Cognition: Reverse Engineering the Mind.* MIT Press.

Technical reports, invited articles, and other unreviewed publications

501. Tenenbaum, J.B., & **Griffiths, T.L.** (2001). Some specifics about generalization. *Behavioral and Brain Sciences, 24*, 772-778. (response to commentaries)

502. Kemp, C., **Griffiths, T.L.**, & Tenenbaum, J.B. (2004). *Discovering latent classes in relational data.* AI Memo 2004-019, Massachusetts Institute of Technology.

503. **Griffiths, T.L.**, & Ghahramani, Z. (2005). *Infinite latent feature models and the Indian buffet process.* Gatsby Technical Report 2005-001, Gatsby Computational Neuroscience Unit, University College London.

504. **Griffiths, T.L.**, & Yuille, A. (2006). A primer on probabilistic inference. *Trends in Cognitive Sciences.* Supplement to special issue on Probabilistic Models of Cognition (volume 10, issue 7).

505. **Griffiths, T.L.**, & Tenenbaum, J.B. (2006). Statistics and the Bayesian mind. *Significance, 3*, 130-133. (invited paper)

506. Smith, K., Kalish, M.L., **Griffiths, T.L.**, & Lewandowsky, S. (2008). Cultural transmission and the evolution of human behaviour: Introduction to the issue. *Philosophical Transactions of the Royal Society, 363,* 3469-3476.

507. **Griffiths, T.L.** (2009). The strengths of – and some of the challenges for – Bayesian models of cognition. *Behavioral and Brain Sciences.* (commentary)

508. **Griffiths, T.L.** (2009). Connecting human and machine learning via probabilistic models of cognition. *InterSpeech 2009.* (invited paper)

509. **Griffiths, T.L.** (2011). Rethinking language: How probabilities shape the words we use. *Proceedings of the National Academy of Sciences, 108,* 3825-3826. (invited commentary)

510. **Griffiths, T.L.**, & Reali F. (2011). Modeling minds as well as populations. *Proceedings of the Royal Society, Series B.* (response to commentary)

511. Xu, F., & **Griffiths, T.L.** (2011). Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition, 120,* 299-301. (introduction to special issue)

512. Chater, N., Goodman, N.D., **Griffiths, T.L.**, Kemp, C., Oaksford, M., & Tenenbaum, J.B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain*

Sciences, 34, 194-196. (commentary)

513. **Griffiths, T.L.**, Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are). *Psychological Bulletin, 138,* 415-422. (comment)

514. Jia, Y., Abbott, J., Austerweil, J., **Griffiths, T.**, & Darrell, T. (2012). *Visually-grounded Bayesian word learning.* Technical Report UCB/EECS-2012-202, EECS Department, University of California, Berkeley.

515. **Griffiths, T.L.** (2013). Bayesian approaches to color category learning. *Encyclopedia of Color Science and Technology.* New York: Springer.

516. Goodman, N.D., Frank, M.C., **Griffiths, T.L.**, Tenenbaum, J.B., Battaglia, P., & Hamrick, J. (2015). Relevant and robust. A response to Marcus and Davis. *Psychological Science, 26,* 539-541.

517. **Griffiths, T.L.** (2015). Manifesto for a new (computational) cognitive revolution. *Cognition, 135,* 21-23. (invited paper)

518. Lieder, F., & **Griffiths, T.L.** (2020). Advancing rational analysis to the algorithmic level. *Behavioral and Brain Sciences, 43,* e27. (response to commentaries)

519. Turner, C. R., Morgan, T. J. H., & **Griffiths, T.L.** (2025). Complex brains allow functioning in a complex environment by using information. *Behavior and Brain Sciences, 48*, e96.

INVITED TALKS

2025 Toyota Research and Development Labs, Nagoya, Japan.

Centre for Cognition, Computation and Modelling, Birkbeck, University of London, London, UK.

Symposium on Uncertainty, University of Zurich, Zurich, Switzerland.

Invited symposium, Association for Psychological Science, Washington, DC.

Distinguished Lecture in Neuroscience, Carnegie Mellon University, Pittsburgh, PA.

Societal Decision-Making Institute, Carnegie Mellon University, Pittsburgh, PA.

Data Science Distinguished Speakers Series, University of Chicago, Chicago, IL.

Behavioral Economics Seminar, Booth School of Business, University of Chicago, Chicago, IL.

Invited Symposium, Annual Meeting of the Society for Philosophy and Psychology, Ithaca, NY.

ML Collective, Palo Alto, CA.

Anthropic, San Francisco, CA.

Distinguished Lecture in Data Science, Stanford University, Stanford, CA.

Data Science and AI Institute Spring Symposium, Johns Hopkins University, Baltimore, MD.

Bridge Program on Collaborative AI and Modeling Humans, AAAI, Philadelphia, PA. (keynote)

Bridge Program on Bridging Cognitive Science and AI to Bridge Neuro and Symbolic AI, AAAI, Philadelphia, PA.

Center for Data Science, New York University, New York, NY.

Conference on Language Models, Montreal, Canada. (keynote)

The Next Turing Test, University of Cambridge, Cambridge, UK.

AI and Evolutionary Reasoning Workshop, Arizona State University, Tempe, AZ.

Whitehead Lecture Series in Cognition, Computation, and Culture, Goldsmiths, University of London, London, UK.

2024 Workshop on Behavioral Machine Learning, NeurIPS, Vancouver, BC. (keynote)

Workshop on Language Gamification, NeurIPS, Vancouver, BC. (keynote)

Empirical Methods in Natural Language Processing, Miami, FL. (keynote)

Annual meeting of the Artficial and Natural Intelligence Institute, Columbia University, NY. (keynote)

Workshop on Naturalistic Approaches to Artificial Intelligence, Institute for Pure and Applied Mathematics, University of California, Los Angeles, CA.

Brain and Cognitive Science Colloquium, Massachusetts Institute of Technology, Cambridge, MA.

Kempner Institute Colloquium, Harvard University, Cambridge, MA.

Google DeepMind, London, UK.

Workshop on Neuroscience and Artificial Intelligence, Norway.

Conference on Statistical Learning, San Sebastian, Spain. (keynote)

Department of Psychology, University of Arizona, Tucson, AZ.

2023 Columbia Seminar on Cognitive/Behavioral Neuroscience, Columbia University, New York, NY.

Center for Mind, Brain, and Culture Colloquium, Emory University, Atlanta, GA.

Data Science and AI Seminar, University of Georgia, Athens, GA.

Decision Experience and Behavior (DEB) Seminar, University of Haifa, Haifa, Israel.

Cognitive Science Colloquium, Departement d'Etudes Cognitives, École Normale Supérieure, Paris, France.

Neuroeconomics Seminar, University of Zurich, Switzerland.

Chen Institute Symposium, Caltech, Los Angeles, CA.

AI Center for Research, New Jersey Institute of Technology, Newark, NJ.

Humanizing the Sustainable Smart City Workshop, Stockholm, Sweden.

AI, Cognition, and the Economy Workshop, Microsoft Research, New York, NY.

ML in NYC talk, Flatiron Institute, New York, NY.

33rd Advanced School in Economic Theory, Hebrew University Jerusalem, Israel.

Engineering colloquium, Cambridge University, Cambridge UK.

UniReps workshop, Neural Information Processing Systems conference, New Orleans, LA.

2022 Panelist, Methods and Measurement, First Global Scientific Conference on Human Flourishing.

Association for Cognitive Science Conference, Delhi, India. (keynote)

Precision-Convergence Webinar, McGill Centre for the Convergence of Health and Economics (MC-CHE), Montreal, Canada.

Conference on Digital Experimentation (CODE), Massachusetts Institute of Technology, Cambridge, MA. (keynote)

Joint Seminar, Harvard Business School and Department of Economics, Cambridge, MA.

Society for Neuroeconomics, Crystal City, VA. (keynote)

Edinburgh Lectures on Language, University of Edinburgh.

International Conference on Computational Social Science (IC2S2), Chicago, IL. (keynote)

Mind and Machine Seminar, Bristol University.

Tech talk, Facebook AI Research.

2021 Workshop on "I can't believe it's not better," Neural Information Processing Systems conference.

Workshop on "Ecological reinforcement learning," Neural Information Processing Systems conference.

Colloquium, Computer Science Department, University of Rochester.

Computational Psychiatry group, University College London.

Workshop on "Computational Cognition," University of Osnabruck.

Crowder lecture, Department of Psychology, Yale University.

Eastern European Machine Learning Summer School.

Toyota Research Institute Machine Assisted Cognition group.

Tech Talk, Facebook Reality Labs.

Center for Human-Compatible Artificial Intelligence, University of California, Berkeley, CA

Rutgers Perceptual and Cognitive Science Forum (keynote), Rutgers University, New Brunswick, NJ.

Workshop on "Surprise, Curiosity, and Reward" École Polytechnique Fédérale de Lausanne.

Cognitive psychology talk series, University of California, Los Angeles, CA.

2020 Workshop on "The Future of Linguistics," Max Planck Institute, Nijmegen, the Netherlands.

Human Machine Intelligence talk series, University of Virginia, VA.

Colloquium, Department of Psychology, University of British Columbia, Vancouver, Canada.

Decision-making and Reinforcement Learning talk series, Max Planck Institute, Tübingen, Germany.

AI4All summer school, Princeton, NJ.

Workshop on "Cognitive Effort," Annual Conference of the Cognitive Science Society.

Diverse Intelligences Summer Institute, St. Andrews, Scotland.

Workshop on "Human in the Loop Learning," International Conference on Machine Learning.

Workshop on "Machine Learning, Theory, and Method in the Social Sciences," Institute for Advanced Study, Princeton, NJ. (public lecture)

Workshop on "Minds and Machines," University of California, Santa Barbara, Santa Barbara, CA.

Workshop on "Virtual Labs," University of Pennsylvania, Philadelphia, PA.

2019 Workshop on "Learning from Rich Experience," Neural Information Processing Systems conference, Vancouver, Canada.

Decision-making colloquium, Wharton Business School, University of Pennsylvania, Philadelphia, PA.

Annual meeting of the Society for Judgment and Decision-making (JDM), Montreal, Canada. (keynote)

Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland.

Workshop on "Heuristics, Hacks, and Habits," Annual Conference of the Cognitive Science Society, Montreal, Canada.

Amazon TechTalk, Seattle, WA.

Reinforcement Learning and Decision-Making, Montreal, Canada. (keynote)

Workshop on "AI and Cognitive Development," Facebook Artificial Intelligence Research, New York, NY.

Janet Taylor Spence Symposium, Association for Psychological Science, Washington, DC.

Computational Social Science Colloquium, University of Chicago, Chicago, IL.

2018 Microsoft Research, New York, NY.

Decision and Cognition talk series, Columbia University, New York, NY.

Gartner Symposium, Gold Coast, Australia. (keynote)

Hong Kong AI Summit, Hong Kong, China. (keynote)

Research talk, Microsoft Research New York, New York, NY.

Cognition and Decision-Making Colloquium, Columbia University, New York, NY.

Cognition-Perception Colloquium, New York University, New York, NY.

Amazon Research Scientist Summit, Semiahmoo, WA. (keynote)

Workshop on Exploration and Exploitation, Annual Conference of the Cognitive Science Society, Madison, WI.

"Beyond deep learning" workshop, Brown University, Providence, RI.

Roger N. Shepard Visiting Scholar, University of Arizona, Tucson, AZ.

Sloan-Nomis workshop on "Attention and decision-making", New York, NY.

Affective Brain lab meeting, University College London, London, UK.

Sensors group, Uber, San Francisco, CA.

2017 Workshop on Cognitively-Inspired Artificial Intelligence, Neural Information Processing Systems conference, Long Beach, CA.

Sloan-Nomis workshop on "Attention and decision-making".

Debate on "Big data and the mind," Northwestern University, Evanston, IL.

Cognitive Computational Neuroscience conference, New York, NY. (keynote)

TEDx Sydney, Sydney, Australia.

Institute of Neuroinformatics, Universität Zürich, Switzerland.

Simons Institute workshop on "Representation learning", Berkeley, CA.

2016 The Commonwealth Club, San Francisco, CA.

Bay Area ACM chapter, Menlo Park, CA.

OpenAI, San Francisco, CA.

Stripe, San Francisco, CA.

Facebook, Menlo Park, CA.

Rotman School of Management, University of Toronto.

Google Book Talks, Mountain View, CA.

The Commonwealth Club of Silicon Valley, Santa Clara, CA.

California Institute of Integral Studies, San Francisco, CA.

Cloudera, San Francisco, CA.

Department of Psychology, Carnegie Mellon University.

Mind lecture, University of Kansas.

Center for Cognitive Science, University of Minnesota.

2015 Center for Statistics and Machine Learning, Princeton University.

Brain day, University of Waterloo, Canada.

Psychology colloquium, University of Pennsylvania.

Teuber lecture, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Organizational Behavior group, Stanford Graduate School of Business.

Cognitive Science Keynote, Yale University.

2014 Decision Making Conference, Bristol, UK. (keynote)

Working group on collective cognition, Santa Fe Institute, Santa Fe, NM.

Translational Neuroscience Unit, ETH Zürich, Switzerland.

Cognitive Science colloquium, Central European University, Budapest, Hungary.

Institute of Neuroinformatics, Universität Zürich, Switzerland.

IARPA workshop on "Cognitive Science 2.0," Fort Meade, MD.

2013 Mind, Brain, and Computation colloquium, Stanford University.

Department of Statistics, Duke University.

Distinguished Speakers in Cognitive Science Lecture Series, Michigan State University.

Departmental colloquium, Department of Psychology, Princeton University.

Workshop on Integrating Approaches to Computational Cognition, National Science Foundation.

Sage Junior Fellows Workshop, University of California, Santa Barbara.

Computational Social Sciences colloquium, University of Massachusetts, Amherst.

2012 Plenary symposium on "30 Years of Marr's Levels of Analysis," Annual Conference of the Cognitive Science Society.

Evolution of Language conference, Kyoto, Japan. (keynote)

Computational and Systems Neuroscience (CoSyNe) conference, Salt Lake City, Utah. (keynote)

Debate on Bayesian models of Cognition, Cognitive Science Program, Northwestern University.

Cognition and Language Group, Stanford University.

2011 Researching Communication Conference, University of Western Sydney, Sydney, Australia. (keynote)

Department of Linguistics, University of Maryland.

Department of Psychology, Cornell University.

International Conference on Artificial Neural Networks, Helsinki, Finland. (keynote)

Stanford Psychology of Language Talk, Stanford University.

Swartz Institute for Theoretical Neuroscience, Yale University.

California Cognitive Science Students Conference, University of California, Berkeley (keynote).

Cognitive Science Department, University of California, San Diego.

Mind, Brain, and Computation Colloquium, Stanford University.

2010 Symposium on "The cognition and language of color," Optical Society of America Fall Vision Meeting, Rochester, NY.

Workshop on "Computational models of the mind," Stanford University, Stanford, CA.

Annual Meeting of the Society for Mathematical Psychology, Portland, OR.

Workshop on "Language as an evolutionary system," University of Edinburgh.

School of Informatics, University of Edinburgh.

Society of Experimental Psychologists, Philadelphia, PA.

Department of Anthropology, University of California, Los Angeles, Los Angeles, CA.

Cognitive Science Center, University of Minnesota, Minneapolis, MN.

Human, Social, Culture, and Behavior Modeling group, Naval Postgraduate School, Monterey, CA.

Workshop on "The sampling hypothesis," Computational and Systems Neuroscience (CoSyNe) conference, Park City, UT.

School of Psychology, University of Western Australia.

Institute for Research in Cognitive Science, University of Pennsylvania.

2009 Workshop on "Nonparametric Bayes," Neural Information Processing Systems conference, Vancouver, BC.

InterSpeech 2009, Brighton, UK. (keynote)

Workshop on "Pedagogical reasoning," 31st Annual Conference of the Cognitive Science Society, Amsterdam, Netherlands.

Child Language Research Forum, Linguistic Society of America Summer Institute, Berkeley, CA. (keynote)

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA.

Workshop on "Human and machine learning," Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA.

2008 Empirical Methods in Natural Language Processing conference, Honolulu, HI. (keynote)

Workshop on "The cognitive science of induction and confirmation," Venice, Italy.

International Meeting of the Psychometric Society, Durham, NH. (keynote)

Quantitative talk series, Psychology Department, University of California, Davis, Davis, CA.

Symposium on "Bayesian models of perception," Annual Meeting of the Vision Sciences Society, Naples, FL.

Workshop on "Language and Cognition," University of Chicago, Chicago, IL.

Cognitive Science Department, University of Arizona, Tucson, AZ.

Workshop on "Core cognitive developmental mechanisms of understanding social causation and the establishment of conceptual representations of causal and intentional agency and action," Center for Advanced Study in the Behavioral Sciences, Stanford University, Palo Alto, CA.

Computer Science Department, University of Utah, Salt Lake City, UT.

Computer Science Department, Brigham Young University, Provo, UT.

Cognitive Science Department, University of California, Merced, Merced, CA.

Workshop on "Evolution of psychological categories," Institute for Mathematical Behavioral Sciences, University of California, Irvine, CA.

Workshop on "Spanning the Socio-Cognitive Modeling Gap: From Development to Social Simulation," Massachusetts Institute of Technology, Cambridge, MA.

2007 Distinguished Speaker Series, Center for Machine Learning and Intelligent Systems, University of California, Irvine, Irvine, CA.

Cowles symposium, Cowles Foundation for Research in Economics, Yale University, New Haven, CN.

Natural language processing group, Microsoft Research, Redmond, WA.

Psychology Department, University of California, San Diego, La Jolla, CA.

Center for the Study of Language and Information, Stanford, CA.

Psychology Department, Stanford University, Stanford, CA.

Workshop on "Normative models in neuroscience" Computational and Systems Neuroscience (CoSyNe) conference, Park City, UT.

2006 Department of Psychology, University of California, Los Angeles, Los Angeles, CA.

Department of Statistics, University of California, Los Angeles, Los Angeles, CA.

Bayes focus week, Statistics and Mathematical Sciences Institute, Research Triangle Park, NC.

NeuroCritical Care Conference, Baltimore, MD. (keynote)

Center for Mind, Brain, and Computation, Stanford University, Stanford, CA.

AI group, SRI, Palo Alto, CA.

NSF Science of Learning Center conference on "Generalization of knowledge," University of Colorado, Boulder, CO.

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY.

Department of Psychology, Yale University, New Haven, CN.

2005 Workshop on "Bayesian natural language processing" at the Neural Information Processing Systems conference, Whistler, BC.

Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA.

"Empirical inference" symposium, Max Planck Institute for Biological Cybernetics, Tubingen, Germany.

Language Evolution and Computation Research Unit, Edinburgh University, Edinburgh, Scotland.

Brain Sciences Program, Brown University, Providence, RI.

2004 Institute of Cognitive and Brain Sciences seminar, UC Berkeley, Berkeley, CA.

"Hot topics" workshop on Visualization and Analysis of High Dimensional Data, Mathematical Sciences Research Institute, Berkeley, CA.

Department of Psychology, Harvard University, Cambridge, MA.

Gatsby Computational Neuroscience Unit, University College London.

Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI.

2003 Computer Science Department, University of California, Berkeley, CA.

Psychology Department, University of California, Irvine, CA.

Sackler Colloquium on "Mapping knowledge domains," National Academy of Sciences, Irvine, CA.

NTT Communication Sciences Laboratory, Osaka, Japan.

2002 Psychology Department, University of Colorado, Boulder, CO.

Psychology Department, Indiana University, Bloomington, IN.

Applied statistics workshop, Center for Behavioral Research in the Social Sciences, Harvard University, Cambridge, MA.

2001 Psychology Department, University of California, San Diego, CA.

OTHER TALKS AND CONFERENCE PRESENTATIONS

2025 Symposium on "The role of language in human and machine intelligence," Annual Conference of the Cognitive Science Society, San Francisco, CA.

2024 Science of Magic conference, Las Vegas, NV.

2023 Annual meeting of the Society of Experimental Psychologists, Pennsylvania, PA.

2021 International Conference on Thinking, Paris, France.

2019 Annual Conference of the Society for Philosophy and Psychology, San Diego, CA.

Annual Meeting of the Society of Experimental Psychologists, New Brunswick, NJ.

2016 Annual Meeting of the Psychonomic Society, Boston, MA.

2015 Symposium on generative and discriminative models, Annual Conference of the Cognitive Science Society, Pasadena, CA.

2014 Society of Experimental Psychologists, Los Angeles, CA.

Association for Psychological Science conference, San Francisco, CA.

Workshop on deep learning, Annual Conference of the Cognitive Science Society, Quebec City, Canada.

2013 Uncertainty in Artificial Intelligence conference, Seattle, WA.

2012 Symposium on "Psychonomics without experiments", Annual Meeting of the Psychonomic Society, Minneapolis, MN.

2011 Society for Philosophy and Psychology, Montreal, Canada.

2010 Annual Summer Interdisciplinary Conference, Bend, OR.

Australasian Mathematical Psychology Conference, Margaret River, Western Australia.

2009 31st Annual Conference of the Cognitive Science Society, Amsterdam, The Netherlands.

2008 Biennial Meeting of the Philosophy of Science Association, Pittsburgh, PA.

International Conference on Thinking, Venice, Italy.

Annual meeting of the Society for Mathematical Psychology, Washington, DC.

30th Annual Conference of the Cognitive Science Society, Washington, DC.

2007 Neural Information Processing Systems conference, Vancouver, BC.

Annual meeting of the Society for Mathematical Psychology, Irvine, CA.

29th Annual Conference of the Cognitive Science Society, Nashville, TN.

Cognitive Development Society, Santa Fe, NM.

Society for Research in Child Development, Boston, MA.

2006 Neural Information Processings Systems Conference, Vancouver, BC.

Annual meeting of the Psychonomic Society, Houston, TX.

Annual meeting of the Society for Mathematical Psychology, Vancouver, BC.

28th Annual Conference of the Cognitive Science Society, Vancouver, BC.

Eastern Psychological Association, Baltimore, MD.

2005 Neural Information Processing Systems conference, Vancouver, BC.

Annual meeting of the Society for Judgment and Decision-Making, Toronto, ON.

Annual meeting of the Psychonomic Society, Toronto, ON.

27th Annual Conference of the Cognitive Science Society, Stresa, Italy.

2004 Workshop on "Probabilistic models of categorization," Neural Information Processing Systems conference, Whistler, BC.

Neural Information Processing Systems conference, Vancouver, BC.

Annual meeting of the Psychonomic Society, Minneapolis, MN.

Annual Cape Cod conference on Monte Carlo Methods, Cambridge, MA.

Society for Philosophy and Psychology conference, Barcelona, Spain.

Annual Summer Interdisciplinary Conference, Cavalese, Italy.

26th Annual Conference of the Cognitive Science Society, Chicago, IL.

2003 Workshop on "Syntax, Semantics, and Statistics," Neural Information Processing Systems conference, Whistler, BC.

Neural Information Processing Systems conference, Vancouver, BC.

25th Annual Conference of the Cognitive Science Society, Boston, MA.

DIMACS workshop on "Complexity and inference," Rutgers University, Piscataway, NJ.

2002 Neural Information Processing Systems conference, Vancouver, BC.

24th Annual Conference of the Cognitive Science Society, Fairfax, VA.

2001 23rd Annual Conference of the Cognitive Science Society, Edinburgh, Scotland.

Neural Information Processing Systems conference, Denver, CO.

Workshop on "Causal learning and inference in humans and machines," Neural Information Processing Systems conference, Denver, CO.

2000 22nd Annual Conference of the Cognitive Science Society, Philadelphia, PA.

Neural Information Processing Systems conference, Denver, CO.

PROFESSIONAL ACTIVITIES

Society leadership

2023- Governing board member, Cognitive Science Society.

2023- Council member on behalf of the Cognitive Science Society, Federation of Associations in the Behavioral and Brain Sciences.

Editorial and reviewing

2023- Section editor, *Open Encyclopedia of Cognitive Science.*

2017- Editorial board member, *Open Mind.*

2010- Consulting editor, *Psychological Review.*

2009- Editorial board member, *Journal of Machine Learning Research.*

2009- Editorial board member, *Cognitive Science.*

2007- Program committee member, Annual Conference of the Cognitive Science Society.

2021- Senior area chair, Neural Information Processing Systems conference.

2013 Member of search committee for new editor, *Psychonomic Bulletin & Review.*

2010-2011 Guest editor for special issue of *Cognition* (with Fei Xu).

2008 Guest editor for special issue of *Philosophical Transactions of the Royal Society* (with Kenny Smith, Mike Kalish, and Steve Lewandowsky).

2006-2009 Consulting editor, *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

2005-2012 Member of tutorial organizing committee, Annual Conference of the Cognitive Science Society.

2005-2006 Area chair for cognitive science and graphical models, Neural Information Processing Systems conference.

Ad hoc reviewer and panelist for the National Science Foundation (multiple programs), ad hoc reviewer for equivalent organizations in the United Kingdom, Australia, European Union, and Canada. Ad hoc reviewer for *Cognitive Science*, *Cognitive Psychology*, *Psychonomic Bulletin and Review*, *Psychological Review*, *Journal of Machine Learning Research*, *Annals of Applied Statistics*, *Nature Human Behavior*, *Memory and Cognition*, *Neurocomputing*, *Psychological Science*, *Cognition*, *Advances in Applied Mathematics*, *Journal of Mathematical Psychology*, *Psychological Bulletin*, *PLoS Computational Biology*, *Behavioral and Brain Sciences*, *Proceedings of the Royal Society*, *Journal of the Royal Society: Interface*, *Journal of Experimental Psychology: General*, *Complexity*, *PLoS One*, *Acta Psychologica*, *European Journal of Cognitive Psychology*, *Journal of Memory and Language*, *IEEE Transactions on Audio, Speech, and Language Processing*, *Journal of Artificial Intelligence Research*, *Adaptive Behavior*, *Interaction Studies*, *Computational Linguistics*, *Language Learning*, *Trends in Cognitive Science*, *Proceedings of the National Academy of Sciences*, *Science*, *Nature*, the Annual Conference of the Cognitive Science society, the International Conference on Machine Learning, the Artificial Intelligence and Statistics conference, the International Joint Conference on Artificial Intelligence, the National Conference on Artificial Intelligence, the Uncertainty in Artificial Intelligence conference, the Annual Meeting of the Association for Computational Linguistics, the Empirical Methods in Natural Language Processing conference, and the Neural Information Processing Systems conference.

Workshop and symposium organization

2025 Co-organizer, workshop on "Meta-reasoning: Deciding which game to play, which problem to solve, and when to quit," Annual Conference of the Cognitive Science Society.

Co-organizer, workshop on "Reasoning Across Minds and Machines," Annual Conference of the Cognitive Science Society.

2024 Co-organizer, workshop on "Naturalistic Approaches to Artificial Intelligence," Institute for Pure and Applied Mathematics, University of California, Los Angeles.

Co-organizer, workshop on "From Symbols to Signals," The Royal Society, London, UK.

2023 Co-organizer, workshop on "Large Language Models meet Cognitive Science," Annual Meeting of the Cognitive Science Society.

2022 Co-organizer, workshop on "Shared Visual Representations in Human and Machine Intelligence," Neural Information Processing Systems conference.

2021 Co-organizer, workshop on "Shared Visual Representations in Human and Machine Intelligence," Neural Information Processing Systems conference.

Co-organizer, workshop on "Human and Machine Decisions," Neural Information Processing Systems conference.

2020 Co-organizer, workshop on "Scaling cognitive science," Annual Conference of the Cognitive Science Society.

2019 Co-organizer, workshop on "Shared Visual Representations in Human and Machine Intelligence," Neural Information Processing Systems conference.

Co-organizer, workshop on "Scaling cognitive science," Princeton University, Princeton NJ.

2015 Co-organizer, workshop on "Bounded optimality and metareasoning," Neural Information Processing Systems conference.

2010 Co-organizer, workshop on "Transfer learning by learning rich generative models," Neural Information Processing Systems conference.

2009 Co-organizer, workshop on "Bounded-rational analyses of human cognition," Neural Information Processing Systems conference.

Co-organizer, workshop on "Probabilistic models of cognitive development," Banff International Research Station, Banff, Canada.

2008 Co-organizer, workshop on "Connecting probabilistic models of cognition and neural networks," University of California, Berkeley, Berkeley, CA.

2007 Co-organizer, symposium on "Modern Monte Carlo methods," Meeting of the Society for Mathematical Psychology.

2003 Co-organizer, workshop on "Syntax, semantics, and statistics," Neural Information Processing Systems conference.

2001 Co-organizer, workshop on "Causal learning and inference in humans and machines," Neural Information Processing Systems conference.

External tutorials

2022 Sloan-Nomis Summer School on Cognitive Foundations of Economic Behavior, Luzern, Switzerland.

2018 Sloan-Nomis Summer School on Cognitive Foundations of Economic Behavior, Luzern, Switzerland.

2017 Co-organizer, Data on the Mind Summer School, University of California, Berkeley.

2011 Tutorials on Causality and Monte Carlo as part of the Graduate Summer School on probabilistic models of cogntion at the Institute for Pure and Applied Mathematical, University of California, Los Angeles.

2010 Co-presenter, tutorial on "Bayesian models of inductive learning," Annual Conference of the Cognitive Science Society.

Tutorials on Causality, Nonparametric Bayes, and Monte Carlo methods at Machine Learning and Cognitive Science Summer School.

Tutorial on "Bayesian models of cognition," Australasian Mathematical Psychology Conference.

2008 Co-presenter, tutorial on "Bayesian models of inductive learning," Annual Conference of the Cognitive Science Society.

2007 Tutorials on graphical models, Monte Carlo, categorization, causal induction, and assorted other topics as part of the Graduate Summer School on probabilistic models of cognition at the Institute for Pure and Applied Mathematics, University of California, Los Angeles.

2006 Co-presenter, tutorial on "Bayesian models of inductive learning," Annual Conference of the Cognitive Science Society.

2004 Co-presenter, tutorial on "Bayesian models of inductive learning," Annual Conference of the Cognitive Science Society.

Media coverage

Research mentioned in *The Economist*, *The Atlantic*, *New Scientist*, *The New York Times Magazine*, *San Jose Mercurcy News*, *Psychology Today*, *Slate*, and *Cosmopolitan*, and on National Public Radio, BBC Radio, Canadian Broadcasting Corporation Radio One, *Scientific American* podcast, and the television program *Criminal Minds*, as well as a variety of science blogs. *Algorithms to Live By* was a question on the game show *Jeopardy* and inspired a song by the Brooklyn art-rock collective Foyer Red.

Profiled in *IEEE Intelligent Systems Magazine* as one of the "AI Ten to Watch" and *American Psychologist* as recipient of Distinguished Scientific Award for Early Career Contribution to Psychology.

LAB ALUMNI

Graduate students

Sharon Goldwater (Reader, University of Edinburgh)
Adam Sanborn (Assistant Professor, University of Warwick)
Lei Shi (Associate, McKinsey & Company)
Chris Lucas (Reader, University of Edinburgh)
Naomi Feldman (Assistant Professor, University of Maryland)
Kevin Canini (software engineer, Google)
Jing Xu (Assistant Professor, University of Georgia)
Saiwing Yeung (faculty member, Beijing Institute of Technology)
Joseph Williams (Assistant Professor, University of Toronto)
Joe Austerweil (Associate Professor, University of Wisconsin)
Daphna Buchsbaum (Assistant Professor, Brown University)
Anna Rafferty (Associate Professor, Carleton College)
Joshua Abbott (software engineer, Adobe)

M Pacer (software engineer, Netflix)
Jessica Hamrick (researcher, DeepMind)
Stephan Meylan (postdoctoral researcher, University of California, Berkeley)
David Bourgin (software engineer, Adobe)
Falk Lieder (Assistant Professor, University of California, Los Angeles)
Thomas Langlois (postdoctoral researcher, Princeton)
Joshua Peterson (postdoctoral researcher, Princeton)
Vael Gates (postdoctoral researcher, Stanford University)
Rachel Jansen (postdoctoral researcher, NASA)
Erin Grant (postdoctoral fellow, University College London)
Michael Chang (researcher, OpenAI)
Fred Callaway (Assistant Professor, Dartmouth College)
Ruairidh Battleday (postdoctoral researcher, Harvard University)
Rachit Dubey (Assistant Professor, University of California, Los Angeles)
Mayank Agrawal (co-founder, RoundTable)
Carlos Correa (postdoctoral researcher, New York University)
Sreejan Kumar (postdoctoral researcher, New York University)
Ted Sumers (technical staff, Anthropic)
Xuechunzi Bai (Assistant Professor, University of Chicago)

Postdoctoral researchers

Florencia Reali (faculty member, Universidad de Los Andes)
Wolf Vanpaemel (faculty member, Katholieke Universiteit Leuven)
Anne Hsu (lecturer, Queen Mary, University of London)
Elizabeth Bonawitz (Associate Professor, Harvard University)
Luke Maurits (postdoctoral researcher, University of Auckland)
Tom Morgan (Assistant Professor, Arizona State University)
Alex Paxton (Assistant Professor, University of Connecticut)
Jordan Suchow (Assistant Professor, Stevens Institute of Technology)
Aida Nematzadeh (researcher, DeepMind)
Daniel Reichman (Assistant Professor, Worcester Polytechnic Institute)
Nori Jacoby (Assistant Professor, Cornell University)
Bill Thompson (Assistant Professor, University of California, Berkeley)
Ishita Dasgupta (researcher, DeepMind)
Qiong Zhang (Assistant Professor, Rutgers University)
Mark Ho (Assistant Professor, New York University)
Bas van Opheusden (AI Research Scientist, OpenAI)
Robert Hawkins (Assistant Professor, Stanford University)
Thomas Langlois (postdoctoral researcher, University of Texas, Austin)
Joshua Peterson (Assistant Professor, Boston University)
Bill Thompson (Assistant Professor, University of California, Berkeley)
Tom McCoy (Assistant Professor, Yale University)
Evan Russek (Assistant Professor, Hunter College)
Ilia Sucholutsky (Assistant Professor, Purdue University)
Bonan Zhao (Lecturer, University of Edinburgh)
Jian-Qiao Zhu (Assistant Professor, University of Hong Kong)