CURRICULUM VITAE   THOMAS L. GRIFFITHS

PERSONAL DETAILS

Electronic mail:	`tomg@princeton.edu`
Telephone:	(510) 642 7134 (office)
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

July 2018 -	Henry R. Luce Professor of Information Technology, Consciousness, and Culture, Departments of Psychology and Computer Science, Princeton University
July, 2017 - June, 2018	Class of 1951 Professor, Miller Institute for Basic Research in Science University of California, Berkeley
July, 2015 - June, 2017	Professor, Department of Psychology and Cognitive Science Program University of California, Berkeley
July, 2010 - January, 2018	Director, Institute of Cognitive and Brain Sciences University of California, Berkeley
July, 2010 - June 2015	Associate Professor, Department of Psychology and Cognitive Science Program University of California, Berkeley
July, 2006 - June, 2010	Assistant Professor, Department of Psychology and Cognitive Science Program University of California, Berkeley
January, 2005 - June, 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

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

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

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

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

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

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 (32,868 citations, h index of 78 via Google Scholar)

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.")

Journal articles

2. 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.

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

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

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

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

7. 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.

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

9. 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.

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

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

12. 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.

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

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

15. 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.

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

17. 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.

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

19. **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.

20. 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.

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

22. 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.

23. **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.

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

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

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

27. 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.

28. 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.

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

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

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

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

33. 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.

34. 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.

35. 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.

36. 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)

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

38. 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.

39. **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.

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

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

42. 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.

43. 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.

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

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

46. 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.

47. **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.

48. **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.

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

50. 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.

51. **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.

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

53. **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.

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

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

56. 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.

57. 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.

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

59. 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.

60. 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.

61. 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.

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

63. 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.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

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

70. 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.

71. 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.

72. 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.

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

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

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

76. **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.

77. 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.

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

79. Yeung, S., & Griffiths **T.L.** (2015). Identifying expectations about the strength of causal relationships. *Cognitive Psychology, 76,* 1-29.

80. 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.

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

82. 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.

83. 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.

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

85. 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.

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

87. 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.

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

89. 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.*

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

91. 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.

92. 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.

93. 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.

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

95. 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.

96. 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.

97. 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.

98. 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.

99. 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.

100. 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.*

101. 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.

102. 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.

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

104. 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.

105. 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.

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

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

108. 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.

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

Peer-reviewed conference papers

110. **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.*

111. **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.*

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

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

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

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

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

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

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

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

120. **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)

121. 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)

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

123. 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.*

124. 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.*

125. **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)

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

127. 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.*

128. **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.*

129. 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.*

130. 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.*

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

132. 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.

133. 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.*

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

135. 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).*

136. 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).*

137. 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).*

138. **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.*

139. 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)

140. 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.*

141. 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.*

142. 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).*

143. 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.*

144. 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).*

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

146. 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.*

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

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

149. 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.*

150. **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.*

151. 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.*

152. 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.*

153. 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)

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

155. 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.*

156. 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.*

157. 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.*

158. 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.*

159. 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.*

160. 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).*

161. **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.*

162. 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.*

163. 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.*

164. 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.*

165. Austerweil, J. & **Griffiths, T.L.** (2009). Analyzing human feature learning as nonparametric Bayesian

inference. *Advances in Neural Information Processing Systems 21.*

166. 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).*

167. 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.*

168. 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.*

169. 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.*

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

171. 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.*

172. 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.*

173. 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.*

174. 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.*

175. 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.*

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

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

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

179. 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)

180. 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.*

181. 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.*

182. 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.*

183. 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.*

184. 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.*

185. 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.*

186. 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.*

187. 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.*

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

189. 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.*

190. 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.*

191. 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).*

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

193. 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.*

194. 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.*

195. 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.*

196. 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.*

197. 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.*

198. 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.*

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

200. 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.*

201. 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.*

202. 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.*

203. 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.*

204. 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.*

205. 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.*

206. 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.*

207. **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.*

208. 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.*

209. 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.*

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

211. 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.*

212. 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.*

213. 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.*

214. 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.*

215. 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).*

216. 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.*

217. 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.*

218. 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.*

219. 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.*

220. 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.*

221. 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.*

222. 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.*

223. 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.*

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

225. 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.*

226. 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.*

227. 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.*

228. 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.*

229. 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.*

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

231. 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.*

232. 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.*

233. 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).*

234. 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.*

235. 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.*

236. 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.*

237. 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.*

238. 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.*

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

240. 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.*

241. 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.*

242. 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.

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

244. 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.*

245. 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.*

246. 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.*

247. 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.*

248. 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.*

249. 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.*

250. 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.*

251. 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.*

252. 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.*

253. 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.*

254. 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).*

255. 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.*

256. 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.*

257. 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.*

258. 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.*

259. 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.*

260. 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.*

261. 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.*

262. 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.*

263. 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.*

264. 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).*

265. 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).*

Book chapters

266. 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.

267. 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.

268. **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.

269. 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.

270. **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.

271. 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.

272. 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.

273. **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.

274. **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.

275. 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.

276. **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.

277. **Griffiths, T.L.,** Sanborn, A.N., Canini, K.R., Navarro, D.J., & Tenenbaum, J.B. (2011). Nonparamet-

ric Bayesian models of category learning. In E. M. Pothos & A. J. Wills (Eds.) *Formal approaches in categorization.* Cambridge, UK: Cambridge University Press.

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

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

280. 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.

281. 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.

282. 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.

283. 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.

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

Technical reports, invited articles, and other unreviewed publications

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

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

287. **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.

288. **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).

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

290. 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.

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

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

293. **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)

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

295. 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)

296. 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)

297. **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)

298. 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.

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

300. 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.

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

INVITED TALKS

2018 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)

FooCamp, Sebastopol, CA.

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

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

Editorial and reviewing

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.

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, 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*, *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

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.

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.

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.

UNIVERSITY AND DEPARTMENT SERVICE

University of California, Berkeley

Director of the Institute of Cognitive and Brain Sciences (2010-18).

Member of Advisory Board for Social Science Matrix (2014-18).

Chair of faculty search committee in Higher Level Cognition (2013-2015).

Member of Executive Committee for the Institute of Cognitive and Brain Sciences (2006-2010).

Member of the Computational Modeling Committee for the Cognitive Science Program (2006-2007).

Member of the Curriculum Committee for the Department of Psychology (2007-2013).

Other institutions

Member of Advisory Board for Pittsburgh Science of Learning Center (2010-2015).

TEACHING

University courses

2018 Psychology 290Q: Cognitive development for computer scientists (lecturer)

2017 Psychology 290Q: Probabilistic models of cognition (lecturer)

2016 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)
      Computer Science 294: Human-compatible AI (lecturer)

2015 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)
      Psychology 210E: Foundations of Cognition (lecturer)
      Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)

2014 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)

2013 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)
      Psychology 290Q: Computational models of cognition (lecturer)
      Psychology 210E: Foundations of Cognition (lecturer)

2011 Psychology 128/290Q: Probabilistic models of cognition (lecturer).

2010 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)
      Psychology 290Q: Computational models of cognition (lecturer)
      Psychology 210E: Foundations of Cognition (lecturer)

2009 Cognitive Science/Education C1: Introduction to Cognitive Science (lecturer)
      Psychology 128/290Q: Probabilistic models of cognition (lecturer)

2008 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)
      Psychology 290Q: Computational models of cognition (lecturer)
      Psychology 210E: Foundations of Cognition (lecturer)

2007 Cognitive Science/Education C1: Introduction to Cognitive Science (lecturer)

2006 Cognitive Science C131/Psychology C123: Computational models of cognition (lecturer)
      Psychology 290Q: Human and machine learning (lecturer)
      Cognitive Science 168: Human and machine learning (lecturer)

2005 Cognitive Science 128: Computational cognitive science (lecturer)

2002 Psychology 290: Graduate research methods (coordinator, lecturer)

2001 Psychology 290: Graduate research methods (teaching assistant)
      Psychology 253: Graduate statistics (teaching assistant)
      Psychology 1: Introduction to psychology (teaching assistant)

2000 Psychology 253: Graduate statistics (teaching assistant)

Psychology 40: Cognitive psychology (guest lecturer)

External tutorials

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.

ADVISING AND LAB ALUMNI

Undergraduate honors theses

Linsey Smith (graduate student, Northwestern University)
Aaron Beppu (software engineer, Etsy.com)
Brett Goldstein (analyst, Accenture Tech Labs)
Andy Horng (programmer, Recommind)
Nick Lewis (graduate student, Indiana University)
Jared Lorince (graduate student, Indiana University)
Jay Martin (graduate student, New York University)
Julia Ying (graduate student, Cornell University)
Rebecca Neumann (research analyst, RiskIQ)
Avi Press (software engineer, Pandora)
Benj Shapiro (researcher, Pandora)
Tiffany Hwu (graduate student, University of California, Irvine)
Priyam Das (graduate student, University of California, Irvine)

Graduate students

Sharon Goldwater (Reader, University of Edinburgh)
Adam Sanborn (Assistant Professor, University of Warwick)
Lei Shi (Associate, McKinsey & Company)
Chris Lucas (Chancellor's Fellow, University of Edinburgh)
Naomi Feldman (Assistant Professor, University of Maryland)
Kevin Canini (Software Engineer, Google)
Jing Xu (postdoctoral researcher, Johns Hopkins University)
Saiwing Yeung (faculty member, Beijing Institute of Technology)

Joseph Williams (Assistant Professor, University of Toronto)
Joe Austerweil (Assistant Professor, University of Wisconsin)
Daphna Buchsbaum (Assistant Professor, University of Toronto)
Anna Rafferty (Assistant Professor, Carleton College)
Joshua Abbott (postdoctoral researcher, University of Melbourne)
M Pacer (Software Engineer, Netflix)
Jessica Hamrick (researcher, DeepMind)
Stephan Hamrick (postdoctoral researcher, Duke University)

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)
Liz Bonawitz (Assistant Professor, Rutgers 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)