

Jake Russin

jake_russin@brown.edu

Personal website

Google Scholar

Last updated: 8/3/26

CURRENT APPOINTMENT

Brown University Postdoctoral Research Associate Computer Science; Cognitive and Psychological Sciences; Carney Institute; ARIA Institute Advisors: Dr. Michael J. Frank and Dr. Ellie Pavlick	2023 – present
--	----------------

EDUCATION

PhD in Psychology, University of California, Davis Thesis: <i>Compositionality and Cognitive Control in Neural Networks</i> Advisor: Dr. Randall O'Reilly (transferred from CU Boulder in 2019)	2017 – 2023
Non-Degree Students Program, Yale University Full course load in statistics, mathematics, machine learning	2016 – 2017
BA in Neuroscience and Philosophy, Colorado College Thesis: <i>Reasoning in the Prefrontal Cortex</i>	2010 – 2014

OTHER EMPLOYMENT

Meta AI <i>Research Intern, Advisors: Dr. Brenden Lake and Dr. Dieuwke Hupkes</i>	Summer 2022 New York, NY
Microsoft Research <i>Research Intern, Advisors: Dr. Paul Smolensky and Roland Fernandez</i>	Summer 2020 Redmond, WA (virtual)
Montreal Institute for Learning Algorithms (MILA) <i>Visiting Student Researcher, Advisors: Dr. Yoshua Bengio and Dr. Jason Jo</i>	Summer 2018 Montreal, Canada
National Institutes of Health <i>Postbac Intramural Research Training Award, Advisor: Dr. Carolyn Beebe Smith</i>	2015 – 2016 Bethesda, MD
Institute of Cognitive Science, University of Colorado, Boulder <i>Research Assistant, Advisor: Dr. Albert Kim</i>	2014 – 2015 Boulder, CO

SELECTED AWARDS AND GRANTS

T32 Training program fellowship in computational psychiatry <i>National Institute of Mental Health</i>	2025-present
Summer Seminars in Neuroscience and Philosophy Fellow (\$17,720) <i>Duke University, John Templeton Foundation</i>	2022-2023
T32 Training program fellowship in learning, memory, and plasticity <i>National Institute of Mental Health</i>	2020-2021
NSF Graduate Research Fellowships Program honorable mention <i>National Science Foundation</i>	2018
Postbac Intramural Research Training Award <i>National Institutes of Health</i>	2015-2016

Phi Beta Kappa Society	2014
The J. Glenn Gray Award in Philosophy, <i>Colorado College</i>	2014
Distinction in Neuroscience, Distinction in Philosophy, <i>Colorado College</i>	2014
Dean's List, junior and senior years, <i>Colorado College</i>	2013, 2014
Psi Chi Honor Society	2013

PUBLICATIONS

AVAILABLE PREPRINTS

Russin, J.*, McGrath, S. W.*, Williams, D. (2024) From Frege to ChatGPT: Compositionality in language, cognition, and deep neural networks. Forthcoming in *Neuroscience and Philosophy II* [arXiv]

Park S., Zolfaghar M.*, **Russin, J.***, Boorman, E., O'Reilly, R. C. (2022) The representational geometry of cognitive maps under dynamic cognitive control. *Under review at Nature Communications*. [bioRxiv]

PEER-REVIEWED JOURNAL ARTICLES AND CONFERENCE PAPERS

Russin, J., Pavlick, E., Frank, M. J. (2025) Parallel trade-offs in human cognition and neural networks: The dynamic interplay between in-context and in-weight learning. *Proceedings of the National Academy of Sciences*, 125(35). [PNAS]

Russin, J.*, Kim, J.*, Pavlick, E., Frank, M. J. (2025) Step-by-step analogical reasoning in humans and neural networks. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society*. [CogSci]

McGrath, S. W., **Russin, J.** (2024) Multiple realizability and the rise of deep learning. *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*. [CogSci]

Russin, J., McGrath, S. W., Pavlick, E., Frank, M. J. (2024) Is human compositionality meta-learned? *Commentary in Behavioral and Brain Sciences*. Target article: *Meta-Learned Models Cognition*. [BBS]

Webb T. W., Frankland, S. M., Altabaa, A., Krishnamurthy, K., Campbell, D., **Russin, J.**, O'Reilly, R., Lafferty, J., Cohen, J. D. The relational bottleneck as an inductive bias for efficient abstraction. *Trends in Cognitive Sciences*. [TICS]

McGrath, S. W.*, **Russin, J.***, Pavlick, E., Feiman, R. (2024) How can deep neural networks inform theory in psychological science? *Current Directions in Psychological Sciences*. [CDPS]

McGrath, S. W.*, **Russin, J.***, Pavlick, E., Feiman, R. (2023) Properties of LoTs: The footprints or the bear itself? *Commentary in Behavioral and Brain Sciences*. Target article: *The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences*. [BBS]

Chakravarthy, A.*, **Russin, J.***, O'Reilly, R. C. (2022) Systematicity emerges in transformers when abstract grammatical roles guide attention. *Proceedings of the 2022 Conference of the NA Chapter of the Association for Computational Linguistics, Student Research Workshop*. [NAACL SRW]

Zolfaghar M.*, **Russin, J.***, Park S.*, Boorman, E., O'Reilly, R. C. (2022) The geometry of map-like representations under dynamic cognitive control. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*. [CogSci]

Russin, J., Zolfaghar M., Park S., Boorman, E., O'Reilly, R. C. (2022) A neural network model of continual learning with cognitive control. *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*. [CogSci]

O'Reilly, R. C., Ranganath, C., **Russin, J.** (2022) The structure of systematicity in the brain. *Current Directions in Psychological Science*. [CDPS]

*Equal contribution

Russin, J., O'Reilly, R. C. (2021) Deep predictive learning with local gradient information. *Abstract accepted for poster presentation at Bernstein Conference 2021*. [Bernstein]

Russin, J., Fernandez, R., Palangi, H., Rosen, E., Jojic, N., Smolensky, P., Gao, J. (2021). Compositional processing emerges in neural networks solving math problems. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*. [CogSci]

Russin, J.*, Zolfaghar, M.*, Park, S. A., Boorman, E., O'Reilly, R. C. (2021). Complementary structure-learning neural networks for relational reasoning. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*. [CogSci]

O'Reilly, R. C., **Russin, J.**, Zolfaghar, M., Rohrlich, J. (2021). Deep predictive learning in neocortex and pulvinar. *Journal of Cognitive Neuroscience*, 33(6), 1158–1196. [JCN]

Russin, J., O'Reilly, R. C., Bengio, Y. (2020). Deep learning needs a prefrontal cortex. *Bridging AI and Cognitive Science (BAICS) Workshop, ICLR 2020, 11*. [BAICS]

Russin, J., Jo, J., O'Reilly, R. C., Bengio, Y. (2020). Systematicity in a recurrent neural network by factorizing syntax and semantics. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society, 7*. [CogSci]

O'Reilly, R. C., Nair, A., **Russin, J.**, Herd, S. A. (2020). How sequential interactive processing within frontostriatal loops supports a continuum of habitual to controlled processing. *Frontiers in Psychology, 11*. [frontiers]

Cooke, S. K., **Russin, J.**, Moulton, K., Nadel, J., Loutaev, I., Gu, Q., Li, Z., Smith, C. B. (2019). Effects of the presence and absence of amino acids on translation, signaling, and long-term depression in hippocampal slices from Fmr1 knockout mice. *Journal of Neurochemistry*. [JNC]

BOOK CHAPTERS

O'Reilly, R. C., **Russin, J.**, Herd, S. A. (2019). Computational models of motivated frontal function. *Handbook of Clinical Neurology*, 163, 317–332. [ScienceDirect]

IN PREPARATION

Russin, J.*, Kim, J.*, Pavlick, E., Frank, M. J. (*in prep*) Two routes to cognitive flexibility: geometric and algebraic generalization in humans and LLMs.

Russin, J., Bera, K., Pavlick, E., Frank, M. J. (*in prep*) Transformers capture individual differences in reinforcement learning via in-context learning.

CONFERENCE PRESENTATIONS

Russin, J. (2026, August) Transformers capture individual differences in reinforcement learning via in-context learning. Poster presented at: Cognitive Computational Neuroscience conference, New York University. Spotlight (top 10%).

Russin, J. (2026, March) Transformers capture individual differences in reinforcement learning via in-context learning. Poster presented at: Cosyne 2026, Lisbon, Portugal.

Russin, J., Kim, J. (2025, July) Step-by-step analogical reasoning in humans and neural networks. Poster presented at: CogSci 2025, San Francisco, CA.

Russin, J., Kim, J. (2025, June) Step-by-step analogical reasoning in humans and neural networks. Poster presented at: RLDM 2025, Dublin, Ireland.

Russin, J. (2025, March) Compositionality in neural networks capable of in-context learning. Talk presented at: Towards an Algorithmic Framework for Studying Compositionality in Biological and Artificial Neural Networks, Cosyne 2025 workshop, Mont-Tremblant, Quebec, CA.

- Russin, J.** (2024, August) Emergent Cognitive Flexibility in Large Language Models. Talk presented virtually at: Robophilosophy 2024, Aarhus University.
- Russin, J.** (2024, August) From Frege to chatGPT: Compositionality in language, cognition, and deep neural networks. Talk presented at: Summer Seminars in Neuroscience and Philosophy, Duke.
- Russin, J.** (2024, August) Curriculum Effects and Compositionality Emerge with In-Context Learning in Neural Networks. Poster presented at: Cognitive Computational Neuroscience conference, MIT.
- Russin, J.** and McGrath, S. (2024, July) Compositionality in Large Language Models. Talk presented at: Descartes Lectures Workshops 2024, Tilburg, the Netherlands.
- McGrath, S. and **Russin, J.** (2024, July) Multiple Realizability and the Rise of Deep Learning. Poster presented at: CogSci 2024, Rotterdam, the Netherlands.
- Russin, J.** (2024, July) Human Curriculum Effects Emerge with In-Context Learning in Neural Networks. Talk presented at: CogSci 2024, Rotterdam, the Netherlands.
- Russin, J.** (2024, May) Human Curriculum Effects Emerge with In-Context Learning in Neural Networks. Talk presented at: Control Processes 2024, Birmingham, UK.
- Russin, J.** and McGrath, S. (2023, June). Can Deep Learning Inform Explanation in Cognitive Science? Talk presented at: SPP 2023, Pittsburgh, PA.
- Russin, J.** (2022, July). The Geometry of Map-Like Representations under Dynamic Cognitive Control. Talk presented at: CogSci 2022, Toronto, CA.
- Russin, J.** (2022, July). A Neural Network Model of Continual Learning with Cognitive Control. Talk presented at: CogSci 2022, Toronto, CA.
- Russin, J.** (2022, July). Systematicity Emerges in Transformers when Abstract Grammatical Roles Guide Attention. Poster presented at: NAACL SRW 2022, Seattle, WA.
- Russin, J.** (2022, June). A Neural Network Model of Continual Learning with Cognitive Control. Poster presented at: RLDM 2022, Providence, RI.
- Russin, J.** (2022, March). The Geometry of Map-Like Representations under Dynamic Cognitive Control. Poster presented at: COSYNE 2022, Lisbon, Portugal.
- Russin, J.** (2021, September). Deep predictive learning with local gradient information. Poster presented at: Bernstein Conference 2021.
- Russin, J.** (2021, July). Complementary structure-learning neural networks for relational reasoning. Poster presented at: 43rd Annual Meeting of the Cognitive Science Society.
- Russin, J.** (2021, July). Compositional processing emerges in neural networks solving math problems. Poster presented at: 43rd Annual Meeting of the Cognitive Science Society.
- Russin, J.** (2020, April) Deep learning needs a prefrontal cortex. Talk presented at: Bridging AI and Cognitive Science, ICLR 2020 Workshop, Addis Ababa, Ethiopia, April 26, 2020.
- Russin, J.** (2020, July). Systematicity in a recurrent neural network by factorizing syntax and semantics. Talk presented at: 42nd Annual Meeting of the Cognitive Science Society.
- Russin, J.** (2020, July). Compositional generalization by factorizing alignment and translation. Talk presented at: ACL 2020, Student Research Workshop, Seattle, WA.

INVITED TALKS

- 10/2025 **Russin, J.** Parallel trade-offs in human cognition and neural networks: The dynamic interplay between in-context and in-weight learning. MindRL Hub.

05/2025	Russin, J. The dynamic interplay between in-context and in-weight learning in humans and neural networks. Schuck lab, University of Hamburg.
04/2022	Russin, J. Continual learning and map-like representations in neural networks. Human information Processing lab (Summerfield), Oxford University, Oxford, UK.
04/2022	Russin, J. Map-like representations and relational reasoning in neural networks. The Reasoning Lab (Khemlani), Naval Research Laboratory, Washington DC
07/2021	Russin, J. Structure-learning and compositionality in neural networks. Laboratory of Neural Computation and Cognition (Frank), Brown University, Providence, RI.
03/2021	Russin, J. Complementary structure-learning systems. PCCN Brownbag, Department of Psychology, University of California Davis, Davis, CA.
05/2020	Russin, J. Systematicity in neural networks. PCCN Brownbag, Department of Psychology, University of California Davis, Davis, CA.
04/2020	Russin, J. Generalization, structure, and cognitive maps. Memory Meeting, Center for Neuroscience, University of California Davis, Davis, CA.
09/2019	Russin, J. Temporal abstraction in reinforcement learning. Learning and Decision-Making Lab, Center for Mind and Brain, University of California Davis, Davis, CA.
05/2019	Russin, J. Systematicity, selective attention, and the prefrontal cortex. 38th Annual Ekstrand Mini-Convention, University of Colorado, Boulder, Boulder, CO.
12/2018	Russin, J. Compositional Attention Networks. Statistics, Optimization, and Machine Learning Seminar, University of Colorado, Boulder, Boulder, CO.
10/2018	Russin, J. Temporal abstraction in reinforcement learning. Kilpatrick Group, Department of Applied Mathematics, University of Colorado, Boulder, Boulder, CO.
05/2018	Russin, J. Temporal abstraction in reinforcement learning. 37th Annual Ekstrand Mini-Convention, University of Colorado, Boulder, Boulder, CO.
05/2014	Russin, J. Reasoning in the prefrontal cortex. Psychology and Neuroscience Poster Day, Colorado College, Colorado Springs, CO.

TEACHING

9/24-present	SOLAR undergraduate research club, Brown University <i>Help lead weekly meeting for LUNAR lab undergraduates to present research.</i>
Spring 2026	Teaching assistant: Computational Cognitive Neuroscience, Brown University. <i>Led weekly lab sections, gave a guest lecture, graded.</i>
Jan. 2025	Primer on large language models for cognitive scientists, Brown University. <i>Lectured on LLMs for Cognitive and Psychological Sciences Retreat 2025</i>
Spring 2018	Teaching assistant: Research Methods in Psychology, CU Boulder. <i>Lectured and led discussions in weekly lab sections, graded, held weekly office hours.</i>
Fall 2017	Teaching assistant: History of Psychology, CU Boulder. <i>Graded, attended lectures, held weekly office hours.</i>

MENTORING

2024 – present	Tianze (Etha) Hua (PhD student, Computer Science, Brown) <i>Research mentor for projects on mechanistic interpretability in LLMs</i>
----------------	---

2024 – present	Joonhwa Kim (PhD student, Neuroscience, Brown) <i>Research mentor for projects involving analogical reasoning in humans and LLMs</i>
2023 – present	Krishn Bera (PhD student, Cognitive and Psychological Sciences, Brown) <i>Research mentor for projects involving LLMs and cognitive control</i>
2021 – 2023	Ayush Chakravarthy (undergraduate, Computer Science, UC Davis) <i>Research mentor for projects in deep learning and natural language processing.</i>
2020 – 2022	Andrew Carlson (PhD student, Psychology, UC Davis) <i>Research mentor and advisor on career development, publishing, presentations, etc.</i>
2021	Graduate student mentor for CogSci 2021 Mentoring Program
2020 – 2021	Taiqi He (junior specialist postbac, UC Davis) <i>Research mentor for projects in deep learning and natural language processing.</i>

SERVICE

2026 – present	Reviewing Editor: <i>eLife</i>
2019 – present	Ad hoc reviewer: <i>Nature</i> , <i>PNAS</i> , <i>Nature Neuroscience</i> , <i>Nature Human Behavior</i> , <i>Nature Communications Biology</i> , <i>Neuron</i> , <i>TICS</i> , <i>Current Opinion in Behavioral Sciences</i> , <i>Psychonomic Bulletin and Review</i>
2020 – present	Conference reviewer: CogSci conference (2020, 2021, 2022, 2023, 2024, 2025), CCN conference (2024, 2026), RLDM conference (2025)
2026	Workshop organizer: “Mechanistic interpretability in brains and machines: shared concepts and converging methods” COSYNE 2026
2023 – 2024	Carney Brain Science External Postdoc Seminar, Speaker Selection Committee
2022 – 2024	Reading group on AI in philosophy and cognitive science: led weekly zoom meeting for researchers at multiple institutions (UC Davis, Brown, UPenn, Stanford)