

# Marcel Binz

Ingolstädter Landstraße 1, 85764 Oberschleißheim, Germany

✉ marcel.binz@helmholtz-munich.de 🌐 marcelbinz.github.io 🔊 marcelbinz ☎ (+49)15203605574

## RESEARCH INTERESTS

Cognitive Science; Machine Learning; Meta-Learning; Resource Rationality; Large Language Models; Deep Learning; Bayesian Inference; Information Theory; Decision-Making; Reinforcement Learning

## POSITION

**Helmholtz Munich**, Institute for Human-Centered AI 12/2023 - present  
Research scientist and deputy head

## EXPERIENCE

**Max Planck Institute for Biological Cybernetics**, PI: Dr. Eric Schulz 02/2021 - 11/2023  
Postdoctoral researcher

**Harvard University**, PI: Prof. Samuel Gershman 09/2019 - 12/2019  
Research visit

**Facebook Inc.** 06/2016 - 12/2016  
Research internship

**Eberhard Karls Universität Tübingen**, PI: Prof. Martin Butz 04/2015 - 08/2015  
Research assistant

## EDUCATION

**Philipps-Universität Marburg**, PI: Prof. Dominik Endres 2018 - 2021  
Dr. rer. nat. (Psychology)

**KTH Royal Institute of Technology, Stockholm** 2015 - 2018  
M.Sc. (Machine Learning)

**Eberhard Karls Universität Tübingen** 2012 - 2015  
B.Sc. (Cognitive Science)

## PEER-REVIEWED PUBLICATIONS

*Note that in machine learning, unlike most other fields, conference publications are rigorously reviewed. Proceedings of selective conferences are considered archival and comparable to journals.*

**Binz, M.**, Alaniz, S., Roskies, A., Aczel, B., Bergstrom, C., Allen, C., Schad, D., Wulff, D. U., West, J., Zhang, Q., Shiffrin, R., Gershman, S. J., Popov, V., Bender, E. M., Marelli, M., Botvinick, M. M., Akata, Z. and Schulz, E., accepted. How should the advent of large language models affect the practice of science? *Proceedings of the National Academy of Sciences (PNAS)*.

Demircan, C., Saanum, T., Pettini, L., **Binz, M.**, Baczkowski, B. M., Doeller, C., Garvert, M. and Schulz, E., 2024. Evaluating alignment between humans and neural network representations in image-based learning tasks. *Conference on Neural Information Processing Systems (NeurIPS)*.

Hussain, Z., **Binz, M.**, Mata, R. and Wulff, D. U., 2024. A tutorial on open-source large language models for behavioral science. *Behavior Research Methods*.

- Coda-Forno, J., **Binz, M.**, Wang, J. X. and Schulz, E., 2024. CogBench: a large language model walks into a psychology lab. *International Conference on Machine Learning (ICML)*.
- Jagadish, A. K., Coda-Forno, J., Thalmann, M., Schulz, E. and **Binz, M.**, 2024. Human-like category learning by injecting ecological priors from large language models into neural networks. *International Conference on Machine Learning (ICML)*.
- Schubert, J. A., Jagadish, A. K., **Binz, M.** and Schulz, E., 2024. In-context learning agents are asymmetric belief updaters. *International Conference on Machine Learning (ICML)*.
- Binz, M.** and Schulz, E., 2024. Turning large language models into cognitive models. *International Conference on Learning Representations (ICLR)*.
- Binz, M.**, Dasgupta, I., Jagadish, A., Botvinick, M., Wang, J. X. and Schulz, E., 2023. Meta-learned models of cognition. *Behavioral and Brain Sciences (BBS)*.
- Coda-Forno, J., **Binz, M.**, Akata, Z., Botvinick, M., Wang, J. X. and Schulz, E., 2023. Meta-in-context learning in large language models. *Conference on Neural Information Processing Systems (NeurIPS)*.
- Saanum, T., Éltető, N., Dayan, P., **Binz, M.** and Schulz, E., 2023. Reinforcement learning with simple sequence priors. *Conference on Neural Information Processing Systems (NeurIPS)*.
- Schulze Buschoff, L. M., Schulz, E. and **Binz, M.**, 2023. The acquisition of physical knowledge in generative neural networks. *International Conference on Machine Learning (ICML)*.
- Binz, M.** and Schulz, E., 2023. Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences (PNAS)*.
- Binz, M.** and Schulz, E., 2022. Reconstructing the Einstellung effect. *Computational Brain & Behavior*.
- Binz, M.** and Schulz, E., 2022. Modeling human exploration through resource-rational reinforcement learning. *Conference on Neural Information Processing Systems (NeurIPS)*. **Selected as Oral**.
- Binz, M.**, Gershman, S.J., Schulz, E. and Endres, D., 2022. Heuristics from bounded meta-learned inference. *Psychological Review*.
- Brändle, F., **Binz, M.** and Schulz, E., 2022. Exploration beyond bandits. *The Drive for Knowledge: The Science of Human Information Seeking*. Cambridge University Press.

## PREPRINTS

---

- Binz, M.**, Akata, E., Bethge, M., Brändle, F., Callaway, F., Coda-Forno, J., . . . and Schulz, E., 2024. Centaur: a foundation model of human cognition.
- Demircan, C., Saanum, T., Jagadish, A. K., **Binz, M.** and Schulz, E., 2024. Sparse autoencoders reveal temporal difference learning in large language models.
- Jagadish, A. K., **Binz, M.**, Saanum, T., Wang, J. X. and Schulz, E., 2023. Zero-shot compositional reinforcement learning in humans.
- Coda-Forno, J., Witte, K., Jagadish, A. K., **Binz, M.**, Akata, Z. and Schulz, E., 2023. Inducing anxiety in large language models increases exploration and bias.

## NON-ARCHIVAL PUBLICATIONS

---

- Schubert, J. A., Jagadish, A. K., **Binz, M.** and Schulz, E., 2023. A rational analysis of the optimism bias using meta-reinforcement learning. *Conference on Cognitive Computational Neuroscience (CCN 2023)*.
- Jagadish, A. K., Saanum, T., Wang, J. X., **Binz, M.** and Schulz, E., 2022. Probing compositional

inference in natural and artificial agents. *5th Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM 2022)*.

Demircan, C., Pettini, L., Saanum, T., **Binz, M.**, Baczkowski, B. M., Doeller, C., . . . and Schulz, E., 2022. Decision-making with naturalistic options. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

**Binz, M.** and Endres, D., 2019. Emulating human developmental stages with bayesian neural networks. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

**Binz, M.** and Endres, D., 2019. Where do heuristics come from?. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Butz, M. V., Simonic, M., **Binz, M.**, Einig, J., Ehrenfeld, S. and Schrod, F., 2016. Is it living? Insights from modeling event-oriented, self-motivated, acting, learning and conversing game agents. *Proceedings of the Annual Meeting of the Cognitive Science Society*.

**Binz, M.**, Otte, S. and Zell, A., 2015. On the applicability of recurrent neural networks for pattern recognition in electroencephalography signals. *Workshop New Challenges in Neural Computation*.

## TEACHING

---

<b>Computational cognitive science</b> , Eberhard Karls University of Tübingen Lecturer	2022, 2023
<b>International Interdisciplinary Computational Cognitive Science Summer School</b> Lecturer	2022, 2023
<b>Bayesian statistics and machine learning</b> , Philipps-Universität Marburg Lecturer	2020
<b>Theoretical neuroscience</b> , Philipps-Universität Marburg Lecturer	2019, 2020
<b>Deep learning in data science</b> , KTH Royal Institute of Technology Teaching assistant	2017

## SUPERVISION (PHD STUDENTS)

---

<b>Julian Coda-Forno</b> (co-supervised with Eric Schulz and Jane Wang) Meta-Learning in large language models	2022 - present
<b>Akshay Kumar Jagadish</b> (co-supervised with Eric Schulz) Reverse-engineering adaptive principles of cognition	2021 - present

## SUPERVISION (MASTER AND BACHELOR STUDENTS)

---

<b>Elif Kara</b> Heuristic decision-making in the wild	2024
<b>Johannes Schubert</b> Investigating the optimism bias using meta-reinforcement learning	2023
<b>Luca Schulze Buschoff</b> Development as decompression	2022
<b>Akshay Kumar Jagadish</b> Compositional generalization in meta-reinforcement learning	2021
<b>Gwen Hirsch</b>	2020

Comparing meta-learners with human performance in a continual learning framework	
<b>Hauke Niehaus</b>	<b>2019</b>
Simulating decision-making deficits in a deep meta-reinforcement-learning agent	

## WORKSHOP ORGANIZATION

---

<b>In-context learning in natural and artificial intelligence</b> , Rotterdam	<b>2024</b>
The Annual Meeting of the Cognitive Science Society	
<b>Meta-learned models of cognition</b> , Freiburg	<b>2022</b>
The Biannual Conference of the German Cognitive Science Society	

## MEMBERSHIPS

---

<b>ELLIS Society</b>	<b>2024 - present</b>
<b>Cognitive Science Society</b>	<b>2024 - present</b>

## REVIEWING

---

<b>Nature Communications</b>	<b>2024 - present</b>
<b>Open Mind</b>	<b>2024 - present</b>
<b>Nature</b>	<b>2023 - present</b>
<b>Nature Human Behaviour</b>	<b>2023 - present</b>
<b>International Conference on Learning Representations (ICLR)</b>	<b>2023 - present</b>
<b>Conference on Neural Information Processing Systems (NeurIPS)</b>	<b>2023 - present</b>
<b>Behavior Research Methods</b>	<b>2023 - present</b>
<b>Trends in Cognitive Sciences</b>	<b>2023 - present</b>
<b>Conference on Cognitive Computational Neuroscience</b>	<b>2023 - present</b>
<b>Proceedings of the National Academy of Sciences (PNAS)</b>	<b>2022 - present</b>
<b>Psychological Review</b>	<b>2022 - present</b>
<b>Computational Brain &amp; Behavior</b>	<b>2022 - present</b>
<b>Annual Meeting of the Cognitive Science Society</b>	<b>2021 - present</b>

## INVITED TALKS

---

<b>CCCM seminar series</b> , Birkbeck	<b>2025</b>
<b>TeaP symposium on large language models in psychological research</b> , Frankfurt	<b>2025</b>
<b>NeurIPS Pre-Workshop on Behavioral ML</b>	<b>2024</b>
<b>MindRL Hub</b>	<b>2024</b>
<b>Google DeepMind</b> , London	<b>2024</b>
<b>University of Oxford</b> , Oxford	<b>2024</b>
<b>DGPs symposium on large language models in psychological research</b> , Vienna	<b>2024</b>
<b>Rational altruism lab meeting</b> , Los Angeles	<b>2024</b>
<b>Higher cognition in large language models symposium</b> , Rotterdam	<b>2024</b>
<b>Connecting minds and machines symposium</b> , Munich	<b>2024</b>

Annual Summer Interdisciplinary Conference, Molveno	2024
Departmental research seminar, Milan	2024
Cognition, values & behaviour and crowd cognition joint lab meeting, Munich	2024
Cognitive sciences colloquium, Irvine	2024
Bosch neuro-symbolic AI focus group	2024
Cognition, brain, & behavior research seminar, Harvard	2023
nEuro-economics seminar series, Paris	2023
Digital change symposium, Kloster Seeon	2023
International Interdisciplinary Computational Cognitive Science Summer School	2023
Language models in judgment and decision making research symposium, Vienna	2023
Large language models meet cognitive science workshop, Sydney	2023
Neuro-cognitive modeling group lab meeting, Tübingen	2023
International Titisee Conference on NeuroAI, Titisee	2023
Colloquium of the Institute of Cognitive Science, Osnabrück	2023
Reinforcement learning and decision-making seminar, Tübingen	2023
Conference on Neural Information Processing Systems, New Orleans	2022
Memory, judgement and decision-making seminar, Mannheim	2022
International Interdisciplinary Computational Cognitive Science Summer School	2022
Conference of the German Cognitive Science Society, Freiburg	2022
Human and machine cognition lab meeting, Tübingen	2022
Reinforcement learning and decision-making seminar, Tübingen	2021
Joint lab retreat: Summerfield, Schuck, Schulz	2021
Colloquium of the Institute for Neuroinformatics, Bochum	2019

## AWARDS

---

<b>German Cognitive Science Society Best Publication Award</b>	2018 - 2022
Best publication in cognitive science by a young investigator	
<b>EuroCogSci 2019 Best Poster Award</b>	2019
Best poster presentation	
<b>DMV-Abiturpreis</b>	2010
Excellent performance in high school mathematics	

## GRANTS

---

<b>Google Gemma Academic Program</b>	2024
\$5000 in credits for the Google Cloud Platform	
<b>Scientific Inference and Statistical Inference Conference</b>	2023
Funding for travel and accommodation	
<b>International Titisee Conference on NeuroAI</b>	2023
Funding for travel and accommodation	
<b>German Academic Exchange Service (DAAD) Scholarship</b>	2019
Funding for a three month research visit at Harvard University	
<b>Summer Institute on Bounded Rationality</b>	2019

Funding for travel and accommodation

## PUBLIC OUTREACH

---

<b>Science Journal for Kids</b> version of Using cognitive psychology to understand GPT-3 [Link]	2023
<b>Podcast on our paper</b> Using cognitive psychology to understand GPT-3 [Link]	2023
<b>Interview with the newspaper</b> Tagesspiegel [Link]	2023
<b>Talk for an event on short stories about AI</b> hosted by the newspaper Tagblatt [Link]	2023
<b>KFZ Science Slam</b> [Link]	2019
<b>Mario Lives! AAI Video Competition</b> winning submission [Link; one million views on YouTube]	2015