Dave F. Kleinschmidt

dave.f.kleinschmidt@gmail.com
www.davekleinschmidt.com
github.com/kleinschmidt

I'm a software engineer, data scientist, and cognitive scientist with more than 10 years of experience making complex ideas, data, and systems legible to experts and beginners alike. I have a track record of shipping software that people actually use, a passion for making computational tools broadly useful for humans through open source software, and a knack for straightforward and intuitive presentation of difficult computational ideas in both software and written/oral forms.

Experience

2022–present **Senior Algorithm/Research Software Engineer**, *Beacon Biosignals*, Remote
As IC and tech lead of Quantitative Tools Team, built out Beacon's **platform for algorithm development and quantitative work**: data provenance/versioning system used by all algorithm and life science teams, distributed ML training/evaluation capability, distributed batch building for large time-series biosignal datasets (OndaBatches.jl). Tech lead for Julia-language runtime for Ray.io.

2021–2022 **Research Scientist**, *Beacon Biosignals*, Remote
Developed, applied, and interpreted novel statistical and machine learning models to provide insight into human EEG data for clinical and industry partners. Tech lead and project manager for quantitative science team handling Beacon's first large contract: cleaned, featurized, and analyzed 10,000s of overnight EEG recordings from a Phase 3 clinical trial.

2013–present **Open source maintainer**, *JuliaStats organization*, Distributed/remote
Maintain and develop open source software for statistical modeling in the Julia language. Write code and documentation, review contributions, manage releases and automated testing/CI. Primary developer of **StatsModels.jl**, used in 100+ packages to transform tabular data to numerical arrays.

2018–2021 **Assistant Professor of Psychology**, *Rutgers University*, New Brunswick, NJ
Directed computational cognitive science lab; Supervised graduate research assistants and staff. Designed, implemented (custom node.js+MongoDB backend), and analyzed data (using R and Julia) from **online behavioral experiments**. Designed and implement machine learning models of cognition.

2016–2018 **CV Starr Postdoctoral Fellow**, *Princeton Neuroscience Institute*, Princeton, NJ
Implemented **custom Bayesian nonparametric models** of human perception and categorization. Co-designed and ran **weekly open workshop on statistical philosophy and methods** for PhD students and other post-docs (regression, hierarchical models, Bayesian methods)

2010–2016 **PhD research**, *Brain and Cognitive Sciences, University of Rochester*, Rochester, NY
Created **novel Bayesian theory of perceptual learning for speech**, implemented with MCMC sampler in custom R code and Stan. Developed and deployed custom web-based auditory psychophysics experiments in javascript. Lead machine learning analysis of multimodal data from brain imaging (fMRI) experiments of human perceptual learning.

2009–2010 **Baggett Fellow**, *Linguistics, University of Maryland*, College Park, MD
Designed and taught one-day **workshop on mixed-effects models and ANOVA**. Research on penalized matrix factorization (sparse coding) models of statistical learning

Education

2010–2016 **Ph.D. Brain and Cognitive Sciences**, *University of Rochester*, Rochester, NY

2005–2009 **B.A. Mathematics, concentration Cognitive Science**, *Williams College*, Williamstown, MA, *Summa cum laude*, highest honors in Cognitive Science

Skills

Statistics	Linear/logistic regression, mixed-effects models, nonparametrics, Bayesian methods	Distributed systems	Kubernetes, stream/batch processing, Julia multiprocessing, Ray.io
Software Engineering	Git, Github (Actions), Docker, AWS (EC2, ECR, EKS, SQS), Arrow	Machine Learning	Predictive models, classification, clustering, NLP, deep learning

Programming Languages

Julia	DataFrames.jl, GLM.jl, MixedModels.jl, Makie.jl, Flux.jl	R	tidyverse (dplyr/purrr/tidyr), ggplot2, lme4/brms, rstan
Python	numpy, pandas, scikit-learn, ray	Javascript	Node.js/express, frontend/JQuery
Linux	bash, zsh, git, server/desktop	*etc.*	C++, (Postre)SQL, MongoDB, Matlab, Java, Lisp, Perl, LaTeX

Selected Publications

2020 Wu*, M.-H., **Kleinschmidt*, D. F.,** Emberson, L., Doko, D., Edelman, S., Jacobs, R., & Raizada, R. Cortical transformation of stimulus-space in order to linearize a linearly inseparable task. *Journal of Cognitive Neuroscience*, *Early Access*, 1–13. https://doi.org/10.1162/jocn_a_01533

2019 **Kleinschmidt, D. F.** Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, *34*(1), 43–68. https://doi.org/10.1080/23273798.2018.1500698

Kleinschmidt, D. F., & Hemmer, P. A Bayesian model of memory in a multi-context environment. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Conference of the Cognitive Science Society*. Cognitive Science Society. osf.io/vuksn/

Yarkoni, T., Markiewicz, C., de la Vega, A., Gorgolewski, K., Salo, T., Halchenko, Y., McNamara, Q., DeStasio, K., Poline, J.-B., Petrov, D., Hayot-Sasson, V., Nielson, D., Carlin, J., Kiar, G., Whitaker, K., DuPre, E., Wagner, A., Tirrell, L., Jas, M., … Blair, R. PyBIDS: Python tools for BIDS datasets. *Journal of Open Source Software*, *4*(40), 1294. https://doi.org/10.21105/joss.01294

2018 **Kleinschmidt, D. F.** Learning distributions as they come: Particle filter models for online distributional learning of phonetic categories. In T. T. Rogers, X. Rau, X. Zhu, & C. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 1933–1938). Cognitive Science Society. https://doi.org/10.31234/osf.io/dymc8

Kleinschmidt, D. F., Weatherholtz, K., & Jaeger, T. F. Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science*, *10*(4), 818–834. https://doi.org/10.1111/tops.12331

2016 **Kleinschmidt, D. F.,** & Jaeger, T. F. Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic Bulletin & Review*, *23*(3), 678–691. https://doi.org/10.3758/s13423-015-0943-z

Pajak, B., Fine, A. B., **Kleinschmidt, D. F.,** & Jaeger, T. F. Learning additional languages as hierarchical probabilistic inference: Insights from first language processing. *Language Learning*, *66*(4), 900–944. https://doi.org/10.1111/lang.12168

2015 **Kleinschmidt, D. F.,** & Jaeger, T. F. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2). https://doi.org/10.1037/a0038695

2014 Salverda, A. P., **Kleinschmidt, D. F.,** & Tanenhaus, M. K. Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Zaki, S. R., **& Kleinschmidt, D. F.** Procedural memory effects in categorization: Evidence for multiple systems or task complexity? *Memory & cognition*, *42*(3), 508–24. https://doi.org/10.3758/s13421-013-0375-9

2011 Croft, W., Bhattacharya, T., **Kleinschmidt, D. F.,** Smith, D. E., & Jaeger, T. F. Greenbergian universals, diachrony, and statistical analyses. *Linguistic Typology*, *15*(2), 433–453. https://doi.org/10.1515/LITY.2011.029

* Indicates equal contributions.