Benjamin Plaut

bjplaut@gmail.com Google Scholar

Education

PhD in Computer Science, Stanford University *Sept 2016 – Sept 2021*

Thesis: *Algorithms for Fair Public and Private Resource Allocation*
Thesis advisor: Ashish Goel

Bachelor of Science, Carnegie Mellon University *Aug 2012 – May 2016*

Thesis: *Algorithms for Social Good: Kidney Exchange*
Thesis advisor: Tuomas Sandholm
Allen Newell Award for Excellence in Undergraduate Research (best thesis in Computer Science)

Work Experience

Postdoctoral Researcher, UC Berkeley *Sept 2023 – present*

- Part of the Center for Human-Compatible AI
- Mentored by Stuart Russell
- Studying AI safety, with a focus on generalization

Data Scientist, Lyft *June 2021 – May 2023*

- Co-lead for multi-year initiative with 30+ people
- Created cross-org experimentation principles to ensure scientific rigor (used by 20+ people)
- Grew annual profit and revenue by millions of dollars
- Used supervised learning and linear programming to create a novel way of matching riders and drivers
- Gave "Intro to Effective Altruism" talk (link)

Research Intern, Google *June 2020 – Nov 2020*

- Designed and implemented reinforcement learning algorithm for finding optimal routing policies
- Proved sublinear regret bound
- Evaluated algorithm's regret empirically

Research Intern, Google *June 2019 – Sept 2019*

- Investigated possible definitions of fairness in two-sided markets
- Designed a matching algorithm with provable fairness guarantees

Publications

The author order for my PhD work is alphabetical by field convention. See the key below.
Lead author: *
Senior author: †

- **B Plaut***, Hanlin Zhu, Stuart Russell†. Avoiding Catastrophe in Online Learning by Asking for Help. Working paper. (link)
- **B Plaut***, Khanh Nguyen, Tu Trinh. Probabilities of Chat LLMs Are Miscalibrated but Still Predict Correctness on Multiple-Choice Q&A Working paper. (link)
- T Trinh*, M Danesh, K Nguyen, **B Plaut**†. Getting By Goal Misgeneralization With a Little Help From a Mentor. NeurIPS 2024 Workshop on Safe and Trustworthy Agents. (link)
- A Goel†, **B Plaut***. Counteracting Inequality in Markets via Convex Pricing. WINE 2020. (link)
- S Gollapudi†, K Kollias, **B Plaut***. Almost Envy-free Repeated Matching in Two-sided Markets. WINE 2020. (link)
- B Plaut. Optimal Nash Equilibria for Bandwidth Allocation. WINE 2020. (link)

- N Immorlica†, **B Plaut***, EG Weyl†. Equality of Power and Fair Public Decision-making. WINE 2019. (link)
- A Goel†, R Hulett, **B Plaut***. Markets Beyond Nash Welfare for Leontief Utilities. WINE 2019. (link)
- **B Plaut*** and T Roughgarden†. Communication Complexity of Discrete Fair Division. SODA 2019, SICOMP 2020. (link)
- N Garg, A Goel†, **B Plaut***. Markets for Public Decision-making. WINE 2018. (link)
- **B Plaut*** and T Roughgarden†. Almost Envy-Freeness with General Valuations. SODA 2018, SIDMA 2020. (link)
- H Xu*, **B Plaut**, X Zhu, M Chen, U Mavinkurve, A Maiti, G Song, K Murari, and M Mandal†, Direct Observation of Folding Energy Landscape of RNA Hairpin at Mechanical Loading Rates. *The Journal of Physical Chemistry*, 2017. (link)
- JP Dickerson, D Manlove†, **B Plaut**, T Sandholm†, and J Trimble*. Position-Indexed Formulations for Kidney Exchange. EC 2016. (link)
- **B Plaut***, JP Dickerson, and T Sandholm†. Fast Optimal Clearing of Capped-Chain Barter Exchanges. AAAI 2016. (link)

Other Experience

President, Co-founder, and primary instructor, Cardinal West Coast Swing *Nov 2018 – Aug 2020*

- Co-founded a west coast swing club at Stanford University.
- Designed a curriculum for beginner lessons
- Taught weekly lessons
- Trained and supervised new teachers

Teaching Assistant, Stanford University *Mar 2018 – June 2018*

- Duties included office hours, advising students on a final project, grading, and overall course structuring

Teaching Assistant, Carnegie Mellon University *Jan 2014 – Dec 2015*

- Taught recitation, ran review sessions, mentored students on a final project
- Also was "Head of Staff Morale" for a staff of 40 TAs. Organized social events, coordinated purchases, etc.

Languages

Programming Languages	Python, C++, SQL, Unix, LaTeX, Matlab
Natural Languages	English (native), Spanish (advanced proficiency)

Personal projects

- Music composition – I write music under the name Melancholy Flower (pun on "melon cauliflower"). Check out my Spotify!
- RAMbrandt – An algorithmic art-generator based on Markov chains, computer vision, and vector calculus.
- Ballroom Dance Video Tutorial – Database containing demonstration videos for various ballroom dance videos