

JAMES HONAKER

james@hona.kr
http://hona.kr

Education

- 2004 **Harvard University**, PhD, Department of Government.
1997 **Caltech**, BS, Social Science.

Positions

- 2020–Present Research Scientist, **Facebook**
2020–Present Chief Privacy Engineer, **OpenDP**
2020–Present Associate, **Harvard John A. Paulson School of Engineering and Applied Sciences**
2019 Lecturer on Computer Science, **Harvard University**
2016–2020 Research Associate, **Center for Research on Computation and Society, Harvard John A. Paulson School of Engineering and Applied Sciences**
2012–16 Senior Research Scientist, **Institute for Quantitative Social Science, Harvard University**
2009–12 Lecturer, **Pennsylvania State University**
2004–09 Assistant Professor, **UCLA**.
2001–04 Acting Assistant Professor, **UCLA**.

Publications

- 2019 D’Orazio, Vito, James Honaker, Raman Prasad, Michael Shoemate. “Modeling and Forecasting Armed Conflict: AutoML with Human-Guided Machine Learning” *Proceedings of 3rd International Workshop on Big Data Analytics for Cyber Intelligence and Defense 2019 (BDA4CID 2019) at IEEE BigData 2019*, Los Angeles, CA, December 2019.
- 2019 Gil, Yolanda, James Honaker, Shikhar Gupta, Yibo Ma, Vito D’Orazio, Daniel Garijo, Shruti Gadewar, Qifan Yang, Neda Jahanshad. “Towards Human-Guided Machine Learning” *Proceedings of the 24th ACM International Conference on Intelligent User Interfaces (IUI)*, Los Angeles, CA, March 2019.
- 2019 Plutzer, Eric, Michael B. Berkman, James Honaker, Christopher Ojeda, and Anne Whitesell. “Measuring Complex State Policies: Pitfalls and Considerations, with an Application to Race and Welfare Policy.” *Policy Studies Journal* doi: 10.1111/psj.12307
- 2018 Wood, Alexandra, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O’Brien, Thomas Steinke, and Salil Vadhan. “Differential Privacy: A Primer for a Non-technical Audience” *Vanderbilt Journal of Entertainment & Technology Law* 21(1):209-275.
- 2017 Blackwell, Matthew, James Honaker and Gary King. “A Unified Approach to Measurement Error and Missing Data: Overview and Applications” *Sociological Methods and Research* 46(3):303-341.

- 2017 Blackwell, Matthew, James Honaker and Gary King. "A Unified Approach to Measurement Error and Missing Data: Details and Extensions" *Sociological Methods and Research* 46(3):342-369.
- 2015 Crosas, Mercè, Gary King, James Honaker and Latanya Sweeney "Automating Open Science for Big Data" *Annals of the American Academy of Political and Social Science* 659(1): 260–273.
- 2014 Honaker, James and Vito D’Orazio. "Statistical Modeling by Gesture: A graphical, browser-based statistical interface for data repositories" *Extended Proceedings of ACM Hypertext 2014*.
- 2011 Honaker, James, Gary King and Matthew Blackwell. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7): 1-47.
- 2010 Honaker, James and Gary King. "What to do About Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54(2): 561-581.
- 2008 Frisina, Laurin, Michael Herron, James Honaker and Jeffrey B. Lewis. "Ballot Formats, Touchscreens, and Undervotes: A Study of the 2006 Midterm Elections in Florida." *Election Law Journal* 7(1):25-47.
- 2002 Honaker, James, Jonathan Katz and Gary King. "A Fast, Easy, and Efficient Estimator for Multiparty Electoral Data" *Political Analysis* 10(1):84-100.
- 2001 King, Gary, James Honaker, Anne Joseph and Kenneth Scheve. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation" *American Political Science Review* 95(1):49-69.

Statistical Software

- 2020-2021 *SmartNoise* (with Microsoft and Harvard Privacy Tools Group) Differentially private curator system for Azure cloud computation.
<https://smartnoise.org>
- 2017-2021 *AutoML-RPC* A flexible abstraction API between user interfaces (GUIs and notebooks) and automated machine learning (AutoML) engines, using an ML pipeline description language and Protobuff. Initially developed under the Data-Driven Discovery of Models (D3M) DARPA program.
<https://gitlab.com/datadrivendiscovery/automl-rpc/>
- 2015-2020 *PSI (Ψ) A Private data Sharing Interface* (with Privacy Tools Group) This prototype system allows researchers with sensitive datasets to make differentially private statistics about their data available through data repositories using the Dataverse platform.
<https://privacytools.seas.harvard.edu/psi>
- 2014-2021 *TwoRavens* (with Vito D’Orazio) A gesture-driven, browser based, thin-client interface for conducting statistical analysis on datasets in archives and repositories and differentially private secure data.
<http://2ra.vn>
- 2014-2016 *Zelig (version 5)* (with Christine Choirat, Christopher Gandrud, Kosuke Imai, Gary King and Olivia Lau) A library for statistical inference in R that provides access to dozens of statistical models through a common call structure, with

utilities for simplified model interpretation. Also wrote the related modules *ZeligChoice* and *ZeligEI*.

<http://zeligproject.org>

2005-2012 *Amelia II*: A Program for Missing Data (with Matthew Blackwell and Gary King) **Winner of the 2014 Best Statistical Software Award of the Society for Political Methodology.**

<http://gking.harvard.edu/amelia>

1998-2004 *Amelia*: A Program for Missing Data (with Anne Joseph, Gary King and Ken Scheve)

Also, contributor/maintainer to *Zelig(v4)*, *ei*, *Consilience*, *ReadMe*, *WhatIf*, *cem*, *D³M*.

Working Papers

2017 Thomas Brawner and James Honaker. "Bootstrap Inference and Differential Privacy" Presented at Theory and Practice of Differential Privacy 2017, Dallas TX., October 2017.

2016 Marco Gaboardi, James Honaker, Gary King, Jack Murtagh, Kobbi Nissim, Jonathan Ullman and Salil Vadhan "PSI (Ψ): A Private data Sharing Interface" Presented at Theory and Practice of Differential Privacy 2016, New York, June 2016.

2015 Vito D'Orazio, James Honaker and Gary King. "Differential Privacy for Social Science Inference" Presented at Summer Meetings of the Society for Political Methodology, Rochester, July 2015.

2015 Honaker, James. "Efficient Use of Differentially Private Binary Trees." Presented at TPDP 2015 - Theory and Practice of Differential Privacy, London, April 2015.

2014 Honaker, James. "Unemployment and Violence in Northern Ireland: a missing data model for ecological inference." Presented at the Summer Meetings of the Society for Political Methodology, Tallahassee, July 2005.

2014 "The Structure of Autocratic Rule." (With Barbara Geddes, Joseph Wright) Presented at the European Political Science Association Annual Meeting, Barcelona, June 2013. Annual Meetings of the American Political Science Association, Washington D. C., August 2014.

2013 "Sorting Algorithms for Qualitative Data to Recover Latent Dimensions with Crowdsourced Judgments: Measuring State Policies for Welfare Eligibility under TANF" Presented at the Summer Meetings of the Society for Political Methodology, Charlotte, July 2011.

External Funding

- 2020-2021 Sloan Foundation, “OpenDP: An Open-Source Suite of Differential Privacy Tools” (\$750,000 Co-PI with Salil Vadhan, Mercè Crosas, Gary King)
- 2020-2021 DARPA, “CSL: Sotto Voce: Federated Machine Learning for Speech Recognition with Two-sided Privacy” (\$1,000,000 PI with Steve Chong, Prasanna Muthukumar)
- 2019-2020 Sloan Foundation, “OpenDP: An Open-Source Suite of Differential Privacy Tools” (\$884,838 Co-PI with Salil Vadhan, Mercè Crosas, Gary King)
- 2019 US Census Bureau, “Supplement to Formal Privacy Models and Title 13: Enabling private data analysis for the social and economic sciences” (\$398,120 (\$195,000 to Harvard SEAS): author and senior investigator; PIs Kobbi Nissim, Adam Smith and Salil Vadhan)
- 2017-2021 DARPA, “D3M: TwoRavens: Intuitive Statistical Exploration, Model Extraction, and Curation” (\$1,700,000; PI with Co-PI Vito D’Orazio)
- 2016-2020 National Science Foundation, “TWC: Large: Collaborative: Computing Over Distributed Sensitive Data” (\$1,700,000; Co-PI with Steve Chong, Marco Gaboardi, Kobbi Nissim and Salil Vadhan; CNS-1565387)
- 2015-2017 Sloan Foundation, “Applying Theoretical Advances in Privacy to Computational Social Science Practice” (\$868,954 (\$175,000 to IQSS): author and senior investigator; PIs Micah Altman and Salil Vadhan)
- 2012-2015 Minerva Research Initiative, “Regional Dynamics and Regime Disruption in Dictatorships” (\$412,000; Co-PI with Joseph Wright)
- 2011-2014 National Science Foundation, “Multi-Level Policy Responsiveness to Public Opinion” (\$291,210; Co-PI with Eric Plutzer and Michael Berkman; SES-1059723)

Internal Grants

- 2020-2021 Harvard Data Science Initiative, “Ensuring Privacy in COVID-19 Epidemiological Mobility Data Sets” (\$100,000 Co-PI with Salil Vadhan, Satchit Balsari, Caroline Buckee, Mercè Crosas, Gary King)

Invited Talks

- “Dataverse and OpenDP: Tools for Privacy-Protective Analysis in the Cloud” Red Hat Research Day 2020. Webinar and Panel, 2020.
- “OpenDP: a community effort to build open-source tools for differential privacy” Apple Privacy Preserving Machine Learning (PPML) Workshop. Webinar, 2020.
- “TwoRaven and Human-Guided Machine Learning.” Center for Analytics, State Department. Webinar, 2020.
- “OpenDP: An Open-Source Suite of Differential Privacy Tools” Symposium on Data Science and Statistics, American Statistical Association. Webinar, 2020.
- “TwoRavens and Human-Guided Machine Learning for Conflict Event Data” Bureau of Conflict and Stabilization Operations, State Department, Washington DC., 2020.
- “TwoRavens and D3M for Food Insecurity Analysis” Bill and Melinda Gates

Foundation, Seattle WA, 2020.

“Experiences Building PSI” Microsoft, Cambridge MA, 2019.

“Discovery, Replication and Reuse of Sensitive Scientific Data with Differential Privacy” BBN Science Development Program, Cambridge MA, 2019;

“Discovery, Replication and Reuse of Sensitive Scientific Data with PSI” Simons Institute for the Theory of Computing, UC Berkeley, 2019;

“Applied Privacy: Sensitive Data, Attacks and Differential Privacy” Harvard School of Public Health 2018;

“Bootstrap Inference and Differential Privacy: Standard Errors for Free” United States Census Bureau, Washington DC 2018;

“PSI (Ψ): A Private data Sharing Interface” Boston College Law School Program on Innovation and Entrepreneurship; IBM, Cambridge 2018;

“Interactive Data Manipulation with Data Driven Documents (d3) and Javascript” Harvard School of Public Health 2017;

“PSI (Ψ): A Private data Sharing Interface” Privacy Tools for Data Sharing: Lessons Learned and Directions Forward; Harvard University 2017;

“TwoRavens: Intuitive Statistical Exploration, Model Extraction, and Curation” Dataverse Community Meeting, Cambridge 2017;

“Differential Privacy in the Scientific Data Repository” DIMACS/Northeast Big Data Hub Workshop on Overcoming Barriers to Data Sharing including Privacy and Fairness, Rutgers University 2017;

“PSI (Ψ): A Private data Sharing Interface” Dataverse Community Meeting, Harvard Medical School, Boston 2016;

“Introduction to Multiple Imputation for Missing Data Problems” Pew Foundation, Washington DC 2016;

“Multiple Imputation Approaches to Missingness and Mismeasurement in Panel Studies” Food and Drug Administration (FDA), Silver Spring 2015;

“Missing Data” Harvard Graduate School of Education, 2014;

“Sorting Algorithms for Qualitative Data to Recover Latent Dimensions with Crowdsourced Judgements” Applied Statistics Workshop, Harvard 2014;

“The Promise of Computational Methods and Big Data” Hertie School of Governance, Berlin 2013;

“Imputation and Overimputation” MIT Guest Lecture 2013;

“Mechanical Turk for Text Annotation and Judgment Tasks” Quantitative Social Science Initiative, Penn State, 2012;

“Multiple Imputation and Overimputation” Applied Quantitative Methods Workshop, Northwestern University, 2011;

“Unemployment and Violence in Northern Ireland: A Missing Data Model for Ecological Inference” Quantitative Social Science Initiative, Penn State, 2010;

“The Ecology of Preference” Institute for Mathematical Behavioral Science, UC Irvine, 2005.

Teaching Experience

Harvard University:

GRADUATE COURSES

CS208 Applied Privacy for Data Science

(Course overall evaluation 4.8/5; Instructor overall evaluation 4.8/5)

Penn State University:

GRADUATE COURSES

Maximum Likelihood Models
Machine Learning and Robust Model Estimation
Time-Series Cross-Sectional Methods
Mathematical Methods for Political Science

UNDERGRADUATE COURSES

Introduction to Quantitative Analysis
Studies in the Statistical Analysis of Political Data: Political Violence

University of California, Los Angeles:

GRADUATE COURSES

Maximum Likelihood Models
Quantitative Tools for Comparative Politics
Evolutionary Game Theory
Time-Series Cross-Sectional Methods

UNDERGRADUATE COURSES

Introduction to Quantitative Analysis
Studies in the Statistical Analysis of Political Data: Political Violence
Game Theoretic Models in International Relations: Tournaments and Evolution.

American Political Science Association Annual Meeting, Atlanta, 1999.

Short Course Instructor, "What to Do About Missing Data?"

Thesis Committees (and initial position)

Lisa Blaydes (Stanford University)
Vito D'Orazio (The University of Texas at Dallas)
Kim George (Texas A & M)
Julia Gray (University of Pittsburg)
Drew Linzer (Emory University)
Rob Salmond (Michigan University)
Jaekwon Suh (Truman State University)
Stan Hok-Wui Wong (Chinese University of Hong Kong)
James Yonamine (Allstate)

Postdoctoral Fellows

Thomas Brawner (VP and Principal Data Scientist, Terra Crunch)
Vito D'Orazio (Assistant Professor, University of Texas at Dallas)

Research Fellows

Christian Covington (Graduate School, Univ. of Waterloo)

Staff Programmers (since Harvard, and current position)

Michael Shoemate (current)
Raman Prasad (current)
Ethan Cowen (current)

Ira Globus-Harris (Graduate School, U Penn)
Kripanshu Bhargava (Software Developer, Egen Solutions)
Matthew Owen (Software Engineering Manager, Moat)

Research Assistants (since Harvard: Undergraduate and REU Fellows, and current position)

Nabib Ahmed
Andreea Antuca (graduate school, Oxford)
Christian Baehr (Junior Data Analyst, Aiddata, College of William and Mary)
Fanny Chow
Katherine Clayton
Holly Cunningham
Caper Gooden (graduate school, The Johns Hopkins University)
Audrey Haque (current Harvard Masters Student)
Jason Huang (current Harvard Undergrad)
Justin Kaashoek (current Harvard Undergrad)
Hyun Woo Lim (graduate school, Washington University, PS)
Michael Lopiccolo
Wassim Marrakchi (current Harvard Undergrad)
Jack Murtagh (graduate school, Harvard CS)
Dan Muise (graduate school, Stanford)
Michael Lackner
Lipika Ramaswamy (Leapyear Technologies)
Grace Rehaut (Analyst, Goldin Associates LLC)
Kathryn Taylor
Julia Vasile (graduate school, Applied Math and Statistics, Stonybrook)
Clara Wang (graduate school, Yenching Scholar at Peking University)
Lancelot Watthieu (Software Engineer, NanoSemi Inc)