Persuasion at Scale: Machine Learning, Causality and the Information Ecosystem

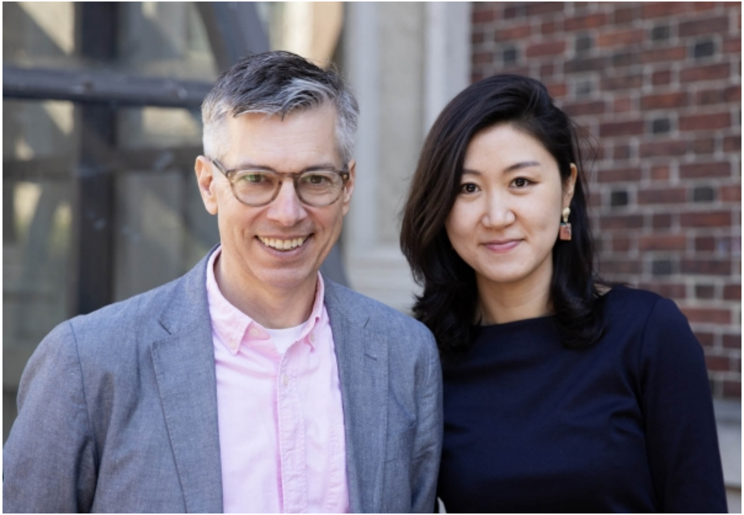
Lessons Learned from teaching at the edge of a new field

Chris Wiggins; work with Eunji Kim

Columbia University; course site = bit.ly/PScaleClass

2025-06-26

Instructors: Wiggins & Kim



Outline

Field Building: The Mathematics of Interpreting the World and Changing It

Separate fields, growing closer daily

- **Political Science/Social Science**: Builds 'theories' of how individuals and groups behave
- **Causal Inference**: Measures what actually causes behavioral change
- **Marketing**: Optimizes content to drive consumer behavior, esp in industry and with abundant resources
- **Media Studies**: Analyzes how content affects audiences and how this changes over time
- **Machine Learning**: Builds algorithmic systems for optimization and personalization
- **GenAI**: (esp. LLM): Generates content

The Problem: Overlapping phenomena, same data, same arena, same ethical stakes—but taught in isolation

Student Reality

- Students read the news, follow politics, use social media
- Just as polarized as broader society, no unified worldview
- Their questions:
  - What should I believe about media influence?
  - Are algorithms really manipulating them?
  - How does persuasion actually work?
- **Our goal**: Students approach these questions *without hype or fear*

To be understood, not feared: Think rigorously about a concerning topic

How can all of these be true simultaneously?

- **Optimization Methods Work**: A/B testing, bandits, ML improve metrics, yet:
- **Marketing Effects Are Finite**: Real persuasion smaller than expected, and
- **Inference Is Finite**: Even perfect data has attribution limits, yet
- **Detection Can Be Rigorous**: Strong mathematical and statistical methods exist, yet:
- **'Has Math' vs. 'Objective'**: The math doesn't eliminate subjective choices, just hides them

Observation v. Intervention; Understanding v. Optimization

‘The philosophers have only interpreted the world, in various ways; the point is to change it.‘

In curricula, methods for understanding persuasion and building persuasion systems are different subjects. (Causal inference and ML)

They’re happening in same information ecosystem.

Persistent Theme of Academic-Industry Complememtary

Academic Side:

- Rigorous theory
- Careful identification
- External validity concerns
- Publication incentives

Industry Side:

- Massive scale data and resources
- Sales not rigorous theory
- Business constraints
- Optimization focus

Overlapping mathematical foundations, overlapping methodological challenges

View from publishing/journalism

CI and ML used to

- **Understand** reader behavior (academic mode)
- **Optimize** subscription flows (industry mode)
- **Measure** content effectiveness (research mode)
- **Build** recommendation systems (engineering mode)

Teaching (and research) opportunity

see this as one integrated framework, not separate domains, certainly the mathematics of persuasion and optimization has overlap

What We Taught: Causality, ML, and AI in Information Operations

The Curriculum Arc

Five-part journey from measurement to AI-generated persuasion

Part 1: How Do We Quantify a Messy World?

The Universal Challenge

Converting human behavior into analyzable variables

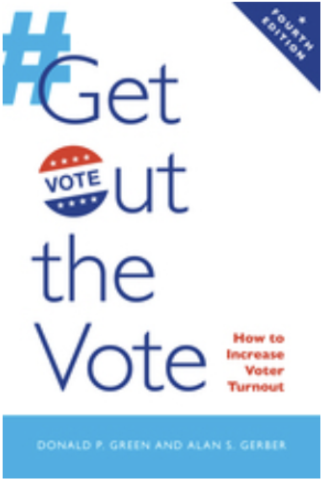
- **Political Science**: "Coding" political persuasion from survey data
- **Tech Practice**: "engagement" as catchall for user behavior that can be made into SQL
- **ML Systems**: Requires complex reality to be turned into abundant features
- **throughout**: bias and limitations of interpretation

Part 2: Causality and Experiments

The Gold Standard

Randomized controlled trials—sandbox and field experiments

- **RCTs**: Industry (1908-present) and academic view $1/\sqrt{N}$ reality
- **Tech Practice** and A/B Testing: Continuous experimentation, attribution complexity
- **The eBay example**: \$50+ million in "effective" search ads showed minimal causal impact (Lewis and Rao 2023)



Get Out the Vote: How to Increase Voter Turnout

Donald P. Green

Alan S. Gerber

Copyright Date: 2019

Edition: 4

Published by: Brookings Institution Press

Pages: 253

Part 3: When Experiments Aren't Possible

The Causal Inference Toolkit

IV, DiD, RDD, and using time as "cause"—but also the surprising persistence of OLS

- **Instructor Surprise**: How prevalent OLS still is in practice
- **Student Reading**: Krueger & Lewis-Beck's "Is OLS Dead?"
- **Reality Check**: Even with sophisticated causal methods, simple regression remains ubiquitous
- **The Pragmatic Truth**: Sometimes OLS is what you can actually implement

Imperfect compliance

Code/data examples, old and new

colab.research.google.com/drive/1RM_irpmRb80pn5KOJf4Y5eSVE7dWMwNv

Angrist 1990

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all

```r
library(haven)    # for reading Stata .dta files
```

```r
# -------------------------------
# First Stage: dmdcdat.dta
# -------------------------------
# Read the first-stage dataset directly from the URL.
dmdc_url <- "https://economics.mit.edu/sites/default/files/inline-files/dmdcdat.dta"
dmdc <- read_dta(dmdc_url)

# Convert to a data frame (if needed)
dmdc <- as.data.frame(dmdc)
```

```r
# -------------------------------
# Reduced Form: cwhsc_new.dta
# -------------------------------
# Read the reduced form dataset directly from the URL.
cwhsc_url <- "https://economics.mit.edu/sites/default/files/inline-files/cwhsc_new.dta"
cwhsc <- read_dta(cwhsc_url)
cwhsc <- as.data.frame(cwhsc)
```

Explanation:

- **Filtering the Data:** The code filters the dataset to include only white individuals (`race == 1`) and those in intervals 1 and 3. In Angrist's 1990 study, he focused on white males to assess the impact of Vietnam-era military service on lifetime earnings, using the draft lottery as an instrumental variable.
- **Grouping and Summarizing:** By grouping the data by birth year and treatment status, and then calculating the mean probability of service (`ps_r`), the code mirrors the first stage of Angrist's two-stage least squares (2SLS) estimation. This stage examines the relationship between the instrumental variable (draft eligibility) and the endogenous explanatory variable (veteran status).
- **Calculating Differences (gamma1):** The difference in mean service probabilities between treated and control groups (`diff = int_1 - int_3`) represents the causal effect of draft eligibility on the likelihood of serving in the military. In Angrist's framework, this is denoted as gamma1 and is essential for estimating the causal impact of veteran status on earnings.

Article

https://doi.org/10.1038/s41467-022-35576-9

Exposure to the Russian Internet Research Agency foreign influence campaign on Twitter in the 2016 US election and its relationship to attitudes and voting behavior

Received: 11 March 2021

Accepted: 12 December 2022

Gregory Eady[1], Tom Paskhalis[2], Jan Zilinsky[3], Richard Bonneau[4], Jonathan Nagler[5] & Joshua A. Tucker[5]

Code/data examples, old and new

Received: 11 March 2021

Accepted: 12 December 2022

Published online: 09 January 2023

Check for updates

Gregory Eady [1], **Tom Paskhalis** [2], **Jan Zilinsky** [3], **Richard Bonneau** [4], **Jonathan Nagler** [5] & **Joshua A. Tucker** [5]

There is widespread concern that foreign actors are using social media to interfere in elections worldwide. Yet data have been unavailable to investigate links between exposure to foreign influence campaigns and political behavior. Using longitudinal survey data from US respondents linked to their Twitter feeds, we quantify the relationship between exposure to the Russian foreign influence campaign and attitudes and voting behavior in the 2016 US election. We demonstrate, first, that exposure to Russian disinformation accounts was heavily concentrated: only 1% of users accounted for 70% of exposures. Second, exposure was concentrated among users who strongly identified as Republicans. Third, exposure to the Russian influence campaign was eclipsed by content from domestic news media and politicians. Finally, we find no evidence of a meaningful relationship between exposure to the Russian foreign influence campaign and changes in attitudes, polarization, or voting behavior. The results have implications for understanding the limits of election interference campaigns on social media.

Caveat: "CI doesn't work" – Facebook, 2018

A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook*

Brett R. Gordon
Kellogg School of Management
Northwestern University

Florian Zettelmeyer
Kellogg School of Management
Northwestern University and NBER

Neha Bhargava
Facebook

Dan Chapsky
Facebook

April 12, 2018

Caveat: "CI doesn't work" – Facebook, 2018

Abstract

Measuring the causal effects of digital advertising remains challenging despite the availability of granular data. Unobservable factors make exposure endogenous, and advertising's effect on outcomes tends to be small. In principle, these concerns could be addressed using randomized controlled trials (RCTs). In practice, few online ad campaigns rely on RCTs, and instead use observational methods to estimate ad effects. We assess empirically whether the variation in data typically available in the advertising industry enables observational methods to recover the causal effects of online advertising. This analysis is of particular interest because of recent, large improvements in observational methods for causal inference (Imbens and Rubin 2015). Using data from 15 US advertising experiments at Facebook comprising 500 million user-experiment observations and 1.6 billion ad impressions, we contrast the experimental results to those obtained from multiple observational models. The observational methods often fail to produce the same effects as the randomized experiments, even after conditioning on extensive demographic and behavioral variables. We also characterize the incremental explanatory power our data would require to enable observational methods to successfully measure advertising effects. Our findings suggest that commonly used observational approaches based on the data usually available in the industry often fail to accurately measure the true effect of advertising.

Keywords: Digital Advertising, Field Experiments, Causal Inference, Observational Methods, Advertising Measurement.

Part 4: From Understanding to Optimizing

The Progression

A/B testing $\rightarrow$ bandits $\rightarrow$ contextual bandits $\rightarrow$ recommendation systems

- **Exploration vs. Exploitation**: Balancing learning and optimization
- **Multi-Armed Bandits**: Dynamic allocation based on outcomes
- **Contextual Bandits**: Personalized treatment assignment

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Adaptive Experimental Design: Prospects and Applications in Political Science

Molly Offer-Westort Stanford Graduate School of Business
Alexander Coppock Yale University
Donald P. Green Columbia University

Abstract: *Experimental researchers in political science frequently face the problem of inferring which of several treatment arms is most effective. They may also seek to estimate mean outcomes under that arm, construct confidence intervals, and test hypotheses. Ordinarily, multiarm trials conducted using static designs assign participants to each arm with fixed probabilities. However, a growing statistical literature suggests that adaptive experimental designs that dynamically allocate larger assignment probabilities to more promising treatments are better equipped to discover the best performing arm. Using simulations and empirical applications, we explore the conditions under which such designs hasten the discovery of superior treatments and improve the precision with which their effects are estimated. Recognizing that many scholars seek to assess performance relative to a control condition, we also develop and implement a novel adaptive algorithm that seeks to maximize the precision with which the largest treatment effect is estimated .*

Tech Co. Optimization in Practice

- **Product and Business goals**: From static A/B to contextual bandits
- **Recsys**: Personalized content delivery based on user behavior
- **Governance Implications**: Governance both as ‘coding‘ and as policy

KPI mindset in industry centers A/B testing and optimization

Part 5: How AI Changes Everything

Technical Foundations

ANNs $\rightarrow$ Transformers $\rightarrow$ LLMs $\rightarrow$ Generative AI

- **Explaining ANN, GenAI**: historical and visual overview
- **New Persuasion Mechanisms**: AI-generated content persuades differently
- **Measurement Challenges**: AI $\rightarrow$ content $\rightarrow$ user response (multiple endogenous steps)
- **Recent literature**: Reddit and other recent headlines and articles

Part 5: How AI Changes Everything

Home > Tech

Reddit threatens legal action against AI researchers for 'highly unethical' experiment

Researchers with the University of Zurich used AI bots to generate thousands of comments in r/changemyview.

By Chance Townsend on April 29, 2025

Homework 1: YouTube Watch History Reflection

The Self-Report vs. Behavioral Data Reality Check

Students download their actual YouTube watch history and compare to their self-reported usage

- **Key Discovery**: Systematic discrepancies between stated and revealed preferences
- **Political Focus**: How accurate are self-reports about news/politics consumption?
- **Methods Learning**: Neither self-reports nor digital exhaust are gold standards
- **Broader Implications**: What this means for media studies research methodology

Assignment 1 reveals that asking people directly is NOT the gold standard—but neither are digital traces

Homework 2: Hurricane Blandy and Instrumental Variables

Natural Experiment in Social Media Persuasion

Students use Hurricane Blandy's power outage as instrument for social media usage effects on voting

- **Fictional Scenario**: Hurricane cuts power south of 39th Street, disrupts "Snapface" usage
- **IV Strategy**: Power outage as exogenous shock to social media engagement
- **Methods**: Two-Stage Least Squares (2SLS) implementation in practice
- **Angrist Connection**: Same logic as draft lottery → military service → earnings

Students learn to find causality when randomization happens TO them, not BY them

Homework 3: BlueSky Linguistic Analysis - The Unique Assignment

Real-Time Language Analysis Around Events

Students analyze how language use changes around significant events using actual BlueSky data

- **Data Access**: Complete BlueSky firehose via collaboration with Texas Bluesky Archive
- **Big Query Integration**: All of BlueSky behind Big Query—students know who said what to whom
- **Student Choice**: Pick their own event (election, Signalgate, etc.) and regex patterns
- **Methods**: Difference-in-means vs. Regression Discontinuity Design comparison
- **Real Stakes**: Analyzing actual political discourse as it happens

Unique opportunity: Students get unprecedented access to real social media discourse data

Tools: Slack as Weekly Discussion Engine

Slack: The Continuous Conversation

Every week, students posted comments and questions in Slack channels—which directly informed lectures

- **Live Feedback Loop**: Student questions and observations sourced weekly lecture content
- **Cross-Assignment Discussion**: Students debated homework findings and methodology
- **Real-Time Clarification**: Immediate responses to conceptual confusion
- **Peer Learning**: Students helping each other with technical and theoretical challenges
- **Instructor Responsiveness**: Lectures adapted based on what students were actually struggling with

Required learning new tool, but will be useful post-Columbia

Tools: Google Colab for Hands-On Learning

Colab: Democratizing Computational Analysis

Heavy use of Google Colab both in class and for homework assignments—despite having a completely non-computational final exam

- **In-Class Analysis**: Live coding during lectures to explore concepts immediately
- **Homework Implementation**: All three assignments required substantial Colab work
- **No Setup Barriers**: Students could run complex analyses without local software installation
- **Shared Notebooks**: Easy collaboration and sharing of analytical approaches
- **Real Data Access**: BlueSky, YouTube, Angrist and other papers, synthetic datasets

Students learned by doing—analyzing real data with production-quality tools

The Reproducibility Crisis: A Persistent Theme

Attempting to Reproduce Papers: Mostly Failures

A persistent theme throughout the class was my attempts to reproduce any paper—most failed or had to hide the most interesting data

- **Typical Experience**: Papers with missing data, broken links, proprietary datasets
- **Hidden Complexity**: Code that doesn't run, unclear preprocessing steps
- **Data Unavailability**: Most interesting analyses used data students couldn't access
- **Notable Exception**: Angrist (1990)—could get the data, transform from Stata to Python, put it on Colab, and reproduce results to within 10% error
- **Student Lesson**: Reproducibility is harder than it should be, even for canonical papers

Teaching reproducibility by experiencing its challenges firsthand

The Pedagogical Paradox: 20th Century Final Exam

Blue Book Exam: No Code, Pure Conceptual Understanding

Despite heavy computational focus, final exam was traditional blue book format

Throughout Course:

- Complex Colab analyses
- Real data manipulation
- ML/Statistical programming

Final Exam:

- Pen and paper only
- Conceptual questions
- No computational aids

Lessons Learned: Teaching Truth-Making in an Algorithmic Age

A New Field Emerging?

Separate Fields, Converging Reality

- Political Science/Social Science: Builds 'theories' of how individuals and groups behave
- Causal Inference: Measures what actually causes behavioral change
- Marketing: Optimizes content to drive consumer behavior
- Media Studies: Analyzes how content affects audiences
- Machine Learning: Builds algorithmic systems for optimization and personalization
- GenAI: Generates content at scale

Meeting Students Where They Are

Addressing Real Concerns Directly

- Career anxiety: "Will AI replace me or can I use it?"
- Ethical confusion: "Are these systems manipulating me?"
- Professional preparation: "How do I work with/against these systems?"
- Information overwhelm: "What should I actually believe?"

The Approach: Approach: honest, contemporary, situated, dynamic, and critical

Honest Pedagogy for Contested Terrain

Teaching in the Messiness

- Reproducibility crisis as learning opportunity, not failure
- Academic vs. industry perspectives as complementary, not competing
- Student polarization as starting point, not obstacle
- Uncertainty as pedagogical resource, not problem

The Broader Stakes

Why This Matters Now

- Information ecosystem is the new civic infrastructure
- Students will build, regulate, or report on these systems
- Democracy needs citizens who understand persuasion architectures
- Collective, good-faith sensemaking is important; happens across multiple ‘professions‘

Collective good faith senemaking

Beyond Methods - Teaching Sense-Making

- How facts become truth through research, literature-consensus, and professional norms
- How to navigate contested topics when the stakes are real
- Claim: students can handle complexity if we're realistic and honest about
  - both the benefits and the limits of rigor and
  - the interests of participating actors in sensemaking

Thank you!

Questions and Discussion