UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Essays on Psychology, Politics, and Economics

Permalink

https://escholarship.org/uc/item/53205781

Author

Kim, Woojin

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Essays on Psychology, Politics, and Economics

by

Woojin Kim

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Stefano DellaVigna, Chair

Professor Ned Augenblick

Professor Dmitry Taubinsky

Spring 2024

Abstract

Essays on Psychology, Politics, and Economics

by

Woojin Kim

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Stefano DellaVigna, Chair

This dissertation incorporates two major themes from behavioral economics and political economy. The first theme focuses on how social forces—such as culture, norms, and identity—shape economically meaningful behavior. This work builds on key theories from psychology and economics that embody intuitive ideas: we compare what we have relative to our peers or expectations, and are constrained by social norms yet impose them on others as well. The first two chapters explore these ideas in historically or politically consequential contexts. They evaluate the historic role of commercial television in setting consumer expectations and study the motives behind pressuring others to follow norms on political civic duty.

The second theme in this dissertation moves onto governments as decision-makers and studies the determinants of policy adoption across the United States. The first chapter within this theme describes the general dynamic patterns of policy diffusion for a broad set of U.S. state-level policies. The second then focuses on municipal governments in their adoption of non-partisan, low-cost "nudge" interventions.

To my Father

Contents

1 Introduction **1**

2 Television and American consumerism **3**
2.1 Introduction . . . 4
2.2 Historical background . . . 7
2.3 Data . . . 8
2.4 Empirical strategy . . . 10
2.5 Results . . . 14
2.6 Discussion and conclusion . . . 35

3 "Don't you plan on voting?" The motives and effects of peer pressure in voter mobilization **37**
3.1 Introduction . . . 38
3.2 Experimental design and sample . . . 41
3.3 Results . . . 46
3.4 Discussion . . . 58

4 Policy diffusion and polarization across U.S. states **60**
4.1 Introduction . . . 61
4.2 Case study: Medicaid and Food Stamp Program . . . 64
4.3 Data and summary statistics . . . 65
4.4 Evidence on innovation and diffusion . . . 66
4.5 Evidence relating to models of policy diffusion . . . 72
4.6 Discussion and conclusion . . . 75
4.7 Figures and tables . . . 77

5 Bottlenecks for evidence adoption **95**
5.1 Introduction . . . 96
5.2 Setting and data . . . 99
5.3 Framework . . . 103
5.4 Results . . . 104
5.5 Interpretation and implications . . . 109
5.6 Generalizability of results . . . 112
5.7 Discussion and conclusion . . . 113
5.8 Figures and tables . . . 116

6 Bibliography **134**

A Appendix for Television and American consumerism **148**
A.1 Business categories in the *Census of Business* . . . 148
A.2 Construction Permit-holders vs. Pending Applicants . . . 149
A.3 Appendix figures and tables . . . 151

B Appendix for "Don't you plan on voting?" The motives and effects of peer pressure in voter mobilization **177**
B.1 Additional survey details . . . 177
B.2 Pre-analysis plan details . . . 179
B.3 Appendix figures and tables . . . 181

C Appendix for Policy diffusion and polarization across U.S. states **198**
C.1 Alternate measures of correlated adoptions . . . 198
C.2 Estimating voter preferences using the ANES, GSS, and measures from political science . . . 199
C.3 Appendix figures and tables . . . 204

D Appendix for Bottlenecks for evidence adoption **234**
D.1 Example of BIT trial report . . . 234
D.2 Online forms and communications on city websites . . . 239
D.3 Cosine similarity in BIT trial communications and out-of-sample online forms 241
D.4 Marginal cases of non-adoption . . . 242
D.5 Broad cases of adoption . . . 243
D.6 Forecasting survey . . . 244
D.7 Categorizing trials combining new and pre-existing communications . . . 245
D.8 Appendix figures and tables . . . 246

List of Figures

2.1 Distribution of first local TV station broadcasting date 11
2.2 Differences in retail trends between the Pre- and Post-Freeze groups 18
2.3 1948 expenditure shares and 1948-54 sales growth by retail sector 22
2.4 TV advertising revenue and effect on retail sales 28
2.5 Access to TV and highway construction 33
2.6 Access to TV and birthrate 35
3.1 Experimental design 42
3.2 Message template and guidelines 44
3.3 Willingness-to-pay (WTP) to send Pressure Message 47
3.4 Senders' WTP and beliefs for the Pressure Message 49
3.5 Effect of anonymity by beliefs on the affective effect 52
3.6 Comparison between Pressure and Info Message conditions 55
3.7 Intend-to-treat effects of receiving peer pressure on electoral participation . . 57
4.1 Three policy examples 78
4.2 Case studies of welfare programs 79
4.3 Policy sources 80
4.4 Innovating states 81
4.5 Correlation in geography and politics among adopters (random and observed) 82
4.6 Dynamics of policy diffusion dimensions 83
4.7 Simulated policy diffusion 84
4.8 Event study from switches in state government party control 85
4.9 Comparison to polarization in DW-NOMINATE 86
4.10 Policy-by-policy diffusion patterns 87
5.1 Study design and sample restrictions 117
5.2 Trial-by-trial adoption and effect sizes 118
5.3 Adoption of nudges: Observed compared to benchmarks 119
5.4 Word cloud from open-ended forecasts of adoption determinants 120
5.5 Adoption of nudges by effectiveness 121
5.6 Adoption based on city context 122
5.7 Adoption based on experimental design 123
5.8 Mechanisms behind the effect of pre-existence 124
5.9 Pre-existence and evidence based adoption 125
5.10 Survey evidence on organizational inertia 126
5.11 Counterfactual adoption rules 127
5.12 Hjort et al. (2021) policy adoption experiment: Letter and nudge adoption in treatment and control groups 128
A.1 *Census of Business* record example 152
A.2 Merchandise Line Sales data from 1948 *Census of Business* 153
A.3 Geographic distribution of counties by TV group 154
A.4 Counties from Baum-Snow (2007)'s highway dataset 155
A.5 Distributions of demographics for Pre- and Post-Freeze groups 156
A.6 TV ownership by year 157

A.7 Robustness checks . . . 158
A.8 Distribution of topographical features . . . 159
A.9 Status map of main broadcasting companies and affiliate stations . . . 160
A.10 Leading TV advertisers 1951-54 . . . 161
A.11 TV advertising revenue data from *Advertising Age* . . . 162
A.12 TV access and highway construction: TWFE and problematic weights . . . 163
A.12 Event study of TV access and highway construction (additional specifications) 164
A.13 *Televsion Digest* example excerpts . . . 165
A.14 Distribution of first local TV station broadcasting date (with CP and PA) . 166
B.1 Non-registrants' willingness to share unregistered status . . . 178
B.2 Within-Sender assignment of Direct and Anonymous Messages . . . 182
B.3 Eliciting the Sender's beliefs on the affective effect (Recipient's (dis)like of the message) . . . 183
B.4 Eliciting the Sender's beliefs on the instrumental effect (likelihood of persuading the Recipient) . . . 184
B.5a Eliciting the Sender's willingness-to-pay to send the message . . . 185
B.5b Eliciting the Sender's willingness-to-pay to send the message . . . 186
B.6 Sender-level WTP distributions for Pressure Message . . . 187
B.7 Distribution of Senders' beliefs on the instrumental effect . . . 188
B.8 Intend-to-treat effects of receiving peer pressure on electoral participation (excluding "pre-registered") . . . 189
C.1 Summary statistics: Policy area composition and adoption speed . . . 204
C.2 Innovating states . . . 205
C.3 Correlation in geography and politics among adopters (alternate thresholds) 206
C.4 Stability of closest thirds in each dimension . . . 207
C.5 Simulated policy diffusion . . . 208
C.5 Simulated policy diffusion . . . 209
C.5 Simulated policy diffusion . . . 210
C.6 Robustness checks: Threshold of closest states . . . 211
C.7 Measuring voter policy preference from ANES and GSS . . . 212
C.8 Correlation of policy outcomes: 1980-85 vs. 2005-10 . . . 214
C.9 Categorizing the ideology of policies for event study analysis . . . 215
C.10 Event study from switches in state government party control . . . 216
C.10 Event study from switches in state government party control (de Chaisemartin and D'Haultfoeuille estimator) . . . 217
C.11 Specification checks: Range of measure by number of adopters . . . 218
C.11 Specification checks: Stability in coefficients by number of adopters . . . 219
D.1 Example of adoption of BIT-NA trial . . . 247
D.1 Example of adoption of BIT-NA trial . . . 248
D.2 Availability and updating of public forms/communications on city websites . 249
D.3 Adoption by response times (bin scatters) . . . 250
D.4 BIT recommendations for adoption by statistical significance . . . 251
D.5 Adoption of nudges by pre-existence: Additional results . . . 252
D.6 Receiver operating characteristic (ROC) curves . . . 253
D.7 Interaction effects: Staff retention . . . 254

D.7 Interaction effects: Control take-up 255
D.8a Survey evidence on organizational inertia: Non-adopters vs. adopters 256
D.8b Survey evidence on organizational inertia: Word cloud 257
D.9 Hjort et al. (2021) policy adoption experiment: Policy brief for tax payment reminder letter (reproduced) 258
D.9 Hjort et al. (2021) policy adoption experiment: Template for tax payment reminder letter (reproduced) 259
D.10 Comparisons of forecasts and observed adoption 260

List of Tables

2.1 Summary statistics by TV group . . . 12
2.2 Selection of Pre- vs. Post-Freeze groups on retail sales before the FCC Freeze 13
2.3 Relative retail trends . . . 17
2.4 Heterogeneity of the TV effect on log retail sales growth over 1948-54 . . . 19
2.5 Decomposition of difference in retail sales growth over 1948-54 between Pre- and Post-Freeze groups . . . 24
2.6 Summary statistics on sales, advertising, and visibility by product . . . 27
2.7 Advertising, durables, and the TV effect (1948-54) . . . 31
3.1 Senders' WTP and beliefs across experimental conditions . . . 51
3.2 Senders' WTP and beliefs for the Pressure Message: Heterogeneity . . . 54
4.1a Summary of NBER data set . . . 88
4.1b Summary statistics of policy data sets . . . 88
4.1c Policy areas . . . 89
4.2 Highest and lowest innovators (20%) . . . 90
4.3 Policy diffusion predictors by decade . . . 91
4.4 Heterogeneity in policy diffusion . . . 92
4.5 Models of policy diffusion: Role of migration and voter preferences . . . 93
4.6 Vaccine regulations and COVID-19 policies . . . 94
5.1 Summary of papers on adoption of evidence . . . 129
5.2 Sample characteristics . . . 130
5.3 Determinants of nudge adoptions . . . 131
5.4 Comparison of specific nudge adoption and broad adoption . . . 132
5.5 Hjort et al. (2021) policy adoption experiment: Persuasion rates . . . 133
A.1 Summary statistics for other groups . . . 167
A.2 Access to TV and retail sales (additional specifications) . . . 168
A.3 Pre-trends in median occupational income score or family income . . . 169
A.4a Topography summary statistics . . . 170
A.4b TV access, topography, and bordering counties . . . 170
A.5a Top 10 evening series for 1950-54 seasons by audience size (Nielsen ratings) . 171
A.5b Top 10 evening series for 1954-58 seasons by audience size (Nielsen ratings) . 172
A.6 Raw decomposition of difference in retail sales growth over 1948-54 between Pre- and Post-Freeze groups . . . 173
A.7 TV access year and highway miles in 1947 plan . . . 174
A.8 Access to TV and birthrate . . . 175
A.9 Selection of Construction Permit-holders vs. Pending Applicants on retail sales before the FCC Freeze . . . 176
B.1 Pre-analysis plan specification and results . . . 180
B.2a Balance: Senders . . . 190
B.2b Balance: Non-registered Recipients . . . 191
B.2c Balance: Recipient relation to assigned Sender . . . 192
B.2d Sample attrition . . . 192
B.3 Sample characteristics . . . 193

B.4a Senders' WTP and beliefs across experimental conditions: Alternate specifications and robustness . . . 194
B.4b Senders' WTP and beliefs across experimental conditions (uncensored WTP) 195
B.5 Effect of receiving a Direct Pressure Message on rates of electoral participation 196
B.6 Non-registered Recipients: Experimental outcomes . . . 197
C.1 Summary of recent papers on policy diffusion in political science and examples in economics . . . 220
C.2a SPID sample examples . . . 221
C.2b NBER working paper sample . . . 222
C.2b NBER working paper sample . . . 223
C.3a Summary statistics: Policy outcomes from NBER papers . . . 224
C.3b COVID-19 policies . . . 225
C.3c Vaccine regulations . . . 226
C.4 Highest and lowest innovators of NBER policies (20%) . . . 227
C.5a Examples of states in closest thirds . . . 228
C.5b Examples of states in closest thirds of voter preferences (ANES & GSS) . . . 229
C.6 Robustness checks . . . 230
C.7a Models of policy diffusion: Role of migration and voter preferences (expanded demographics and public opinion) . . . 231
C.7b Models of policy diffusion: Role of migration and voter preferences (expanded ANES & GSS voter preferences) . . . 232
C.8 Event studies . . . 233
D.1 City features of completed vs. abandoned trials . . . 261
D.2 Average nudge treatment effects . . . 262
D.3 Sample characteristics: City features . . . 263
D.4 Forecasts summary . . . 264
D.5 BIT recommendations for trial adoption . . . 265
D.6 Determinants of nudge adoption (robustness) . . . 266
D.7 Sample characteristics: Survey of non-adopters . . . 267
D.8a Hjort et al. (2021) policy adoption experiment: Control and treatment rates by adoption definition . . . 268
D.8b Hjort et al. (2021) policy adoption experiment: Balance in letter adoption (control group) . . . 268

Acknowledgments

For Stefano DellaVigna, the title of "advisor" is far too limited. He embodies what it means to be not just a good economist and mentor, but also a good person. I can only strive to do for others at least half as much as he has done for me.

I learnt almost everything I know about doing research thanks to those at Berkeley. There I met Elizabeth Linos, whose inspiring creativity and congeniality makes working together feel more like leisure. Ned Augenblick, Ricardo Perez-Truglia, and Dmitry Taubinsky were there every week to refine our bad ideas and give us precious hope that something good could come out of them.

I have been blessed with lifelong friends along the way. To name a few, I depended on Eric Koepcke, Maxi Lauletta, Pat Kennedy, and Ross Chu for commiseration, encouragement, and necessary distractions. And with those at Church Without Walls who adopted me, Tammy, and Kiwi into their community, we did life together.

Finally, to my parents for their comforting love, my brother in whose footsteps I still find myself inexplicably following, and to Tammy who is my universe and my home: All this and anything else I accomplish I owe to you.

Chapter 1

Introduction

The first theme of this dissertation focuses on how social forces—such as culture, norms, and identity—shape economically meaningful behavior. This work builds on key theories from behavioral economics that embody intuitive ideas: we compare what we have relative to our peers or expectations, and are constrained by social norms yet impose them on others as well. The next two chapters explore these ideas in historically or politically consequential contexts. The first chapter in this area evaluates the historic role of commercial television in setting consumer expectations. The second studies the motives behind pressuring others to follow norms on political civic duty.

In Chapter 2, I investigate how the introduction of mass commercial television in the postwar era changed American consumer behavior. Media scholars and U.S. historians claim that TV with its unprecedented advertising appeal drew Americans into a culture of upscaling and purchasing products for social status. I test this prevailing theory using newly digitized nationwide county-level retail sales data from the Census of Business series. I compare growth in retail sales between areas with and without local TV service over the unanticipated Federal Communications Commission (FCC) Freeze, which halted the licensing of new TV stations from 1948–52.

I find three results that corroborate TV's long-attributed role in American consumerism. First, during the Freeze, total retail sales in counties with TV access increased by 3–4% more on average than in counties without access. Second, the effect of TV was concentrated in the automobile sector, which alone accounted for a third of the total difference. Third, TV advertising led to higher growth in sales, but only for durable goods, which neatly aligns with both qualitative reports and theories of conspicuous consumption. Historians also propose that the suburban family life popularized in TV sitcoms became the mainstream representation of the American Dream in the 1950s. Consistent with these accounts, I find that the start of TV access coincides with greater activity in local highway construction and birthrates.

Chapter 3 is motivated by the growing role of online peer pressure as a potential tool to mobilize young American voters. It remains unclear, however, whether these voters are willing to pressure their peers and how effective it would be.

In this online field experiment, college students choose whether to send an email that pressures their campus peers to register to vote for the 2020 U.S. Elections. The senders have a negative average willingness-to-pay (WTP) to send the email, meaning they would rather pay than pressure their peers. The experiment reveals two counteracting motives at play. On the positive side, senders believe that pressuring their peers will persuade them to register. On the negative side, they believe that their peers will dislike being pressured, which outweighs the positive motives. However, under anonymous messaging, they become much less sensitive to whether the recipient dislikes being pressured, and their average WTP

turns positive. This pattern of behavior reflects voters who are motivated by the prospect of mobilizing their peers, but are restrained by self-interested concerns from social image costs or retaliation.

While these prior chapters aim to better understand individual decision-making, the next two move onto governments as decision-makers and study the determinants of policy adoption. They do so for two distinct types of policies and governments. The first describes the general dynamic patterns of policy diffusion for a broad set of U.S. state-level policies, spanning an array of topics from marijuana legalization to wildlife conservation. The second examines the adoption of non-partisan, low-cost "nudge" interventions in municipal governments.

Chapter 4 (co-authored with Stefano DellaVigna) presents a series of facts based on a data set of over 700 U.S. state policies spanning the past seven decades. First, for the introduction of new laws, state capacity seems to have a small role, in that larger and richer states are only slightly more likely to innovate policy. Second, the diffusion of policies from 1950 to 2000 is best predicted by proximity—a state is more likely to adopt a policy if nearby states have already done so—as well as similarity in demographics and voter policy preferences. Third, since 2000, political alignment is the strongest predictor of diffusion. Models of learning and correlated preferences can account for the earlier patterns, but the findings for the last two decades indicate a sharply increasing role of party control. We conclude that party polarization has emerged as a key factor recently for policy adoption, plausibly leading to a worse match between state policies and voter preferences.

Chapter 5 (co-authored with Stefano DellaVigna and Elizabeth Linos) is prompted by the increasing use of randomized controlled trials (RCTs) within governments to test innovations. There has been little research, however, about how governments incorporate the results from these experiments into policy-making. We follow up with 30 U.S. cities than collectively ran 73 randomized control trials (RCTs) with a national Nudge Unit. Although 78% of the nudge interventions found positive effects in the trials, the cities adopted just 27% of them into their ongoing communications. The effectiveness of the nudge does not predict adoption. Instead, the largest predictor is whether the nudge was implemented using pre-existing communication, as opposed to new communication. We identify organizational inertia as a leading explanation: changes are more likely to survive when incorporated into pre-existing infrastructure.

Chapter 2

Television and American consumerism

This chapter studies how the introduction of mass commercial television in the postwar era changed American consumer behavior. Media scholars and U.S. historians claim that TV with its unprecedented advertising appeal drew Americans into a culture of upscaling and purchasing products for social status. I test this prevailing theory using newly digitized nationwide county-level retail sales data from the *Census of Business* series. I compare growth in retail sales between areas with and without local TV service over the unanticipated Federal Communications Commission (FCC) Freeze, which halted the licensing of new TV stations from 1948-52. I find three results that corroborate TV's long-attributed role in American consumerism. First, during the Freeze, total retail sales in counties with TV access increased by 3-4% more on average than in counties without access. Second, the effect of TV was concentrated in the automobile sector, which alone accounted for a third of the total difference. Third, TV advertising led to higher growth in sales, but only for durable goods, which neatly aligns with both qualitative reports and theories of conspicuous consumption. Historians also propose that the suburban family life popularized in TV sitcoms became the mainstream representation of the American Dream in the 1950s. Consistent with these accounts, I find that the start of TV access coincides with greater activity in local highway construction and birthrates.

2.1 Introduction

Americans in the 1950s faced two revolutions in daily life: they had money to spare, and television to watch. Around a decade later, Galbraith (1958) in his enduring work *The Affluent Society* lamented the consumerist society America had become, where possessions represented social status and "one man's consumption becomes his neighbor's wish." He held advertisers responsible for turning the discretionary income of the common American into insatiable consumer demand for an ever-expanding line of goods. Historians and media scholars claim that the new medium of television, with its unprecedented advertising power, was a crucial catalyst that drew the American Dream away from ideals and instead toward a "national ethos of consumption" (Samuel, 2001).[1]

In this paper, I provide evidence for this popular historical narrative and study the effect of the introduction of TV on retail sales in the United States. I take newly transcribed nationwide administrative data from the *Census of Business* series published by the Department of Commerce in intermittent years over 1930-58. I use an empirical strategy revolving around the FCC Freeze, which has been employed as early as Hennigan et al. (1982) and extensively in economics since Gentzkow (2006). In late 1948, the FCC unexpectedly announced a "temporary" break in issuing TV licenses amid unforeseen signal interference between stations. Although the hiatus was planned to last only six to nine months, almost four years passed before the FCC resolved all issues. Meanwhile, hundreds of broadcasting applications stalled. TV historian Barnouw (1990) labels this gap as a "laboratory period"; some counties had started to gain TV service ("Pre-Freeze" group), while the others could only wait until the Freeze ended ("Post-Freeze" group). I compare the growth in retail sales between the Pre- and Post-Freeze groups during the FCC Freeze.

I find that retail sales in the Pre-Freeze group grew by 3-4% more on average than in the Post-Freeze group. This difference is statistically significant, and also historically significant. For two decades prior to the Freeze, the growth in retail sales had been indistinguishable between the two groups. On the supply side, the trends in retail stores and workers leading up to the Freeze period were also similar. Even the levels of pre-TV retail sales are orthogonal to the selection of Pre- vs. Post-Freeze counties once I control for population, which was the main criterion that broadcasters targeted. Furthermore, I extend the Freeze-based empirical strategy by recording from historical trade magazines when stations first *applied* for TV broadcasting licenses. I study the subset of stations that had applied for licenses before the Freeze was announced but had yet to be built. Within this subset, I find similar differences between the counties near stations that were arbitrarily allowed by the FCC to go on air during the Freeze and those that had to wait.

Next, I investigate whether TV induced consumers to buy more goods in general, or only in certain retail sectors. I decompose the difference in total sales growth into the ten retail sectors that the *Census of Business* used to classify stores (e.g., automobile, apparel, and food). I find that most of the difference in total sales growth is due to differences in the growth rates within certain sectors. The automobile sector is at the forefront, and alone accounts for a third of the total gap.

I then evaluate advertising as a mechanism for the effect of TV. I use the Merchandise

[1] See also Lipsitz, 1990; Halberstam, 1994; Edgerton, 2007; Kennedy, Cohen, and Bailey, 2010; Spigel, 2013.

Line Sales records from the 1948 *Census of Business* to separate sales within each of the ten retail sectors into more granular product categories, which can then be merged to product-level TV advertising revenue data. I examine the relationship between each product's growth in sales and the amount of revenue it attracted for TV advertising. At first glance, I detect no evidence of advertising effects. However, historians and media scholars describe that the effect of TV advertising varied by product (Sterling and Kittross, 2002), and that durable goods such as cars and household appliances in particular became material symbols of the American Dream (Cullen, 2003; Edgerton, 2007; Berkin et al., 2011). These products were not only directly advertised in commercials, but also depicted through product placement in popular TV shows, which aggrandized their function for the ideal suburban family (Lipsitz, 1990; Halberstam, 1994; Samuel, 2012; Nacey, 2015). When I allow the effect of advertising to differ between durable and consumable goods, I indeed find a significant and positive relationship between TV advertising revenue and sales growth in the Pre-Freeze group during the Freeze period for durable goods. There is no such relationship neither in the Post-Freeze group, which did not have TV access during this time, nor for consumables.

Historians also distinguish the 1950s as a pivotal era when the American Dream became commonly conceived as raising a family in the suburbs, an image that television reinforced in popular sitcoms such as *I Love Lucy* and *Make Room for Daddy* (Mintz and Kellogg, 1988; Lipsitz, 1990; Halberstam, 1994; Weiss, 2000; Samuel, 2001; Cohen, 2003; Samuel, 2012; Spigel, 2013). Motivated by these accounts, I explore the effects of TV on two broader societal outcomes: highway construction (a complement for cars and a precondition for suburbanization) and birthrates. Using data from Baum-Snow (2007), I find that the timing of TV access coincided with the start and expansion of local highway construction. The historical context explained by Baum-Snow indicates that this concurrence was demand-driven: the national plan for the highway system was drafted before the Freeze in 1947, and any highway construction likely reflected local commuting demand until federal funding became widely available in 1956. As for births, there were no initial differences until the Freeze, and then the rates in the counties that could watch TV began to rise relative to those without access. This result reflects the traditional gender roles reinforced in the shows, but may also have been a byproduct of couples being at home together more often to watch TV, before modern contraceptives.

This paper adds to a large literature on the wide-ranging effects of media (for an overview, see DellaVigna and La Ferrara, 2015). A stream of papers within this literature focuses on the medium of television. These papers have studied the effect of television on political outcomes in the U.S. and globally (DellaVigna and Kaplan, 2007; Enikolopov, Petrova, and Zhuravskaya, 2011; Durante, Pinotti, and Tesei, 2019), anti-American beliefs in the Middle East (Gentzkow and Shapiro, 2004), social perceptions of women (Jensen and Oster, 2009), social capital (Olken, 2009), and fertility choices (La Ferrara, Chong, and Duryea, 2012; Kearney and Levine, 2015). Those specifically featuring the FCC Freeze have also explored a diverse set of outcomes in the United States, including crime (Hennigan et al., 1982), voter turnout (Gentzkow, 2006), educational outcomes among children (Gentzkow and Shapiro, 2008), beer markets (George, 2009), polarization (Campante and Hojman, 2013), the superstar effect (Koenig, 2021), and labor supply (Fenton and Koenig, 2021).

This paper is most similar to the analyses in Bursztyn and Cantoni (2016) and Baker and George (2010). Bursztyn and Cantoni (2016) examine the effects of past Western television

exposure on the consumption patterns of East Germans after reunification in Germany. Although they do not find an increase in overall consumption as in this paper, they do find advertising effects on the composition of purchases, in that households spent more on intensely advertised goods. Baker and George (2010) also employ the FCC Freeze strategy in the 1940-50s and find that households with TV access reported taking on more non-mortgage debt to purchase durable goods, especially new cars. Reassuringly, their evidence based on a separate data source (the Survey of Consumer Finances) support the findings in this paper.

The differences between this paper and Baker and George (2010) are worth highlighting, given the apparent similarities in the historical setting and the result on automobiles. The main dependent variable in Baker and George is household debt, not retail spending in general, and their survey data are limited to only observing purchases that can be financed through debt (i.e., durables). Yet ex ante, the effect of TV on consumption may not have been limited to goods typically bought on credit, and could also have propelled sales for consumables such as food, toiletries, and cigarettes, which were the most heavily advertised products on TV at the time. In comparison, the administrative dataset in this paper is geographically exhaustive and reveals the universe of retail dollars and the type of store in which they were spent. With this rich and comprehensive dataset, this paper quantifies the magnitude of the TV effect on the entire retail sector—including the supply side—and then decomposes the effect on sales by business categories. The nationwide coverage also allows the paper to explore the heterogeneity of the effect by county-level characteristics. In addition, the analyses link product-by-product sales with advertising expenditure to directly test the prevailing historical hypotheses of the advertising and conspicuous consumption mechanisms, which Baker and George can only indirectly infer from TV access. Beyond these advantages, this paper considers the effects on highway construction and birthrates to broach broader dimensions of American consumerism during this era.

This paper also relates to the literature on conspicuous consumption (Veblen, 1899) and advertising (see Bagwell, 2007, for an overview). These two topics, however, are rarely discussed together, especially in empirical papers (see Krähmer, 2006, for a theoretical model combining the two). The positive effect for durables and the null effect for consumables in this paper support the theories of Bagwell and Bernheim (1996) and Rayo and Becker (2006), who predict that durables are superior for the purposes of conspicuous consumption. Recently, Shapiro, Hitsch, and Tuchman (2021) find that for two-thirds of brands during 2010-2014, TV advertising had insignificant returns in sales, and that the elasticity does not vary by product module. Their results are not inconsistent with those in this paper, and the scopes widely differ. This paper focuses on the introduction of TV in the postwar era rather than its effect as an established medium in 2010-2014, and also analyzes sectors and product markets as a whole rather than the ROI or elasticity for specific brands.

Beyond the economics literature, this paper fills an empirical gap in the prevailing historical theory that emphasizes the role of television in the origins of modern American consumerism (Lipsitz, 1990; Barnouw, 1990; Halberstam, 1994; Samuel, 2001; Edgerton, 2007; Kennedy, Cohen, and Bailey, 2010; Berkin et al., 2011; Spigel, 2013; Nacey, 2015).

2.2 Historical background

2.2.1 Television and the rise of American consumer culture

Commercial television debuted in the United States on July 1, 1941. After the hiatus during World War II, television spread swiftly across the country into 90 percent of American homes by 1960 (Edgerton, 2007; Berkin et al., 2011). By then, viewers watched an average of 3-3.5 hours of TV a day (Edgerton, 2007). From its onset, citizens and scholars alike acknowledged that the "cultural juggernaut" of television was permeating family life, leisure, education, and politics (Coffin, 1955; Allen and Thompson, 2017). In 1951, *The New York Times* reported that television "is influencing the social and economic habits of the nation to a degree unparalleled since the advent of the automobile" (Edgerton, 2007). By the end of the 1950s, television was deemed the "centerpiece" and the "most important discursive medium" in American culture (Edgerton, 2007; Lipsitz, 1990).

Advertisers and even politicians during this era sought to change the average American's attitude toward consumption through TV programming (Lipsitz, 1990; Barnouw, 1990; Halberstam, 1994; Schor, 1999; Samuel, 2001; Cohen, 2003; Edgerton, 2007; Spigel, 2013; Nacey, 2015). The decades prior to the 1950s had been marred by the Great Depression and resource conservation during the war. For a society to emerge that conjoined new possessions with social progress and prestige, a cultural revolution was required. Television, as American sociologist Lipsitz (1990) claims, was the catalyst:

> *"But television's most important economic function came from its role as an instrument of legitimization for transformations in values initiated by the new economic imperative of postwar America. For Americans to accept the new world of 1950s consumerism, they had to make a break with the past."*

The timing was opportune. The postwar boom meant Americans had cash to spend (Katona, 1964); advertisers, increasingly through TV, gave them plenty of reasons to do so. In 1949, the Vice President of the leading advertising firm J. Walter Thompson Company heralded television's unprecedented ability to show the use of products, rather than mere descriptions of them (Fidler, 1949). Even as early as 1950, industry reports predicted that television would surpass radio in audience penetration (Beckinger, 1950). Surely enough, TV had overtaken radio and newspaper's market share in the advertising business by 1954 (Advertising Age, 2005).

Beyond direct advertising, popular shows such as *I Love Lucy*, *The Adventures of Ozzie and Harriet*, and *Leave It to Beaver* depicted vignettes into the melodrama of "ordinary" suburban families (Leibman, 1995; Nacey, 2015). These fictional households, however, unfailingly owned the newest cars and appliances (Leibman, 1995; Nacey, 2015). By casting these families as the middle-class paragon, programmers were keenly aware that they were setting the standards for material aspirations among their audience—in essence, inspiring viewers to keep up with the on-screen fictional Joneses (Lipsitz, 1990; Schor, 1999).

2.2.2 The FCC Freeze of 1948-1952

The birth and spread of television roused many aspiring broadcasters, but they first had to be licensed by the FCC and assigned a frequency on the signal spectrum to televise shows

and commercials. The main broadcasting companies at the time were ABC, CBS, NBC, and DuMont Television Network. These companies were allowed to build up to five "owned-and-operated" stations to broadcast the programs they produced (Brooks and Marsh, 2003). Independent groups constructed many other stations, but for financial reasons, almost all became affiliates of the main broadcasters and televised shows produced by those large networks (Bogart, 1958; Sterling and Kittross, 2002).

However, on September 29, 1948, the FCC announced that it would postpone issuing licenses for new TV stations due to signal interference problems that it had been "unable to foresee" (Sterling and Kittross, 2002). This "temporary" halt lasted "not for the six to nine months suggested by the FCC Chairman Wayne Coy, but for nearly four years" until April 1952 (Sterling and Kittross, 2002; Edgerton, 2007). This prolonged hiatus was in part due to the Korean War (Barnouw, 1990), and the FCC used this time to settle other burgeoning issues, such as expanding the spectrum range for broadcasting signals, choosing which color television system to adopt, and setting aside some frequencies for educational programming (Sterling and Kittross, 2002; Edgerton, 2007).

The staggered diffusion of TV stations caused by the Freeze naturally partitions U.S. counties into two distinct groups, based on when their first local TV station went on air. The Pre-Freeze counties had local TV stations that began broadcasting before or during the Freeze. Meanwhile, the Post-Freeze counties were forced to wait until the Freeze ended for their first local TV stations to be licensed and begin broadcasting. Given intense demand from both potential viewers and broadcasters, the Post-Freeze counties would have likely gained TV access years earlier in the absence of the Freeze (Barnouw, 1990).

2.3 Data

The main sources of data in this paper are: (i) the *Census of Business* series for county-level retail data, (ii) Gentzkow and Shapiro (2008)'s data on the first broadcasting dates of TV stations at the Designated Market Area (DMA) level, and (iii) county-level highway construction data from Baum-Snow (2007) and number of births from the *County Data Book* (ICPSR, 2012). This section describes each of these data sources in further detail. For county-level demographic variables such as population, sex, income, age, education, etc., I use decennial census data from 1940.

2.3.1 *Census of Business*

Data for each county's retail sales (the main dependent variable of interest), stores, and employment are taken from the *Census of Business* series published by the U.S. Bureau of Census. The Bureau compiled this data by surveying all recognizable retail establishments and cross-checking them with the list of businesses reporting to the Social Security Administration. Given the administrative nature of the data, coverage is thorough.[2] The dataset in this paper includes all six waves of the Census between 1930 and 1960, which were conducted in the years 1930, 1935, 1939, 1948, 1954, and 1958. The data for 1930 were taken from

[2]A follow-up investigation estimated that the size of under-coverage in sales for the 1948 data was 1.3 percent. For counties with only a few establishments in a retail category, the sales data in that category are occasionally withheld to avoid disclosure.

Haines and ICPSR (2010), while the data for the remaining years were manually transcribed and checked against ICPSR (2012) whenever possible. Figure A.1 shows an excerpt of the records.

The Census categorizes stores based on their main retail product into the following business categories: food; eating and drinking places; general merchandise; apparel; furniture, furnishings, and appliances; automotive; gasoline service stations; lumber, building, and hardware; drug and proprietary; and other retail. Appendix A.1 describes each of these sectors. In 1948, the Census also compiled Merchandise Line Sales, which disaggregate the sales within each retail sector into more granular product categories (see Figure A.2). For example, a store that specialized in drugs would be classified under the "Drug and proprietary" sector, but could also sell other products including food. In aggregate, the food sales in this drug store would be counted under the "Drug and proprietary" sector, not the "Food" sector; the Merchandise Line Sales dataset, however, would be able to recover the separate amounts of drug and food sales within the "Drug and proprietary" sector. The Census did not report the Merchandise Line Sales for other years.

2.3.2 Television service diffusion

I use Gentzkow and Shapiro (2008)'s dataset for the diffusion of TV service, which is the main independent variable of interest. The dataset includes the first broadcasting dates of local commercial TV stations at the Designated Market Area (DMA) level, which is the standard unit of geography used in industry to analyze television viewership. There are over 200 DMAs nationwide, and each county is assigned to a DMA based on which local television stations dominate viewership in that county.[3] The DMA measure of TV exposure assumes that all counties in that DMA gain TV access when the first station in the DMA goes on air. This approximation may be less accurate for counties on the boundaries of the DMA or with topographical features that hinder signals.[4] I revisit these geographic complications in Section 2.5.2.

2.3.3 Highway construction and births

Highway construction and birthrates are two additional outcomes explored after the main empirical analyses on the retail sector. The highway construction data are taken from Baum-Snow (2007). A crucial feature of Baum-Snow's empirical strategy is the 1947 national highway plan, which "represents most of the system we see built today." The Baum-Snow

[3]For example, the "San Francisco - Oakland - San Jose" DMA consists of 10 Bay Area counties (Alameda, Contra Costa, Lake, Marin, Mendocino, Napa, San Francisco, San Mateo, Santa Clara, and Sonoma). The first TV station in this DMA was KPIX, which began broadcasting as a CBS affiliate in late 1948 from a tower on top of a large hill in San Francisco. In 1949, ABC joined the television market and built their own station named KGO, while the owners of the *San Francisco Chronicle* also constructed the KRON station as an NBC affiliate. By 1949, the major TV broadcasting networks producing the most popular shows (ABC, CBS, and NBC) were all reaching audiences in the Bay Area market.

[4]Fenton and Koenig (2021) use an Irregular Terrain Model to demonstrate how this approximation can produce both false negatives and false positives. For example, they note that the Topeka and St. Joseph market area's first TV station started broadcasting in 1953, but counties on the outskirts could tune into the station in the neighboring Kansas City DMA as early as 1949.

dataset contains only the counties on this 1947 highway plan (see Figures A.4a and A.4b). The county-level data file in Baum-Snow does not record the miles of highways assigned to each county in the 1947 plan, but does have the miles of actual construction until 2000. I use the actual miles of highways constructed by 1990 as a measure of the miles assigned to each county in the 1947 plan, since the 1947 plan served as the blueprint for the present-day system and Baum-Snow states that "By 1990[...]virtually the entire 1947 plan had been built." For the number of births in each county, I use the *County Data Book* dataset (downloaded from ICPSR, 2012), which records births starting in 1944 and also in 1950, 1954, and 1960.

2.4 Empirical strategy

This section lays the foundations for the empirical analysis. First, in Section 2.4.1, I use the FCC Freeze "discontinuity" to partition counties. In Section 2.4.2, I compare the observable characteristics of the Pre- and Post-Freeze groups. While there are notable differences, in Section 2.4.3, I show that retail sales before the Freeze are similar in the two groups after controlling for population.

2.4.1 The disrupted diffusion of TV

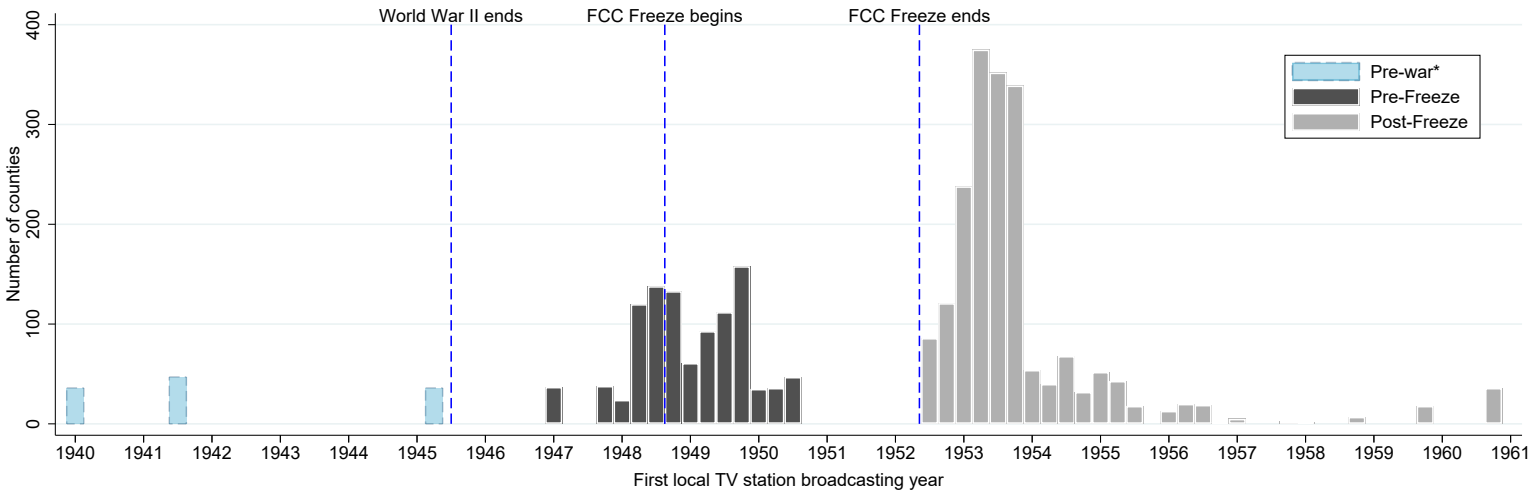
Since the empirical strategy relies on the staggered spread of TV service, I follow Gentzkow (2006), George (2009), and Baker and George (2010) by grouping counties based on when their first local TV station went on air. The Freeze naturally categorizes counties into either (i) the Pre-Freeze group, which gained TV service before or during the Freeze (from 1946 to 1951), or (ii) the Post-Freeze group, which only received TV access after the Freeze ended in 1952.[5] Figure 2.1 shows the distribution of first broadcasting dates across counties. As evident in the figure, the Freeze stemmed the diffusion of TV during its four-year span. Once the Freeze thawed, there was a dramatic spike in counties gaining TV access in 1952 and 1953.

2.4.2 Descriptive statistics: Pre- vs. Post-Freeze group

As Gentzkow (2006) notes, the Freeze disrupted the timing of TV diffusion, but not necessarily the order, since broadcasters first sought licenses for more populous areas that had greater marketing potential. Table 5.2 reports the demographic characteristics of the Pre- and Post-Freeze groups, and Figure A.3 shows the geographic dispersion of TV. There is a concentration of Post-Freeze counties in the northern Mountain and Midwest areas, but most regions contain variation in TV start dates. The Pre-Freeze group has notably higher population, density, and urban proportion on average than the Post-Freeze group. Despite these disparities in the averages, the distributions do overlap as shown in Figures A.5a-A.5k.

[5]The earliest six television markets, which are New York City, Philadelphia, Chicago, Washington D.C., Schenectady (the founding city of General Electric), and Los Angeles, actually began commercial broadcasting before the end of World War II in 1945. This "Pre-war" group is dropped from the analysis, since it gained TV service many years before the FCC Freeze and is much less comparable to both the Pre- and Post-Freeze groups. Table A.1 shows the summary statistics for the Pre-war group.

Figure 2.1: Distribution of first local TV station broadcasting date



This histogram plots the distribution of the first quarter that a local TV station began broadcasting across counties. The *Pre-Freeze* group had a local station that began broadcasting after WWII but before/during the FCC Freeze, which lasted from late-1948 to mid-1952. The first local TV stations for the *Post-Freeze* group were built *after* the Freeze.

*The *Pre-war* group, which had access to television programming before the end of World War II and consists of only six market areas, is excluded from the analysis.

Table 2.1: Summary statistics by TV group

	Pre-Freeze group (1)	Post-Freeze group (2)	(2)-(3)
1948 Retail sales ($1000s)	55155.58	20588.24	34567.34
	[161439.74]	[37189.13]	(7288.09)
1948 Log(retail sales)	9.68	9.19	0.49
	[1.40]	[1.21]	(0.13)
1948 Retail stores	708.10	319.78	388.32
	[1662.63]	[435.04]	(85.32)
1948 Employees per retail store	2.71	2.49	0.22
	[1.31]	[1.18]	(0.10)
Demographic controls			
1940 Population	54132.58	24452.77	29679.81
	[130084.58]	[30734.51]	(6345.03)
1940 Male %	51.07	51.51	-0.43
	[1.73]	[2.13]	(0.22)
1940 Median age	26.74	25.90	0.84
	[3.94]	[3.64]	(0.60)
1940 Median family income ($)	756.75	655.42	101.33
	[354.19]	[300.63]	(41.44)
1940 Persons per sq. mile	200.01	67.58	132.44
	[1053.44]	[387.94]	(53.40)
1940 White %	89.42	87.79	1.63
	[16.27]	[19.04]	(2.76)
1940 Urban %	26.98	19.91	7.07
	[26.41]	[23.28]	(1.89)
1940 High school educated %	19.78	20.15	-0.37
	[7.59]	[7.71]	(1.18)
Counties	1036	1945	
Media markets (DMA)	53	146	

This table shows the average 1948 retail outcomes and 1940 demographic characteristics of counties in the Pre- and Post-Freeze groups. Retail sales and median family income are unadjusted for inflation. Standard deviations are shown in brackets. The last column shows the average difference between the Pre- and Post-Freeze groups with standard errors clustered by DMA (Designated Market Area) in parentheses.

Table 2.2: Selection of Pre- vs. Post-Freeze groups on retail sales before the FCC Freeze

Dep. Var.: Log(retail sales)	(1)	(2)	(3)
Pre-Freeze group	0.48	-0.06	-0.05
	(0.14)	(0.08)	(0.03)
Log(1940 pop.}×$\mathbf{1}\{t = 1930\}$		1.13	1.06
		(0.03)	(0.03)
Log(1940 pop.}×$\mathbf{1}\{t = 1935\}$		1.15	1.07
		(0.03)	(0.02)
Log(1940 pop.}×$\mathbf{1}\{t = 1939\}$		1.18	1.07
		(0.02)	(0.02)
Log(1940 pop.}×$\mathbf{1}\{t = 1948\}$		1.17	1.08
		(0.03)	(0.03)
$\mathbf{1}\{t = 1935\}$	-0.22	-0.41	
	(0.02)	(0.13)	
$\mathbf{1}\{t = 1939\}$	-0.00	-0.45	
	(0.02)	(0.20)	
$\mathbf{1}\{t = 1948\}$	0.66	0.27	
	(0.02)	(0.13)	
Constant	8.53	-2.38	-1.49
	(0.05)	(0.32)	(0.71)
All controls			✓
Observations	11893	11888	11888
Counties	2978	2974	2974
Media markets (DMA)	199	199	199
R^2	0.09	0.81	0.94

This table compares differences in log retail sales between the Pre- and Post-Freeze groups in 1930, 1935, 1939, and 1948 leading up to the FCC Freeze period (1948-52). Estimates are from OLS regressions. "All controls" include all the demographic characteristics listed in Table 5.2 (log population, male %, median age, median family income, log density, white %, urban %, high school educated %, all from 1940) with interactions for each year as well as census-region-by-year fixed effects. The indicators for each year are absorbed by the region-year fixed effects in Column 3. Standard errors shown in parentheses are clustered by DMA.

2.4.3 Selection of Pre-Freeze group

Next, I check whether broadcasters first targeted counties with more active consumers, not just higher populations. Table 2.2 compares the log retail sales between the Pre- and Post-Freeze groups in the years leading up to 1948 before most counties had access to TV. Without any controls in Column 1, the Pre-Freeze group indicator is indeed associated with higher retail sales. However, simply adding a control for population in Column 2 shrinks the coefficient on the indicator to an insignificant value near zero. Column 3 includes the rest of the demographic controls in Table 5.2, which barely moves the coefficient. These findings imply that conditional on demographic controls, the assignment of counties to the Pre- and Post-Freeze groups—and hence to TV exposure during the Freeze—is orthogonal to initial levels of retail sales.

2.4.4 Summary of empirical strategy

Building on these findings and on the prior studies that use this identification strategy, I propose that the FCC Freeze led to exogenous variation in television exposure between Pre- and Post-Freeze groups over 1948-52. Thus, I compare the growth in retail sales between these two groups over years for which retail sales data are available (1930, 1935, 1939, 1948, 1954, and 1958) to identify the effect of TV on retail sales. While Table 2.2 has established that the Pre-Freeze group was not selected on the *levels* of retail sales, I compare the pre-trends before 1948 before the Freeze to verify that the Pre-Freeze group was neither selected on the *growth* of retail sales. I use the next period from 1948-54 to measure the causal effect of TV on retail sales, during which the Pre-Freeze group was "treated" with several additional years of TV access. The last comparison over 1954-58 can reveal whether any differences persisted or became negligible once the Post-Freeze group also gained access to TV.

2.5 Results

This section presents the main results of this paper. First, Section 2.5.1 establishes whether the counties that could watch TV during the Freeze had higher growth in retail activity than counties without TV access, and Section 2.5.2 performs robustness checks. Then, Section 2.5.3 identifies the retail sectors that expanded the most with TV. Section 2.5.4 uses product-level sales and advertising data to evaluate the mechanisms of advertising and conspicuous consumption. Drawing on the retail results and the historical literature, Section 2.5.5 explores broader effects of TV on highway construction and birthrates.

2.5.1 Effect of TV on the retail sector

Did TV increase retail sales? Following previous papers in this empirical setting (e.g., Gentzkow, 2006; George, 2009), I compare the relative trends in the dependent variable between the Pre- and Post-Freeze groups over each period. In particular, I run the regression:

$$\begin{aligned}\log(\text{sales})_{it} = \alpha + \mathbf{\Pi} \boldsymbol{X}_{it} + \delta_1 \text{Yeartrend}_{30\text{-}35} + \delta_2 \text{Yeartrend}_{35\text{-}39} + \delta_3 \text{Yeartrend}_{39\text{-}48} \\ + \delta_4 \text{Yeartrend}_{48\text{-}54} + \delta_5 \text{Yeartrend}_{54\text{-}58} \\ + \text{Pre-Freeze}_i \times (\gamma + \beta_1 \text{Yeartrend}_{30\text{-}35} + \beta_2 \text{Yeartrend}_{35\text{-}39} + \beta_3 \text{Yeartrend}_{39\text{-}48} \\ + \beta_4 \text{Yeartrend}_{48\text{-}54} + \beta_5 \text{Yeartrend}_{54\text{-}58})\end{aligned} \tag{1}$$

where i indexes counties and t the year. The dependent variable $\log(\text{sales})_{it}$ is the natural logarithm of total retail sales. $\boldsymbol{X}_{it}$ is a vector of controls that includes census-region-by-year fixed effects and the base-year demographic variables from Table 5.2 interacted with each year to flexibly control for the time-varying effects of demographics. Pre-Freeze_i is an indicator for whether county i is in the Pre-Freeze group. The Yeartrend_{A-B} variable is coded as "0" until year A and then as "1" from year B onward. Since the model is fully saturated in yeartrend variables, the β coefficients of interest are the difference-in-differences estimates between the Pre- and Post-Freeze groups over each period.

Table 2.3 presents the estimates of Equation 1. Column 1 begins by comparing the raw trends without any controls, and Columns 2-4 add combinations of demographic controls and/or county fixed effects. Notably, the specifications in Columns 2-4 explain most of the variation with R^2 values above 0.80. To visualize the relative trends, Figure 2.2 plots the differences between the Pre- and Post-Freeze groups from the full specification in Column 4. The main results are consistent across specifications.

In all three "pre-treatment" periods leading up to the Freeze (1930-35, 1935-39, 1939-48), the pre-trends in retail sales are statistically indistinguishable between the Pre- and Post-Freeze groups.[6] Therefore, any significant difference in the following Freeze period is unlikely to be driven by pre-existing confounders and would be a deviation from the trends over the previous 18 years.

Nevertheless, over the next 1948-54 period coinciding with the Freeze, growth in retail sales is significantly stronger in the Pre-Freeze group than in the Post-Freeze group, with estimates of the difference ranging from 3.2 to 4.5 log points ($p < 0.05$). This magnitude roughly translates into \$131 additional annual spending for the median Pre-Freeze household in 1954 dollars, or around \$1263 today.[7]

Once the Post-Freeze group also gains TV access in the last 1954-58 period, the difference in retail sales growth returns to being statistically indistinguishable. The fact that the difference does not become negative—that is, the Post-Freeze group does not have higher growth relative to the Pre-Freeze group—indicates the Pre-Freeze group continues on the post-TV consumption trend while the Post-Freeze group catches up. This pattern aligns with the historical context discussed in Section 2.2.1 and with papers in the same setting such as Gentzkow (2006) and Fenton and Koenig (2021), which find that the TV induced a long-lasting change in lifestyle rather than a temporary shock caused by novelty.[8]

Were there any effects on supply? Columns 5-6 estimate the full specification but on supply-side outcomes. The dependent variable is the number of employees per retail store in Column 5 and the natural logarithm of the number of retail stores in Column 6. Reassuringly, there are no pre-trends in both outcomes. Then during the Freeze period (1948-54), the Pre-Freeze group has higher growth in the number of employees per retail store by 0.069, or 2.5% relative to the 1954 average of 2.71. Over the subsequent period (1954-58), the Pre-Freeze group has stronger growth in the number of retail stores by 1.8 log points. These results are suggestive ($p < 0.1$), though not conclusive, of a short-term supply response in labor

[6] In specifications with fixed effects, the Pre-Freeze group shows a slight yet insignificant uptick at the end of the 1939-48 period when its counties begin to gain TV access.

[7] The median volume of retail sales in the Pre-Freeze group was \$14323000 in 1948. In the raw data, retail sales grew by 0.1131 log points on average in real terms over 1948-54 in the Post-Freeze group and 0.1586 log points in the Pre-Freeze group. Therefore, the effect of TV on retail sales for the median Pre-Freeze county can be roughly calculated as $14323000 \times (e^{0.1586} - e^{0.1131}) \approx \$746589$, or \$833799 in 1954 dollars after accounting for an inflation rate of around 11.68%. The median number of households (interpolated between decennial census years assuming a constant growth rate) for the Pre-Freeze group in 1954 was 6359, and hence the effect on the median household is approximately 833799/6359 =\$131. A dollar in 1954 is worth around \$9.64 in 2020 dollars (U.S. Bureau of Labor Statistics, 2021).

[8] These results assume that people in counties with TV access actually watched TV. Television's near monopolization of leisure during this era is well-documented (Barnouw, 1990; Fenton and Koenig, 2021). Furthermore, Figure A.6, which graphs the share of households that own TV sets in the average Pre- and Post-Freeze county across time, provides indirect evidence consistent with the staggered diffusion of TV.

followed by a long-term adjustment in the number of retail stores.

Back on the demand side, Table 2.4 explores the heterogeneity of the TV effect across observable characteristics. Each row represents a separate regression building on the full specification in Column 4 of Table 2.3 with an additional interaction term Pre-Freeze$_i \times$ Yeartrend$_{48-54} \times \mathbf{1}\{x_i \leq \tilde{x}\}$, where x_i is the dimension of interest listed in the leftmost column and $\tilde{x}$ is the Pre-Freeze group median. This additional interaction term allows the TV effect identified over the Freeze period to vary within the Pre-Freeze group conditional on the county being above or below the median of the observable characteristic.

For example, the first row shows that the growth in retail sales for Pre-Freeze counties *below* the median population is 5.9 log points higher than for the Post-Freeze group over the 1948-54 Freeze period. Meanwhile, in the Pre-Freeze counties *above* the median, the growth rate is actually similar to that in the Post-Freeze group (-0.4 log points, $p = 0.83$). Hence, the difference in the relative trends between the below- and above-median Pre-Freeze subgroups is $5.9 - (-0.4) = 6.3$ log points (s.e.=2.2).

Across dimensions, the TV effect within the Pre-Freeze group appears to be concentrated in the less populated, more rural counties with lower initial retail sales and perhaps lower radio penetration. Plausibly, the technology of TV was a more drastic innovation in entertainment and advertising in such areas. The findings also alleviate selection concerns, since these smaller counties are unlikely to be the ones directly targeted by broadcasters.

2.5.2 Robustness checks

This section provides a battery of robustness checks for the main result of the TV effect on retail sales. Figure A.7a-A.7b begins by comparing subsets of Pre- and Post-Freeze counties with stronger arguments for identification.

Figure A.7a emulates a procedure from Gentzkow (2006) and Durante, Pinotti, and Tesei (2019) and estimates the TV effect on the least populated counties within each DMA. The timing of TV access in these smaller counties is more likely to be exogenous since the first TV stations in their vicinity were not targeted at them but rather at the large cities nearby. The figure shows the difference in retail sales growth between the Pre- and Post-Freeze groups over 1948-1954 when the Freeze occurred, while keeping only the smallest p% of counties within each DMA. As p is varied along the horizontal axis, the point estimate of the difference remains stable and becomes statistically significant at around $p = 55\%$.

Figure A.7b next compares the closest counties in the Pre- and Post-Freeze groups in terms of the first local broadcasting date. That is, the sample is restricted to Pre-Freeze counties in the last N DMAs and Post-Freeze counties in the first N DMAs to gain TV access. This analysis is informative if we believe that counties closer in their first broadcasting dates on either side are more comparable. The figure plots the difference in retail sales growth between these subgroups over 1948-1954 and varies N along the horizontal axis. The result stays consistent throughout and becomes significant with just 20 DMAs from each group.

Table C.6 verifies that the TV effect on retail sales is robust to a parametric representation of the TV effect, alternate dependent variables, subsamples of interest, and spatially correlated standard errors. All specifications in this table include the full set of demographic controls and county fixed effects.

Table 2.3: Relative retail trends

	Log(retail sales)				Employees/store	Log(retail stores)
	(1)	(2)	(3)	(4)	(5)	(6)
Pre-Freeze TV group	0.459	-0.059				
	(0.140)	(0.029)				
Differences in growth relative to Post-Freeze group over each period						
Pre-Freeze×Yeartrend$_{30\text{-}35}$	0.010	-0.003	0.018	-0.001	-0.005	0.003
	(0.029)	(0.022)	(0.028)	(0.022)	(0.020)	(0.013)
Pre-Freeze×Yeartrend$_{35\text{-}39}$	0.025	0.003	0.020	0.003	-0.012	-0.007
	(0.020)	(0.016)	(0.020)	(0.016)	(0.017)	(0.010)
Pre-Freeze×Yeartrend$_{39\text{-}48}$	-0.008	0.018	-0.002	0.018	0.019	-0.007
	(0.020)	(0.016)	(0.020)	(0.016)	(0.026)	(0.014)
Pre-Freeze×Yeartrend$_{48\text{-}54}$	0.045	0.035	0.043	0.032	0.069	-0.005
	(0.018)	(0.015)	(0.019)	(0.016)	(0.036)	(0.012)
Pre-Freeze×Yeartrend$_{54\text{-}58}$	0.015	0.006	0.012	0.006	-0.021	0.018
	(0.012)	(0.009)	(0.011)	(0.009)	(0.035)	(0.010)
Yeartrend$_{30\text{-}35}$	-0.224		-0.224			
	(0.021)		(0.021)			
Yeartrend$_{35\text{-}39}$	0.210		0.215			
	(0.016)		(0.016)			
Yeartrend$_{39\text{-}48}$	0.665		0.660			
	(0.015)		(0.015)			
Yeartrend$_{48\text{-}54}$	0.113		0.110			
	(0.012)		(0.012)			
Yeartrend$_{54\text{-}58}$	0.055		0.057			
	(0.008)		(0.007)			
Constant	8.538	-1.669	8.696	8.313	3.375	5.850
	(0.054)	(0.725)	(0.016)	(0.379)	(0.738)	(0.262)
Demographic controls		✓		✓	✓	✓
County FE			✓	✓	✓	✓
Observations	17841	17822	17841	17822	17834	17839
Counties	2981	2974	2981	2974	2974	2974
Media markets (DMA)	199	199	199	199	199	199
R^2	0.13	0.94	0.80	0.83	0.79	0.29

This table compares the relative trends between the Pre- and Post-Freeze groups in log retail sales in Columns 1-4, employees per retail store in Column 5, and log retail stores in Column 6. The omitted group is the Post-Freeze group. The years in this panel data set are 1930, 1935, 1939, 1948, 1954, and 1958. Yeartrend$_{\text{A-B}}$ is a variable coded as "0" until year A and as "1" for years B onward. Since this model is fully saturated in yeartrend variables between each time period, the coefficient on Yeartrend$_{\text{A-B}}$ captures the average growth between year A and year B. The interaction of Yeartrend$_{\text{A-B}}$ and the Pre-Freeze group indicator estimates the difference-in-differences between the Pre- and Post-Freeze groups from year A to year B. Demographic controls include all those listed in Table 5.2 (log population, male %, median age, median family income, log density, white %, urban %, high school educated %, all from 1940) with interactions for each year as well as census-region-by-year fixed effects. When demographic controls are included, the region-year effects absorb the uninteracted Yeartrend$_{\text{A-B}}$ variables. In Columns 3-6, the constant is the average of the county-level fixed effects, which absorb the Pre-Freeze group indicator. Standard errors shown in parentheses are clustered by DMA.

Figure 2.2: Differences in retail trends between the Pre- and Post-Freeze groups



This figure plots the estimates from Columns 4-6 of Table 2.3 and shows the differences between the Pre- and Post-Freeze groups in the growth of log retail sales, log retail stores, and employees per retail stores over the time periods denoted on the horizontal axis. 95% confidence intervals are shown with standard errors clustered by DMA.

Table 2.4: Heterogeneity of the TV effect on log retail sales growth over 1948-54

	Relative trends within Pre-Freeze group		
	Below median	Above median	Below - above diff.
1940 Population	0.059	-0.004	0.063
	(0.018)	(0.020)	(0.022)
1940 Density	0.037	0.025	0.012
	(0.022)	(0.019)	(0.029)
1940 Median family income	0.023	0.044	-0.022
	(0.021)	(0.021)	(0.029)
1940 Urban %	0.053	0.007	0.045
	(0.020)	(0.019)	(0.023)
1940 Male %	0.018	0.044	-0.026
	(0.017)	(0.023)	(0.025)
1940 White %	0.050	0.013	0.037
	(0.029)	(0.017)	(0.036)
1940 High school educated %	0.045	0.017	0.028
	(0.021)	(0.020)	(0.026)
1939 Retail sales	0.053	0.004	0.049
	(0.019)	(0.018)	(0.021)
County mean elevation	0.050	0.014	0.036
	(0.025)	(0.029)	(0.044)
1950 Households with radio %	0.056	0.003	0.053
	(0.025)	(0.017)	(0.031)

Each row represents a separate regression where the counties in the Pre-Freeze group are partitioned into halves based on whether the county is below or above the Pre-Freeze group median of the variable in the leftmost column. The regression takes the same specification as in Column 4 of Table 2.3 but adds an interaction on the Pre-Freeze×Yeartrend$_{48\text{-}54}$ term with an indicator for whether a county is below the median for the Pre-Freeze group. Under the "Below median" ("Above median") column, the table reports the growth in log retail sales for the Pre-Freeze counties that are below (above) the median, relative to the Post-Freeze group. The "Below - above diff." column shows the estimated difference in the relative growth between the Pre-Freeze counties below and above the median. Standard errors shown in parentheses are clustered by DMA.

Column 1 expresses the effect of TV as a linear term in the number of years with TV access, which is a common parametric specification in previous studies (e.g., Gentzkow, 2006; Gentzkow and Shapiro, 2008; Baker and George, 2010). The downside of this specification in this context is that we do not observe retail sales for counties annually, but only in intermittent years when the *Census of Business* was conducted. Nevertheless, the results still find a significant effect of TV, estimating an increase of 0.97 log points (s.e.=0.48) in retail sales for each additional year of TV access.

Columns 2 and 3 examine alternate outcomes to investigate whether the effect of TV on retail sales is driven by differential trends in the number of households or in income. The dependent variable in Column 2 is retail sales per household, where the number of households in each county is interpolated between decennial census years assuming a constant growth rate. The dependent variable in Column 3 is retail sales as a share of income, where each county's income is approximated by interpolating the number of households and the median family income, and then multiplying the two.[9] Despite the measurement error in these approximations, the main patterns remain unchanged; the Pre-Freeze group has significantly

[9]Statistics on the average family income are not available in the public use samples, and complete count

higher growth in retail sales per household and as a share of income only during the 1948-54 Freeze period.

Columns 4 to 6 investigates different subsamples of interest. Column 4 evaluates the "distraction hypothesis", which stipulates that TV had in fact a temporary negative effect on retail sales by distracting consumers from going shopping, which subsided when the novelty wore off. Under this hypothesis, the difference between the Pre- and Post-Freeze groups over 1948-54 may be from TV *undermining* retail sales in the Post-Freeze counties that had just gained TV access, rather than increasing retail sales in the Pre-Freeze group, which may have become accustomed to TV. To test this hypothesis, Column 4 takes the subset of Pre-Freeze counties with first broadcasting dates in 1947 and 1948 and the Post-Freeze counties with first broadcasting dates at least half a year before 1954. There is no evidence for distraction in the Pre-Freeze subgroup over 1939-48, and the difference over 1948-54 remains intact.

Column 5 compares the Construction Permit-holders (CP) and Pending Applicants (PA), which are discussed in detail in Appendix A.2. For these subgroups, there is clear historical evidence that they would have received TV service around the same time, if not for the Freeze. Both had applications on file for TV services in their respective areas when the Freeze was announced. While the Construction Permit-holders gained TV access mostly within a year, the Pending Applicants were delayed by four years. These groups also display similar differences in retail sales trends as the full sample.[10]

Columns 7 and 8 compute Conley (1999) standard errors, which accounts for spatial correlation by allowing the error term to be arbitrarily correlated among counties within some distance threshold. The main result remains significant at the 5% level under a threshold of 350 miles and at the 10% level even under a threshold of 500 miles.

As mentioned in Section 2.3.2, the DMA approximation for the start of TV access may be less accurate for counties on the boundaries or with geographical features that obstruct signals. I inspect these issues in Table A.4a-A.4b. The first panel compares the average elevation and area between the Pre- and Post-Freeze groups at the county- and DMA-level. People living in places with rougher terrain or in more expansive, remote locations may have poorer TV signal but may also purchase more cars (which Section 2.5.3 finds is the primary product driving the TV effect). While there are no significant differences in land area, the Pre-Freeze group reassuringly has lower average elevation. Based on topographical features, the average Pre-Freeze county was better suited to receive TV signals, but was not predisposed to buy automobiles. Figure A.10 shows the distributions of elevation and area both in logs given the severe right skew.

Table A.4b systematically considers the impact of these geographic characteristics and also the heterogeneity for the border counties that straddle two DMAs with different TV start dates. Column 1 replicates the main result from Column 4 of Table 2.3, and Column 2 adds flexible controls for elevation and area by interacting each feature with an indicator for every year in the panel. The results are largely unchanged. Column 3 explores heterogeneity in the TV effect over the Freeze period among counties that are more prone to the

census data have been released only until 1940. Since the census records income data only from 1940 onward, Column 3 does not extrapolate income to years before 1939. Table A.3 finds no pre-trends in proxies for income.

[10]Column 6 is discussed in Section 2.5.5.

DMA approximation error for TV access. The Pre-Freeze counties bordering a Post-Freeze DMA may have been too far from the TV stations in their own DMA to receive service, and conversely, the Post-Freeze counties bordering a Pre-Freeze/Pre-war DMA may have been able to tune into a spillover signal. There is suggestive evidence that the Pre-Freeze counties on the boundaries were indeed too far to experience a TV effect; these counties have marginally significant lower growth in retail sales relative to the non-boundary Pre-Freeze counties. On the other hand, there is no sign of a spillover TV effect for Post-Freeze counties on the border. This is reasonable given that the signal may have been too weak even for counties on the outskirts of the Pre-Freeze DMAs. Column 4 incorporates the topographical controls and finds similar patterns.

Lastly, I investigate whether there were any changes in TV programming or advertising content that coincided with the timing of the Freeze. For example, there may have been a specific TV series that encouraged retail sales during the Freeze, but was then discontinued when the Freeze ended. In this case, the influence of TV would only be a historical episode limited to a few years for the Pre-Freeze counties, rather than a prolonged effect throughout the nation.

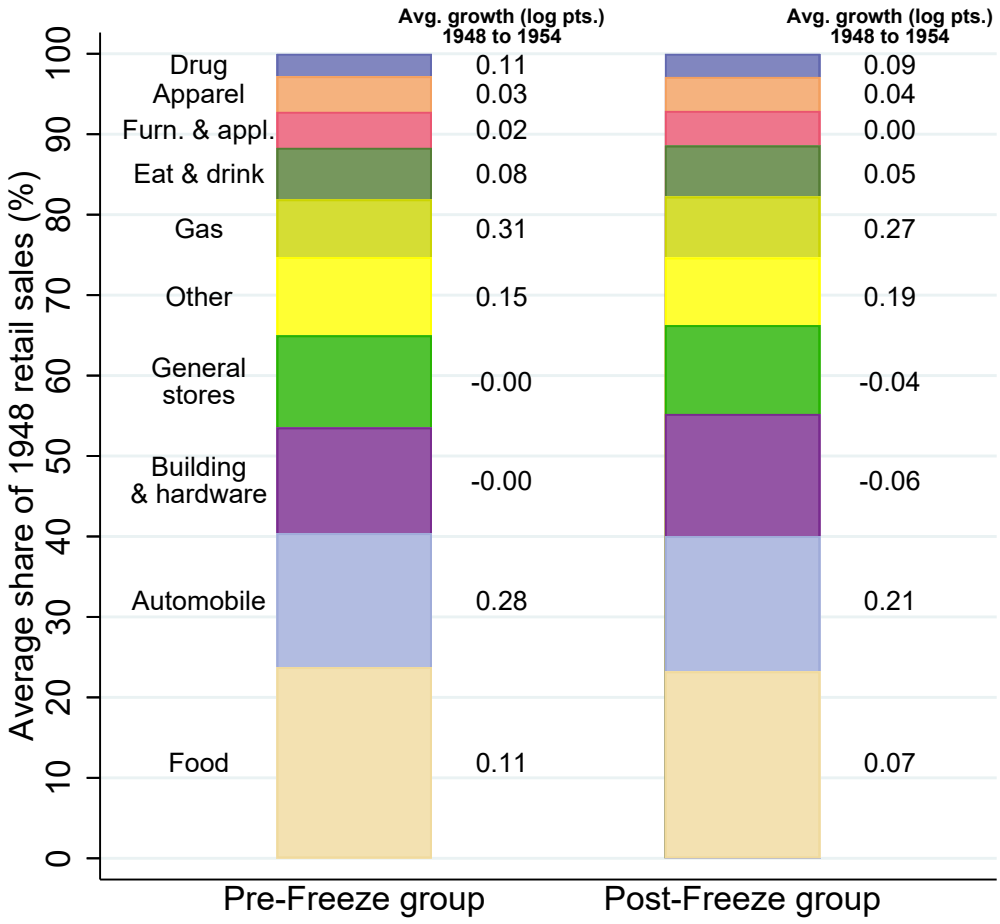
Table A.5a-A.5b compiles the top ten TV series each year from 1950-58 since Nielsen began to collect data on viewership.[11] The popularity of the top TV shows remains stable during and after the Freeze. For example, seven of the top ten programs in the 1952-53 season following the Freeze are in the top ten for the previous season, and the classic 1950s show *I Love Lucy* is in the top three for every year of its airtime from 1951-57. In addition, Figure A.10 (reproduced from an industry magazine) lists the ten brands with highest annual TV advertising expenditures over 1951-54. The leading advertisers are fairly stable across years with no clear overhaul around the Freeze; eight of the top ten in 1952 when the Freeze ended are also among the top ten in 1953.

2.5.3 Effect of TV by retail sector

Did TV viewers buy more products across all retail sectors evenly, or only in certain sectors? This section decomposes the effect of TV on overall sales into the ten retail sectors used by the Census (e.g., automobile, apparel, and food; see Appendix A.1 for details). The contribution of each sector to overall retail sales growth can be disaggregated into two components: (1) the growth rate within the sector, and (2) its initial share of total retail sales. For illustration, if apparel sales grew by 50% but its initial share of total sales was just 1%, then its effect on the growth of overall sales would only be 0.5%. Figure 2.3 summarizes both components for each retail sector and highlights "Food", "Automobile", "Building and hardware", and "General stores" as the large sectors that could drive the TV effect.

[11]The Nielsen ratings measure the average percent of all households with TV sets in the country that watch the show. As the high ratings reflect, there was remarkable homogeneity in the content that audiences watched nationwide in these early flourishing years of network television. Given the prohibitive costs for producing shows and commercials and purchasing airtime, most of the content on TV was sponsored by large companies and was broadcast nationally, in contrast to localized radio content (Bogart, 1958; Sterling and Kittross, 2002). As shown in Figure A.9, by early 1951, the main networks (NBC, CBS, ABC, DuMont) had affiliate stations broadcasting their shows in virtually every television market available at the time. Despite higher viewership later in the decade, the absolute ratings for the top shows decreased presumably as the market became increasingly competitive and diluted with more TV shows.

Figure 2.3: 1948 expenditure shares and 1948-54 sales growth by retail sector



This figure shows each retail sector's initial share of total sales in 1948 and the growth in log sales from 1948 to 1954.

For the formal breakdown, I apply the Oaxaca-Blinder decomposition to the following first-order approximation:[12]

$$\Delta \log(\text{total sales})_i = \nu + \sum_j \omega_j \Delta \log(\text{sales})_{ij} \tag{2}$$

where $\Delta \log(\text{sales})_{ij}$ is the growth in log sales in sector j for county i. The constant ν corrects for the average approximation error, and the coefficient ω_j estimates the average contribution of 1 log point growth in sector j to overall growth, and should roughly correspond to the sector's initial share of sales. The decomposition separates the difference in the growth of total retail sales into sector-level differences in the initial shares ω_j versus differences in the growth $\Delta \log(\text{sales})_{ij}$.

Columns 1 and 3 in Table 2.5 report the average sector-level growth for the Pre- and Post-Freeze groups respectively, and Columns 2 and 4 show the weights (i.e., the coefficients from Eq. 2) reflecting each sector's retail share. Columns 5-7 present the decomposition.

Column 5 lists the amount of the difference in total retail sales growth due to the difference in growth within each sector. The last row shows that these sector-level differences in growth explain most of the difference in total sales growth—3.4 (s.e.=1.7) of the overall 4.3 log points. In particular, stronger sales in the automobile sector contribute 1.3 log points (s.e.=0.4), which amount to around a third of the total gap. In dollar units, the median household in the Pre-Freeze group spent around an extra $40 in the automobile sector in 1954. For comparison, the Ford Customline, the most produced car model in 1954, cost $1744 at base value (Flory, 2008). Then for every 45 or so households, the Pre-Freeze group had one more new car.

Column 6 records the extent to which differences in the weights, accounting for the shares of the sectors, play a role. As Figure 2.3 shows, the initial shares are comparable between the two groups, and thus their contribution to the overall growth differential is minimal. Column 7 accounts for the minor differences in the interaction of sector-level weights and growth rates.[13]

[12]The overall growth in total retail sales is equal to the weighted average of the growth in each sector, where the weights are the initial shares. The log of total sales in county i at time t is the log of the sum of the sales in each sector j. That is, $\log(\text{total sales}_{it}) = \log(\sum_j \text{sales}_{ijt})$. Total sales in the next period $t+1$ can be written as $\log(\text{total sales}_{it+1}) = \log(\sum_j (1+g_{ijt})\text{sales}_{ijt})$, where g_{ijt} is the percent growth in sector j from t to $t+1$. A first-order Taylor approximation provides:

$$\log(\text{total sales}_{it+1}) \approx \log(\text{total sales}_{it}) + \sum_j \left(\frac{\text{sales}_{ijt}}{\text{total sales}_{it}} \right) g_{ijt}$$

i.e., the log growth in total sales is approximately the average growth in each sector weighted by its share in time t. As with total growth, I approximate the sector-level growth rate g_{ijt} using log differences.

[13]Table A.6 shows a decomposition fixing the weights at the average shares in 1948. The results are similar.

Table 2.5: Decomposition of difference in retail sales growth over 1948-54 between Pre- and Post-Freeze groups

	Pre-Freeze group		Post-Freeze group		Diff. in $\overline{\Delta \log(\text{retail sales})}$ due to:		
	$\overline{\Delta \log(\text{sales})}$	Coefficient	$\overline{\Delta \log(\text{sales})}$	Coefficient	$\Delta \log(\text{sales})$	Coefficient	Interaction
Category	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Food	0.108 (0.015)	0.224 (0.022)	0.072 (0.011)	0.252 (0.018)	0.009 (0.005)	-0.002 (0.002)	-0.001 (0.001)
General stores	-0.001 (0.024)	0.118 (0.009)	-0.041 (0.014)	0.098 (0.010)	0.004 (0.003)	-0.001 (0.001)	0.001 (0.001)
Apparel	0.032 (0.012)	0.045 (0.007)	0.039 (0.012)	0.028 (0.005)	-0.000 (0.000)	0.001 (0.000)	-0.000 (0.000)
Furniture & appliances	0.024 (0.017)	0.030 (0.008)	0.002 (0.015)	0.039 (0.005)	0.001 (0.001)	-0.000 (0.000)	-0.000 (0.000)
Automobile	0.283 (0.019)	0.187 (0.012)	0.211 (0.013)	0.183 (0.010)	**0.013** (0.004)	0.001 (0.003)	0.000 (0.001)
Gasoline	0.313 (0.026)	0.101 (0.013)	0.274 (0.024)	0.103 (0.011)	0.004 (0.004)	-0.001 (0.005)	-0.000 (0.001)
Building & hardware	-0.003 (0.021)	0.120 (0.012)	-0.061 (0.017)	0.131 (0.011)	0.008 (0.004)	0.001 (0.001)	-0.001 (0.001)
Eating & drinking	0.083 (0.019)	0.092 (0.032)	0.047 (0.017)	0.087 (0.012)	0.003 (0.002)	0.000 (0.002)	0.000 (0.001)
Drug stores	0.107 (0.010)	0.041 (0.011)	0.091 (0.007)	0.051 (0.010)	0.001 (0.001)	-0.001 (0.001)	-0.000 (0.000)
Other	0.155 (0.030)	0.085 (0.007)	0.187 (0.020)	0.089 (0.008)	-0.003 (0.003)	-0.001 (0.002)	0.000 (0.000)
Constant		-0.003 (0.006)		-0.012 (0.005)		0.009 (0.008)	
Total retail	0.153 (0.015)		0.110 (0.011)		0.034 (0.017)	0.013 (0.007)	-0.004 (0.005)

This table decomposes the difference in log retail sales growth over 1948-54 between the Pre- and Post-Freeze groups into differences in the growth within each retail sector and differences in the weight of each sector. Columns 1 and 3 report the average growth within each sector over 1948-54 for the Pre- and Post-Freeze groups respectively. Columns 2 and 4 show the weights, which are the coefficients from a first-order approximation of each sector's initial share of total retail sales. Columns 5-7 are the results from the Oaxaca-Blinder decomposition of Equation 2. The estimates report the difference in retail sales growth between the Pre- and Post-Freeze groups attributable to differences in each sector's growth (Column 5), weight (Column 6), and interaction between the two terms (Column 7). Missing sector-level growth data are imputed as 0 and included with a missing indicator. Bolded estimates are significant at the 0.005 level (using the Bonferroni correction for multiple comparisons). Standard errors shown in parentheses are clustered by DMA.

2.5.4 Advertising and conspicuous consumption

Why did television increase sales in select retail sectors? Prior studies in economics have focused on the role of advertising in attracting consumers to purchase certain brands and products over others (George, 2009; Bursztyn and Cantoni, 2016). In addition, sociologists and psychologists have argued that television can motivate conspicuous consumption, as the possessions owned by the upper-middle class households portrayed on shows become mainstream markers of status (O'Guinn and Shrum, 1997; Schor, 1999). Historical accounts allude to these mechanisms of advertising and conspicuous consumption in the 1950s, when television emerged as "the most powerful advertising medium in history" (Samuel, 2001) and attaining social status through relative wealth became part of the American Dream (Halberstam, 1994; Cullen, 2003; Samuel, 2012).[14]

This section combines several sources of data to explore the effect of TV through advertising and product visibility, which is a crucial component of the conspicuous consumption model. Table 2.6 compiles the summary statistics from these various datasets. The first dataset draws from issues in the early 1950s of the leading industry magazine *Advertising Agency* (see Figure A.11 for an excerpt). This dataset records nationwide TV advertising revenue over 1951-53 for individual products, such as "Household soaps and cleaners" and "Smoking materials", and serves as a coarse measure of advertising intensity. The fourth column in Table 2.6 lists each product's share of the total TV advertising revenue during this period.

To match the TV advertising revenue data to sales at the product level, I use the Merchandise Line Sales section in the 1948 *Census of Business*. The ten-sector categorization from the preceding Section 2.5.3 is inadequate to match product-level data, since it assigns stores to one of the ten retail sectors based only on the main retail product. However, a single store can sell a variety of products that spans retail sectors. For example, if a grocery store sold drugs, its drug sales would be counted under the "Food" sector rather than the "Drug and proprietary" sector. In remedy, the Merchandise Line Sales dataset recovers sales for the specific products *within* each retail sector (see Figure A.2 for an excerpt).

From the Merchandise Line Sales records, I calculate the share of each product's sales in each retail sector in the second column of Table 2.6. Some products have their sales concentrated in their corresponding sectors; for instance, 95% of the "Automotive" product is sold in stores classified under the "Automobile" retail sector, and 87% of the "Building materials" product is sold in stores classified under the "Building & hardware" sector. On the other hand, at least 40% of the "Apparel & accessories" and "Household equipment & furnishing" products are sold in the "General stores" sector, even though there are separate sectors for "Apparel" and "Furniture & appliances". In the third column, I list each product's share of total sales. The "Food" product comprises the highest share of spending at 30%,

[14]There are also contemporary references to these mechanisms. For example, the 1952 "Hofstra" studies (NBC, 1952a,b) provides survey evidence for the effectiveness of television advertising. Bogart (1958) reports that "television is helping to shape values and set standards simply by depicting a world somewhat more glamorous than present reality. And the perquisites and comforts of this fantasy world become the very objects to which the audience aspires, and which it ultimately acquires." Former First Lady Eleanor Roosevelt's 1961 column titled "What happened to the American Dream?" warns that the younger generations are in danger of abandoning the American Dream rooted in democracy and equality, and instead becoming "too concerned with the ways of earning a living or impressing our neighbors or getting ahead."

followed by "Automotive" (14%) and "Apparel & accessories" (13%).

In the last column of Table 2.6, I report Heffetz (2011)'s index of product visibility and the corresponding product code from Heffetz (2011) in parentheses.[15] Heffetz constructs this index from a national U.S. phone survey in 2004-2005, which asked respondents to rate how long it would take them to notice that their neighbor had spent an above average amount on particular products.[16]

There are two limitations, however, in applying the index to the context of this paper. First, the conspicuousness of products changes over time. The survey respondents in Heffetz (2011) rated tobacco as the most visible product, above cars, clothing, and jewelry. Yet norms regarding smoking were drastically different in the 1950s, when smoking was much more prevalent and perhaps much less conspicuous. Second, some product categories in the advertising and sales data lack a clear match to the categories in Heffetz (2011). For instance, there is no direct match for the "Building materials" product, which I assign the same visibility as "Home furnishings and household items" as a proxy.

Another method to capture product visibility is to sort products into durables and consumables, as shown in parentheses next to each product title. As argued in Bagwell and Bernheim (1996) and Rayo and Becker (2006), durables are superior for conspicuous consumption, since they can be repeatedly flaunted to a wide audience. Furthermore, they are often portable (e.g., jewelry) and can be displayed "during the normal course of social interaction" (Bagwell and Bernheim, 1996). Meanwhile, there are limited opportunities to buy a fancy bottle of champagne in everyday social intercourse, and even so, the gain in social status would be fleeting and localized to those in the vicinity.

[15]The index is taken from the Normalized Mean in Table 3 of Heffetz (2011). Higher values indicate greater visibility.

[16]The main question in the survey asked: *"Imagine that you meet a new person who lives in a household similar to yours. Imagine that their household is not different from other similar households, except that they like to, and do, spend more than average on [product]. Would you notice this about them, and if so, for how long would you have to have known them, to notice it? Would you notice it almost immediately upon meeting them for the first time, a short while after, a while after, only a long while after, or never?"*

Table 2.6: Summary statistics on sales, advertising, and visibility by product

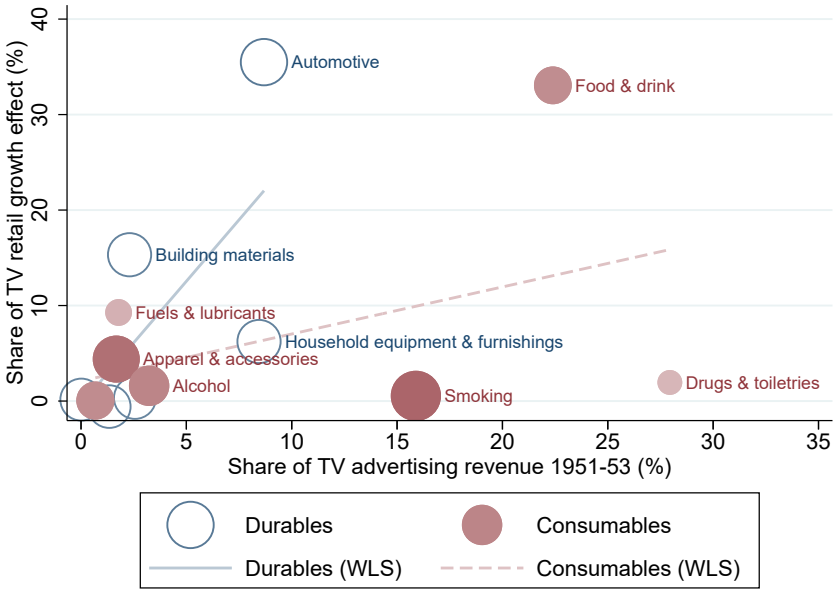
Product (**D**urable/**C**onsumable)	Share of product sales by sector (1948)	% of total sales (1948)	% of TV ad rev (1951-53)	Visibility index (Heffetz, 2011)
Food & drink (**C**)	Food (78%), General stores (3%), Eating & drinking (17%), Drug stores (2%), Other (1%)	29.73	22.39	0.56 (FdO/H[a])
Automotive (**D**)	General stores (1%), Automobile (95%), Gas (3%), Building & hardware (1%)	14.41	8.68	0.73 (Car)
Apparel & accessories (**C**)	General stores (43%), Apparel (57%)	13.22	1.67	0.71 (Clo)
Household equipment & furnishings (**D**)	General stores (40%), Furniture & appliances (52%), Automobile (1%), Gas (< 1%), Building & hardware (5%), Other (2%)	8.21	8.45	0.68 (Fur)
Fuels & lubricants (**C**)	Food (4%), General stores (1%), Automobile (5%), Gas (62%), Building & hardware (3%), Other (26%)	6.55	1.78	0.39 (Gas)
Building materials (**D**)	General stores (11%), Furniture & appliances (< 1%), Building & hardware (87%), Other (1%)	6.53	2.30	0.68 (Fur[b])
Alcohol (**C**)	General stores (< 1%), Eating & drinking (59%), Drug stores (3%), Other (38%)	5.10	3.23	0.61 (AlO/H[a])
Agriculture & farming (**C**)	General stores (5%), Automobile (4%), Building & hardware (40%), Drug stores (1%), Other (51%)	4.17	< 0.01	–
Miscellaneous consumables (**C**)	Food (17%), General stores (12%), Automobile (27%), Gas (15%), Building & hardware (2%), Eating & drinking (5%), Drug stores (4%), Other (19%)	3.86	–	–
Drugs & toiletries (**C**)	General stores (23%), Drug stores (77%)	2.37	27.94	0.36 (Med)
Miscellaneous durables (**D**)	General stores (2%), Furniture & appliances (26%), Automobile (22%), Building & hardware (25%), Other (25%)	1.88	–	–
Smoking (**C**)	Food (5%), General stores (12%), Eating & drinking (19%), Drug stores (37%), Other (26%)	1.30	15.89	0.76 (Cig)
Jewelry, optical goods, cameras (**D**)	General stores (19%), Furniture & appliances (1%), Other (81%)	0.94	1.34	0.67 (Jwl)
Sporting goods & toys (**D**)	General stores (63%), Furniture & appliances (2%), Other (35%)	0.69	0.01	0.66 (Ot1[b])
Radios, TVs, phonographs, musical instruments (**D**)	General stores (21%), Furniture & appliances (75%), Other (4%)	0.62	2.56	0.66 (Ot1)
Office & stationery (**C**)	General stores (57%), Furniture & appliances (2%), Other (41%)	0.43	0.68	0.57 (Bks[b])

This table shows product-level data on sales, TV advertising revenue, and visibility. The product categories are those used in the 1948 *Census of Business* for their merchandise line sales data. "Share of product sales by sector" reports the distribution of each product's sales across the 10 retail sectors as defined by the *Census of Business*. (These are the same 10 retail sectors as in Figure 2.3 and Table 2.5.) "% of total sales" is each product's share of total sales across products. "% of TV ad rev" is each product's share of total TV advertising revenues in 1951-53. "Visibility index" is the normalized visibility index from Table 3 of Heffetz (2011) with the product code reported in parentheses.

[a]Heffetz (2011) distinguishes between consumption at home versus outside for food and alcohol; I take the average of home and outside consumption.

[b]Since there is no exact match to the categories in Heffetz (2011) for these products, I use close categories as proxies. E.g., I assign "Building materials" the same visibility as "Household equipment and furnishings".

Figure 2.4: TV advertising revenue and effect on retail sales



This graph plots each product's share of 1951-53 TV advertising revenues against its share of the TV effect. The share of the TV effect is calculated using the sector-by-sector decomposition in Table 2.5 and the Merchandise Line Sales data in Table 2.6 (see Footnote 17 for details). The size and shading intensity (for consumables only) are proportional to the product visibility index in the last column of Table 2.6. The lines are produced by WLS regressions where the weights are set to product visibility.

Figure 2.4 sketches the relationship between advertising, sales growth, and visibility. I plot each product's share of 1951-53 TV advertising revenue against its share of the TV effect over 1948-54. The size of the points is proportional to the product's visibility index. While the share of TV advertising revenue and the product visibility can be read directly off Table 2.6, the share of the TV effect must be derived using both Tables 2.5 and 2.6. I compute each product's share of the TV effect as the weighted average of each sector's share of the TV effect, where the weights are the product's share of sales in that sector. Each sector's share of the TV effect is calculated from Column 5 of Table 2.5.[17]

Figure 2.4 also plots the WLS regression lines for durables and consumables separately, where the weights are set to product visibility. Overall, the relationship between advertising

[17]For example, Column 5 of Table 2.5 reports that stronger sector-by-sector growth in the Pre-Freeze group led to 3.4 log points higher growth in overall retail sales than the Post-Freeze group, and that the "Automobile" sector accounted for $1.3/3.4 = 38\%$ of that difference. Then, the "Automobile" sector's share of the TV effect is 38%. Tallying similarly for the other sectors, I take the weighted average of these sector-level shares for each product, where the weights are the product's share of sales in that sector:

$$[\text{Share of TV effect}]_k = \sum_j \frac{\text{sales}_{jk}}{\text{total sales}_j} \times \frac{[\text{Col. 5, Table 2.5}]_j}{0.034}$$

k indexes products listed in Table 2.6 and j indexes the ten retail sectors in Table 2.5. The calculation of the weights $\left(\frac{\text{sales}_{jk}}{\text{total sales}_j}\right)$ requires a transformation of Table 2.6 to the sector level. For example, 95% of all "Automotive" *product* sales occurred in the "Automobile" *sector*, but the "Automotive" *product* composed only 87% of all sales within the "Automobile" *sector*.

revenue and share of the TV effect is positive, but much more pronounced for durables than consumables. Within the durables, the "Automotive" and "Household equipment & furnishings" products lead in advertising shares with around 8.5% each. The "Automotive" product alone constitutes over 35% of the TV effect on retail sales growth, and the "Household equipment & furnishings" product has a modest contribution of 6%. "Building materials", which are used for home improvement, hold a sizable 15% share of the TV effect, despite a low share of advertising revenues (<5%). Meanwhile, of the three consumable products with TV advertising revenue shares above 15%, two ("Smoking" and "Drugs & toiletries") have close to zero contribution to overall retail sales growth; "Food & drink" with 33% share of the TV effect is the outlier responsible for lifting the slope of the regression line for consumable products.

Next in Table 2.7, I formally test the patterns from Figure 2.4 in regressions at the county-product level. I regress the change in log sales from 1948-54 on the share of TV advertising revenue, an indicator for durable products, and the interaction term. I compare the effects of these variables between the Pre- and Post-Freeze groups. Specifically, the regression takes the form:

$$\begin{aligned}\Delta_{48-54}\log(\text{sales})_{ip} = \alpha + \boldsymbol{\Pi}\boldsymbol{X_i} + \theta_1 \text{TV ad}\%_p + \theta_2 \text{Durable}_p + \theta_3(\text{TV ad}\%_p \times \text{Durable}_p) \\ + \text{Pre-Freeze}_i \times [\gamma + \tau_1 \text{TV ad}\%_p + \tau_2 \text{Durable}_p + \tau_3(\text{TV ad}\%_p \times \text{Durable}_p)]\end{aligned} \tag{3}$$

where i indexes counties and p products. $\Delta_{48-54}\log(\text{sales})_{ip}$ is the change in log sales for product p in county i from 1948-54, TV ad$\%_p$ is p's share of 1951-53 TV advertising revenues from Table 2.6, and Durable$_p$ is an indicator for whether p is a durable good. As before, Pre-Freeze$_i$ is an indicator for whether county i is in the Pre-Freeze group, and $\boldsymbol{X}_i$ is the vector of county demographics from Table 5.2.

Since the Merchandise Line Sales data are at the national level and available only in 1948, I rely on two approximations. First, I transform sales from the county-sector level to the county-product level by assuming all counties have the same within-sector distribution of sales across products as in the national accounts. For example, the Merchandise Line Sales data report that 87% of national sales in the "Automobile" sector are for the "Automotive" product (see Footnote 17); I assume that the 87% of each county's "Automobile" sector sales are also for the "Automotive" product. Second, I require that these proportions did not change from 1948 to 1954. This second assumption would be violated if, for instance, consumers shopped less at apparel stores after 1948 and bought a higher share of their clothes at department stores in 1954. These two assumptions are admittedly naïve yet necessary given the lack of finer-grain historical data. Accordingly, the results in Table 2.7 are noisy and should be interpreted for their sign and relative magnitudes, not for exact estimates. Nevertheless, the results are consistent with the qualitative literature, which I revisit in Section 2.6.

Column 1 begins by estimating the effect of TV advertising revenue by itself. The coefficient is negative for the Post-Freeze group (-0.16, s.e.=0.07), and zero for the Pre-Freeze group ($-0.16 + 0.15 = -0.01$, s.e.=0.08). Without separating the effect by consumables and durables, we would conclude that advertising had no relation to TV-driven growth in retail sales.

Column 2, however, reveals distinct patterns for consumables and durables. Unexpectedly, consumable products that attracted higher TV advertising revenues had lower sales

growth both in Post-Freeze counties (coef.=-0.31, s.e.=0.06) and to a lesser extent in Pre-Freeze counties (coef.=-0.17, s.e.=0.07). Yet this negative relationship among consumables is driven by the "Fuels & lubricants" product, which is an obvious complement to the booming "Automotive" product in the durables group. Sales for "Fuels & lubricants" grew by 25 log points on average—the most among consumables, followed by 12 log points for "Alcohol"—but had only 1.78% of the TV advertising revenues. To illustrate its influence, Column 3 omits the "Fuels & lubricants" product, which attenuates the negative correlation among consumables for Post-Freeze counties (coef.=-0.16, s.e.=0.07) and eliminates it for Pre-Freeze counties (coef.=-0.01, s.e.=0.08).

For durable goods, on the other hand, Column 2 finds a strong and positive relationship between TV advertising and growth in sales in the Pre-Freeze group (coef.=1.46, s.e.=0.29), but a null effect in the Post-Freeze group (coef.=0.25, s.e.=0.34). Furthermore, the intercept on durable goods in the Pre-Freeze group is estimated as zero: TV access increased sales only for durable goods that were advertised.

Columns 4-6 present robustness checks. The results remain the same when I include county characteristics in Column 4, when I weight observations by the product's visibility index in Column 5 (which drops products that cannot be matched to the index; see the last column of Table 2.6), and when I include county fixed-effects in Column 6.

These results present empirical evidence for the most prominent mechanisms discussed in the literature, but there are many other potential mechanisms that the data do not preclude. For instance, as people spent more time watching TV at home, the value of domestic commodities and residential features such as house size may have increased relative to local urban amenities. As a result, they may have bought cars to live in spacious suburban houses or building materials and household appliances to improve their homes. Such effects would not be due to advertising or conspicuous consumption, but simply due to an improvement in private entertainment technology.

2.5.5 Television and the 1950s American Dream

In addition to consumerist standards of possessions such as the "proverbial two-cars-in-the-garage" (Samuel, 2012), historians describe the public perception of the American Dream evolving in the 1950s to evoke a distinct lifestyle—"the suburban housewife with an up-and-coming husband and a station wagon full of children" (Friedan, 1963). Television was a particularly potent medium at disseminating this rendition of the American Dream (Halberstam, 1994; Samuel, 2001). Classic sitcoms such as *I Love Lucy* (1951-57), *The Adventures of Ozzie and Harriet* (1952-66), *Make Room for Daddy* (1953-71), *Father Knows Best* (1954-63), and *Leave It to Beaver* (1957-63) emphasized suburban living and formulaic gender roles within the nuclear family (Lipsitz, 1990; Weiss, 2000; Cohen, 2003; Samuel, 2012; Spigel, 2013) and "brought the fifties family into living rooms across the nation" (Mintz and Kellogg, 1988). Soon, "the new suburban family ideal was a consensus ideology" (Spigel, 2013).

Table 2.7: Advertising, durables, and the TV effect (1948-54)

Dep. Var.: Δlog(sales) from 1948-54	(1)	(2)	(3)	(4)	(5)	(6)
Share of TV advertising revenue (1951-53)	-0.16	-0.31	-0.16	-0.32	-0.24	-0.32
	(0.07)	(0.06)	(0.07)	(0.06)	(0.06)	(0.06)
Durables		-0.09	-0.05	-0.09	-0.09	-0.08
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Durables × Share of TV ad rev.		0.57	0.41	0.56	0.55	0.47
		(0.31)	(0.30)	(0.31)	(0.30)	(0.30)
Pre-Freeze group	0.01	-0.01	-0.01	-0.01	-0.01	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	
Pre-Freeze group × Share of TV ad rev.	0.15	0.14	0.15	0.14	0.15	0.14
	(0.11)	(0.10)	(0.10)	(0.10)	(0.09)	(0.10)
Pre-Freeze group × Durables		-0.00	-0.00	-0.00	0.00	-0.01
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Pre-Freeze group × Durables × Share of TV ad rev.		1.07	1.06	1.07	1.07	1.19
		(0.40)	(0.39)	(0.40)	(0.40)	(0.39)
Constant	0.07	0.11	0.07	-1.23	0.10	0.11
	(0.02)	(0.01)	(0.02)	(0.45)	(0.01)	(0.00)
Excluding "Fuels & lubricants"			✓			
All controls				✓		✓
Weighted by Heffetz (2011) visibility index					✓	
County FE						✓
Observations	40553	40553	37595	40539	37638	40539
Counties	2962	2962	2960	2961	2962	2961
Media markets (DMA)	199	199	199	199	199	199
R^2	0.00	0.00	0.00	0.02	0.00	0.00

This table shows estimates from OLS regressions. The dependent variable is the change in log sales over 1948-54. Observations in these regressions are at the county-product level. "All controls" include all the demographic variables from Table 5.2 and census-region fixed effects. See Table 2.6 for more details on the categorization of durable goods, the visibility index, and the share of TV advertising revenues. Standard errors clustered by DMA are shown in parentheses.

This section explores the effect of TV on broader dimensions of the 1950s American Dream. I begin with highway construction, which is a natural segue following the result on car sales and is widely considered a precursor for suburbanization. Historical accounts stress the central contribution of cars and highways to mass suburbanization in the postwar era (Samuel, 2001; Cullen, 2003). In economics, Baum-Snow (2007) and Kopecky and Suen (2010) deliver empirical evidence for this narrative. Suburbanization itself has typically been studied as a long-run phenomenon measured across decades,[18] which renders it difficult to link with a few years of exogenous TV exposure during the FCC Freeze. Highways, on the other hand, have been constructed incrementally since the 1940s. Furthermore, Baum-Snow (2007) provides annual, county-level data for highway construction during this era, but not for suburbanization.

The historical context of highway expansion also presents advantages for isolating demand-driven effects. As mentioned in Section 2.3.3, Baum-Snow (2007)'s empirical strategy uses the 1947 national highway plan, which was constructed over decades to form the modern-day

[18]For example, Baum-Snow (2007) compares suburban populations in the U.S. between 1950 and 1990, and Boustan (2010) does so between 1940 and 1970.

system.[19] The timing of events is fortuitous in the context of this paper. Since the 1947 plan was drafted before the FCC Freeze in 1948, the planned supply of highways was assigned before it was known which counties would gain TV access before or after the Freeze. Furthermore, federal funding for highway construction was sparse until the Highway Revenue Act of 1956. Consequently, Baum-Snow (2007) interprets highway construction before the federal funding act as signs that "state and local governments adjusted metropolitan area highway infrastructure at least partly in response to local commuting demand."

To begin with the raw trends, Figure 2.5a graphs the average miles of highway construction started or completed over 1947-58 for the Pre- and Post-Freeze groups.[20] Both groups exhibit the same pattern of exponential growth in highway construction activity, but the start of the upward trajectories coincides with the onset of TV access for each group. The growth begins around 1948 during the Freeze period for the Pre-Freeze group, but only a few years later in 1953 after the Freeze ends for the Post-Freeze group.

Leveraging the annual observations in the dataset, Figure 2.5b plots the event study for the miles of highway construction during 1940-60 around the event of gaining TV access. I include events occurring until 1958, which aligns with the last year in the retail sector analysis.[21] The figure plots four leads and lags of the μ_l coefficients in the following equation:

$$\text{Highway miles}_{it} = \alpha_i + \lambda_t + \sum_l \mu_l \mathbf{1}\{t - E_i = l\} + \nu_{it}$$

The dependent variable Highway miles_{it} is the number of highway miles with construction started or completed in county i by year t, and the variable E_i is the year of the event for gaining TV access. Given the issues with traditional two-way fixed-effects (TWFE) estimators raised in the recent econometrics literature (e.g., Callaway and Sant'Anna, 2020; Chaisemartin and D'Haultfoeuille, 2020; Sant'Anna and Zhao, 2020; Sun and Abraham, 2020; Goodman-Bacon, 2021), I use Callaway and Sant'Anna (2020)'s estimator, which is robust to dynamic treatment effects and staggered treatment timing.[22] The coefficients for the years after the event ($t \geq E_i$) are estimated relative to the year before the event ($E_i - 1$), and the coefficients for the years before the event ($t < E_i$) are estimated relative to the preceding year ($t - 1$).[23] Standard errors are clustered by DMA.

[19] Baum-Snow (2007) explains that in the legislation for the plan, the supply of highways was allocated to promote interregional trade and military operations, rather than to develop suburban areas. The dataset contains only the subset of counties on the 1947 plan, but the main results on retail sales can still be replicated within this subset (see Column 6 of Table C.6).

[20] There was no highway construction in the Post-Freeze group before 1947.

[21] 1238 out of the 1250 Pre- and Post-Freeze counties in the Baum-Snow (2007) dataset gain TV access before 1958.

[22] Estimates are obtained using the `csdid` package in Stata written by Rios-Avila, Sant'Anna, and Callaway. Figure A.12a shows the estimates from the traditional TWFE method. Although the results are not starkly different, the Goodman-Bacon (2021) decomposition in Figure A.12b finds that the half of weights associated with negative effects in the simple TWFE difference-in-differences estimator are contaminated by "problematic" comparisons that use already treated counties as controls for counties that gain TV access later. Therefore, the Callaway and Sant'Anna (2020) estimator is the preferred approach to reliably estimate both the pre-trends and the dynamic treatment effect.

[23] In traditional TWFE event studies, the coefficient for the time period before the event is typically omitted and normalized to 0. The Callaway and Sant'Anna (2020) estimator, on the other hand, does not require an omitted base period.

Figure 2.5: Access to TV and highway construction

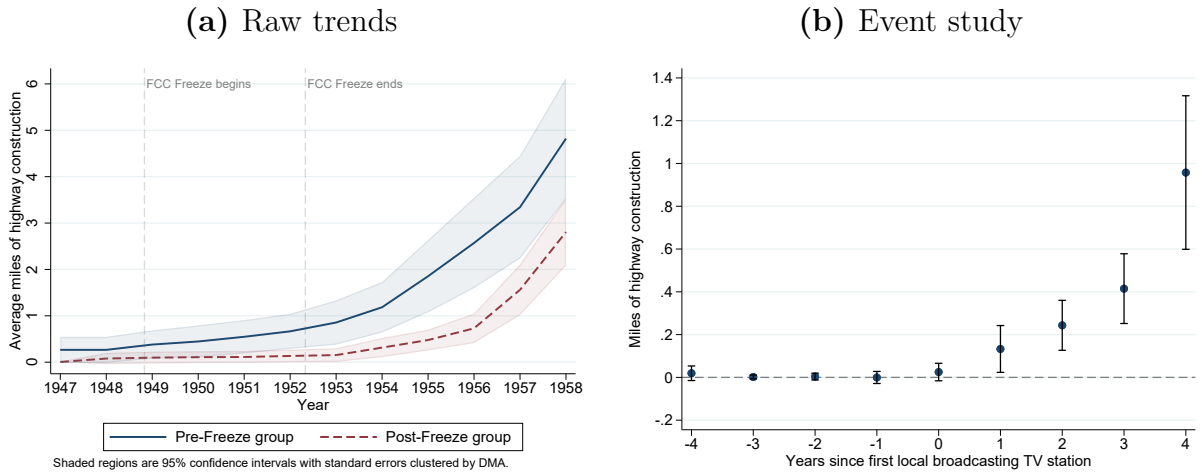


Figure 2.5a shows the average miles of highway construction started or completed across years separately for the Pre- and Post-Freeze groups. The sample consists of the 1238 counties in the Pre- and Post-Freeze groups that have highway data from Baum-Snow (2007) and gain TV access by 1958.
Figure 2.5b plots the event-study coefficients around the year of gaining TV access. Estimates are from the Callaway and Sant'Anna (2020) estimator. Coefficients for the years after the event are estimated relative to the year before the event, and the coefficients for the years before the event are estimated relative to the preceding year. The dataset includes every year from 1940-60. Standard errors for the 95% CIs are clustered by DMA. Year and county fixed effects are included.

The event-study results in Figure 2.5b mirror the trends from the raw data. Before TV access, there is minimal highway construction and the pre-trends are precisely estimated to be nonexistent. After TV access, highway construction expands year by year. Figures A.12c-A.12e first perform robustness checks. The trends remain intact accounting for base-year demographic controls (Figure A.12c), using the log of highway miles as the outcome (Figure A.12d), and only including years up to 1952 to isolate the exogenous variation in TV exposure during the FCC Freeze (Figure A.12e).

To examine the source of the effect, Figures A.12f-A.12h test different dependent variables and time periods. The outcome in Figure A.12f is a binary indicator for whether the county has any highway construction started. This exercise reveals that the effect is not merely driven by an expansion of existing construction projects, but also comes from counties building local highways for the first time. Figures A.12g and A.12h investigate whether the effects arise from the supply or demand side. Figure A.12g modifies the dependent variable to be the miles of current highway construction as a percentage of all the miles assigned to the county in the 1947 plan. This outcome controls for any systemic differences in the planned supply of highways between counties that gained TV access earlier versus later.[24] Meanwhile, Figure A.12h only includes years up to 1956, prior to the federal funding act, when highways were built mainly using state and local funds likely in response to local commuting demand (Baum-Snow, 2007). Both figures exhibit similar increases following TV access and indicate a demand-driven effect.

[24]Table A.7 finds a small negative relationship between the year of gaining TV access and the planned supply of highways across counties when all demographic controls are included.

In the last empirical analysis of this paper, I consider the effect of TV on birthrates. In the economics literature, Jensen and Oster (2009), La Ferrara, Chong, and Duryea (2012), and Kearney and Levine (2015) all find a *negative* effect of TV on fertility across various contexts. Importantly, these papers posit that the effect reflects the content of the popular shows. Jensen and Oster (2009) and La Ferrara, Chong, and Duryea (2012) respectively describe that cable TV shows in India from 2001-03 and novelas in Brazil from 1979-91 depicted progressive themes of upward mobility, female autonomy, and small families. Kearney and Levine (2015) focus on the show *16 and Pregnant*, which the authors explain as portraying the difficulties of teen pregnancy. In sum, the content of the TV shows in these settings seem to transmit information, family values, or cultural norms that persuade *against* raising more children.

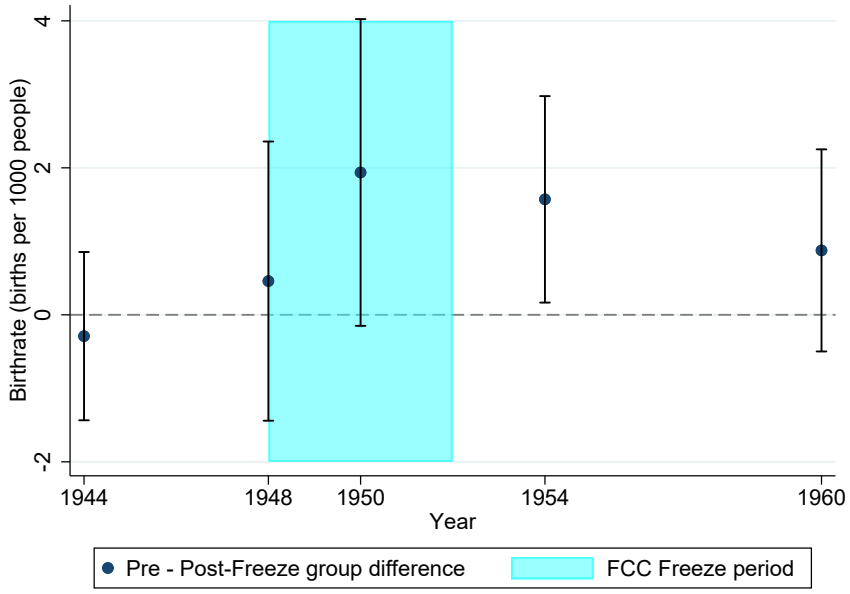
In contrast, U.S. historians deem the 1950s as a period when the prevalent interpretation of the American Dream became envisioned as a domestic lifestyle under the homogeneous ideal of a "male breadwinner and a stay-at-home female homemaker and their two or three children" (Weiss, 2000). Television's role as the emerging centerpiece of popular media during this era was to romanticize this image (Friedan, 1963; Mintz and Kellogg, 1988; Lipsitz, 1990; Samuel, 2012).[25] American historians Mintz and Kellogg (1988) claim that with this cultural regression in gender roles, "Managing a home in the suburbs and raising a family became the goal of most women during the 1950s."

Figure 2.6 shows the raw difference in birthrates between the Pre- and Post-Freeze groups for every recorded year until 1960. Much like the results in the retail sector, there are no differences before the FCC Freeze. During the Freeze, however, Pre-Freeze counties have a marginally higher average birthrate than the Post-Freeze counties. In 1954, the difference has become significant, but tapers off by the end of the decade when TV has spread nationwide. These differences are even more precise under various controls and parametric specifications for TV exposure, reported in Table A.8. This positive relationship between TV and fertility runs counter to previous studies (Jensen and Oster, 2009; La Ferrara, Chong, and Duryea, 2012; Kearney and Levine, 2015), but accentuates the differences in the content of the TV shows and the historical setting. Another explanation is that since television dominated leisure in America during the 1950s for both sexes (Barnouw, 1990; Fenton and Koenig, 2021), births naturally increased from couples spending more time together indoors without modern contraceptives (i.e., an anti-"incapacitation effect", Black, Devereux, and Salvanes, 2008).

[25] One notable example is from the hit show *I Love Lucy*, which revolved around a young middle-class housewife named Lucy. In the second season (1952-53), the real-life pregnancy of starring actress Lucille Ball who played Lucy was weaved into the fictional storyline. When Lucille gave birth in real life, her on-screen character Lucy also gave birth in the show to her first child (TIME, 1953; Barnouw, 1990).

Samuel (2012) writes: "By endorsing traditional gender roles, popular television shows of the 1950s such as *I Love Lucy, Make Room for Daddy,* and *Father Knows Best* served as lessons in a somewhat different but equally powerful Dream, with happiness and success seemingly waiting for those Americans who conformed to what a postwar family should look and act like."

Figure 2.6: Access to TV and birthrate



This figure shows the difference in birthrate (births per 1000) between the Pre- and Post-Freeze groups for the years with birth data available. Estimates for county populations between decennial census years are interpolated assuming a constant growth rate. 95% confidence intervals are shown with standard errors clustered by DMA.

2.6 Discussion and conclusion

This paper documents television's impact on the postwar retail sector through three findings. First, retail sales in counties with TV access during the unexpected FCC Freeze grew by 3-4% more on average over 1948-54 than in counties without TV access. This difference, which appears to be demand-driven, translates to an additional $131 in retail spending for the median household, or around $1300 today. Second, the automobile sector alone accounts for roughly a third of the difference, or around $40 per household, when the average price of a new car was around $1800. Third, the more money that was spent advertising a product on TV, the more the product grew in sales—but only for durable goods and, naturally, only in counties that could watch TV. Beyond the retail sector, this paper finds that access to TV precipitated highway construction and births, complementing previous studies that have also found far-reaching effects of TV in postwar American society on outcomes such as voter turnout (Gentzkow, 2006) and labor supply (Fenton and Koenig, 2021).

These results meld remarkably well with the broader scholarly work on this era. The 1950s in America commenced a change in the culture of consumption, galvanized by an expanding middle class and a common spirit of civic duty to fuel the American economy by buying more commodities (Cohen, 2003; Edgerton, 2007). Departing from Depression frugality and wartime scarcity, the American masses began to consume for wants rather than needs, and these materials wants—most often "a variety of household furnishings and appliances and, of course, a new car"—became "a sign of progress and a matter of status" (Berkin et al., 2011). In such a society embracing conspicuous consumption, Americans

began buying products, especially durable goods, not only for function but also for social status (Samuel, 2012).

Cars in particular became the status symbol par excellence. *Advertising Age* (2003) recounts, "Car owners of the mid-1950s began to see their vehicles as extensions of themselves. Ever-changing designs pushed consumers to replace their cars yearly—not for lack of performance, but for lack of style." Katona affirms in 1964, "In earlier years the car had been recognized, not only as a highly valued possession, but also as the family's most visible status symbol. It was then argued that Americans bought new and fancy cars to keep up with the Joneses[....]In addition, the kind of car selected was believed to reflect the buyer's hidden personality traits."

Television took stage in this era as an unrivaled medium to entertain and advertise to the masses on an unprecedented scale. Samuel (2001) singles out automobile manufacturers in that they "relied heavily on television advertising to get people into local showrooms for a test drive." American historian Lipsitz (1990) recounts a memorable episode of the sitcom *The Goldbergs* (1949-56) in which the mother is persuaded by her children that buying a second car on credit was "the American way".[26] Even popular quiz shows such as *The $64,000 Question* contributed to conflating consumerism with the American Dream (Smith, 1990).[27] Central to television's viability in its early days was its success as the "selling machine in every living room"[28]—and sell it did. While we cannot attribute the rise of American consumerism in the 1950s to TV alone, we can contend that TV was indeed a catalyst, a medium that still influences what we want and buy today.

[26]Full quote: "Conversations about shopping for new furniture in an episode of *The Goldbergs* directed viewers' attention to furnishings in the Goldberg home provided for the show by Macy's department store in exchange for a commercial acknowledgment. The content of the shows themselves offered even more direct emphasis on consumer spending. In one episode of *The Goldbergs*, Molly expresses disapproval of her future daughter-in-law's plan to buy a washing machine on the installment plan. 'I know Papa and me never bought anything unless we had the money to pay for it,' she intones with logic familiar to a generation with memories of the Great Depression. Her son, Sammy, confronts this 'deviance' by saying 'Listen, Ma, almost everybody in this country lives above their means—and everybody enjoys it.' Doubtful at first, Molly eventually learns from her children and announces her conversion to the legitimacy of installment buying, proposing that the family buy two cars in order to 'live above our means—the American way.'"

[27]In a biography of William Paley, the chief executive of CBS during this era, Smith (1990) commentates, "As such, the quiz show meshed perfectly with the materialism of the times. The postwar economy was booming, and wages were rising rapidly. Consumers bought homes, cars, household appliances, novelties, and luxury items. Installment debt reached new heights. And television advertising both celebrated and fanned the buying binge. When quiz show winners marched off with their loot to splurge on Jaguars and mansions, they were fulfilling the American Dream—a dream defined by television."

[28]Quote from NBC, "Strangers into Customers," 1953, excerpted from Halberstam (1994) and Cohen (2003).

Chapter 3

"Don't you plan on voting?" The motives and effects of peer pressure in voter mobilization

Online peer pressure has emerged as a potential tool to mobilize young American voters, but it remains unclear whether these voters are willing to pressure their peers and how effective it would be. This chapter features an online field experiment, in which college students choose whether to send an email that pressures their campus peers to register to vote for the 2020 U.S. Elections. The senders have a negative average willingness-to-pay (WTP) to send the email, meaning they would rather pay than pressure their peers. The experiment reveals two counteracting motives at play. On the positive side, senders believe that pressuring their peers will persuade them to register. On the negative side, they believe that their peers will dislike being pressured, which outweighs the positive motives. However, under anonymous messaging, they become much less sensitive to whether the recipient dislikes being pressured, and their average WTP turns positive. This pattern of behavior reflects voters who are motivated by the prospect of mobilizing their peers, but are restrained by self-interested concerns from social image costs or retaliation.

3.1 Introduction

Messages exerting social pressure—for example, a letter informing neighbors whether you vote—have been especially effective at mobilizing turnout (Gerber and Green, 2017). Yet campaigns avoid sending such messages themselves due to the backlash that follows (Matland and Murray, 2012; Green and Gerber, 2019). Recent efforts to mobilize the pivotal yet chronically lagging young American vote have instead harnessed decentralized voter-to-voter communication via social media and "vote-shaming" apps, where individuals can voluntarily monitor and pressure their peers.[29] This kind of peer engagement has played an important role in theoretical models of voter participation (Ali and Lin, 2013; Levine and Mattozzi, 2020) and has also demonstrated potential in empirical studies (Bond et al., 2012; Haenschen, 2016; Schein et al., 2021). Despite the growing presence of online peer pressure in voter mobilization and enforcement of prosocial behavior in general, we know little about the motives involved. Do those who pressure others believe they are being persuasive? Do they simply want to shame those who don't participate? Are they actually effective?

This paper studies the motives behind pressuring peers to register to vote as well as the effects of the peer pressure on mobilizing voters through an online framed field experiment. With over 1000 participants across four University of California campuses, students who had registered to vote (the Senders) are each matched with three campus peers who had yet to register (the Recipients) a few weeks before the U.S. 2020 Presidential Election. The experiment provides the Senders a template of a pressuring email and elicits their willingness-to-pay (WTP) to send the message to each of the Recipients as the main outcome. The incentive-compatible WTP elicitation provides novel measures of the intensity in preferences both for pressuring (WTP>0) and against pressuring (WTP<0). The experiment also captures the Senders' incentivized beliefs about the instrumental and affective effects of the message on the Recipients.

Senders typically prefer *not* to pressure their peers and have a negative average WTP of -\$0.42 (SE=0.18), but this average masks two counteracting motives revealed in their beliefs. Both the Sender's beliefs about the instrumental and affective effects of their message are positively correlated with their willingness to send the email. Senders who believe that their message is more likely to persuade their Recipient to register to vote (the instrumental effect) or that the Recipient will like their message more (the affective effect) have a significantly higher WTP to send the email. While Senders on average expect their email to be highly persuasive, they also anticipate the Recipient will dislike the message. The latter belief appears to outweigh the former, tipping the average willingness to pressure into the negative.

How should we interpret these patterns? Senders find it costly to pressure their peers—is that evidence of altruism, since most Senders expect the Recipient to dislike the message, or is it due to self-interest? Which of these findings are specific to the context of voter mobilization, as opposed to persuading peers for other purposes?

To further understand the Senders' motives for peer voter mobilization, the experiment features two treatments. The first varies within-Sender and tests the effect of anonymity.

[29]Some examples are Impactive (formerly OutVote), VoteWithMe, and OutreachCircle (formerly VoterCircle). News coverage of these apps include "Can peer pressure defeat Trump?" (*The New York Times*, 2019) and "Your mother, your plumber, your boss—this app tells you their voting records and party affiliations" (*Fortune*, 2018).

The experiment randomly varies whether each of the three Recipients matched to a Sender must be emailed directly (from the Sender's own campus email account) or anonymously (in which case the research team delivers the same email without the Sender's name). The changes under anonymity help to interpret the correlations between the Senders' preferences and beliefs.

Under the anonymity treatment, Senders switch from being willing to pay \$0.42 on average to avoid sending the email directly, to be willing to pay \$0.78 to send the email anonymously instead (diff.=\$1.21, SE=0.16). Communicating anonymously diminishes the cost of pressuring to the extent that the average WTP flips sign from negative to positive, even after controlling for any changes in the beliefs. Senders evidently would like to pressure their peers to register to vote—as long as they don't have to reveal their identity.

This higher willingness to pressure others under anonymity is driven by the majority of Senders who believe the Recipient will dislike the message. They have strong preferences against sending directly, but become indifferent when communicating anonymously, even though they believe the Recipient would dislike the message. This finding implies that the Senders' reluctance to pressure is not primarily motivated by altruism. The likelier explanation lies in self-interested motives, such as avoiding retaliation or social costs from sending an unsolicited, unwelcome message.

Senders are also more sensitive to whether the Recipient will dislike the message when the Recipient is male or someone they know. This heightened sensitivity toward male Recipients only appears when Senders are communicating directly, not anonymously. This distinction indicates that Senders perceive greater social costs or risks of backlash from pressuring a male Recipient. Meanwhile, Senders are more considerate of Recipients whom they know even under anonymity, which suggests stronger altruistic preferences toward personal acquaintances.

The second treatment varies between-Senders and changes the topic of the communication. Instead of persuading the Recipient to register to vote, some Senders are instead assigned to send a "placebo" informative email encouraging Recipients to look up their state assembly district representatives on an online portal. This treatment determines which motives are particularly relevant to the context of peer voter mobilization, as opposed to a merely informative, less pressuring communication for a lower-stakes purpose.

Senders assigned to this informative email treatment still prefer to communicate anonymously, but are significantly less sensitive to their beliefs on both the instrumental and affective effects of their message. Although they generally believe their message will successfully persuade the Recipient to learn who their district representatives are, there is zero correlation between their predicted persuasive effect and their willingness to send the email. They seem to place no value on the instrumental effect of this message, implying that persuading the Recipient to look up district representatives is a trivial objective. They are also less sensitive to whether the Recipient will dislike their message when communicating directly. As this message is less pressuring in tone, they may be less concerned about backlash.

The experimental design also identifies the causal effect of the peer pressure email on the Recipient's actual likelihood of registering to vote and estimates a *negative* point estimate. The estimate is imprecise given the small sample of Recipients, but we can at least rule out the positive optimistic effects that the Senders were expecting. The effects are uncorrelated with the Senders' willingness to send the email and beliefs about their own persuasiveness.

Taken altogether, the experiment finds that registered voters in this sample of young Americans—a population with historically lower turnout—believe pressuring their unregistered peers would be an effective means of mobilization and do indeed have positive preferences for doing so. This is perhaps surprising considering the context of this experiment in California under the U.S. Electoral College system. California has become a liberal stronghold in the last couple decades, with betting odds at 97% for casting its electoral votes to the Democratic candidate in the 2020 Presidential Election (Jones, 2020). Persuading another student to vote, especially one from the University of California system where only 9% identify as conservative (UCUES, 2020), would essentially have no impact on the outcome of the election. For the voters in this sample, the positive motive from the instrumental effect of pressuring peers to participate is thus unlikely to represent an effort to change the outcome of the election toward their preferred candidate. Instead, the motives seem more related to expressive voting, norm enforcement, or civic duty (Brennan and Hamlin, 1998; Feddersen, 2004). In other contexts such as swing states, the potential of influencing which candidate is elected could further enhance the value of the instrumental effect.

Nevertheless, these same voters are also apprehensive about the social costs and risks involved and are ultimately reluctant to pressure their peers directly. One implication is that enabling anonymous communication would increase activity on peer voter mobilization platforms, but at the cost of making participants less mindful about the negative effects of pressure on their peers. Before scaling such efforts, the lack of an effect detected in this study warrants further research to determine whether such messaging is effective enough to justify the welfare costs of giving and receiving peer pressure. Furthermore, the results caution that those who self-select into voluntarily pressuring their peers are likely to be misinformed about their own persuasive ability and may be less cautious about pressuring groups such as women who may seem less likely to retaliate.

In addition to these practical implications for the get-out-to-vote literature, these results relate to broader studies on social pressure, anonymity, and prosocial behavior. In a variety of contexts such as voting, working dutifully, and donating, a series of field experiments have found that people behave more scrupulously when their actions are public.[30] This increase in prosocial behavior, however, does carry welfare costs from being observed by others (DellaVigna et al., 2017; Butera et al., 2022). This paper documents that in addition to the costs of receiving social pressure, there are consequential social costs for directly giving social pressure as well. These results also deepen our understanding of the psychology behind the costs of pressuring peers that are assumed in theoretical models not only of voter mobilization (Levine and Mattozzi, 2020) but also in other contexts such as the workplace (Kandel and Lazear, 1992).

Another set of studies have focused on the mechanism of peer enforcement for prosocial behavior. For instance, in public goods lab games, cooperators punish others at a monetary cost to prevent free-riding (Fehr and Gachter, 2000, 2002). In real life, we often express such enforcement verbally (Ostrom et al., 1992; Bowles and Gintis, 2002). This paper adds

[30]Examples include Bandiera et al. (2005); Gerber et al. (2008); Mas and Moretti (2009); Funk (2010); DellaVigna et al. (2012); Burszytn and Jensen (2015); Perez-Truglia and Troiano (2017); Karing (2023). See Burszytn and Jensen (2017) for a review.

to a group of field experiments that find only a minority of people are willing to speak up when someone behaves antisocially (e.g., Balafoutas and Nikiforakis 2012; Balafoutas et al. 2014). Relative to these studies, this paper captures the distribution in the willingness to communicate pressure, elicits novel measures of important beliefs on the instrumental and affective effects of such pressure, uses actual peers who have yet to take the prosocial action instead of confederates as the targets, and tracks the downstream causal effects of the pressure on the targets' prosocial behavior.

3.2 Experimental design and sample

3.2.1 Overview and logistics

This experiment was conducted a few weeks before the voter registration deadline in California for the 2020 U.S. General Election. A total of 1,059 students who were eligible to register online to vote in California were recruited through behavioral labs from four University of California campuses.[31] In addition to the base compensation of $5, participants could earn up to $7 from the willingness-to-pay (WTP) elicitation and $1 for correct guesses as an incentive to truthfully report their beliefs.

The experimental design and randomization process were all pre-registered (AEARCTR-0006364), except for the Post-election Survey in Section 3.3.3, which was conducted after the main study to better interpret the (surprising) lack of an effect from peer pressure.

Figure 3.1 shows the overview of the experiment. Participants took two Qualtrics surveys within 10 days: (i) the Initial Survey and (ii) the Follow-up Survey. In the Initial Survey, participants reported whether they had registered to vote before starting the study. After their self-reported registration statuses had been verified, the registered participants were all assigned to be "Senders" and faced the decision whether to send an email to other participants. The non-registered participants were all assigned to be "Recipients". The surveys are described in chronological order next.

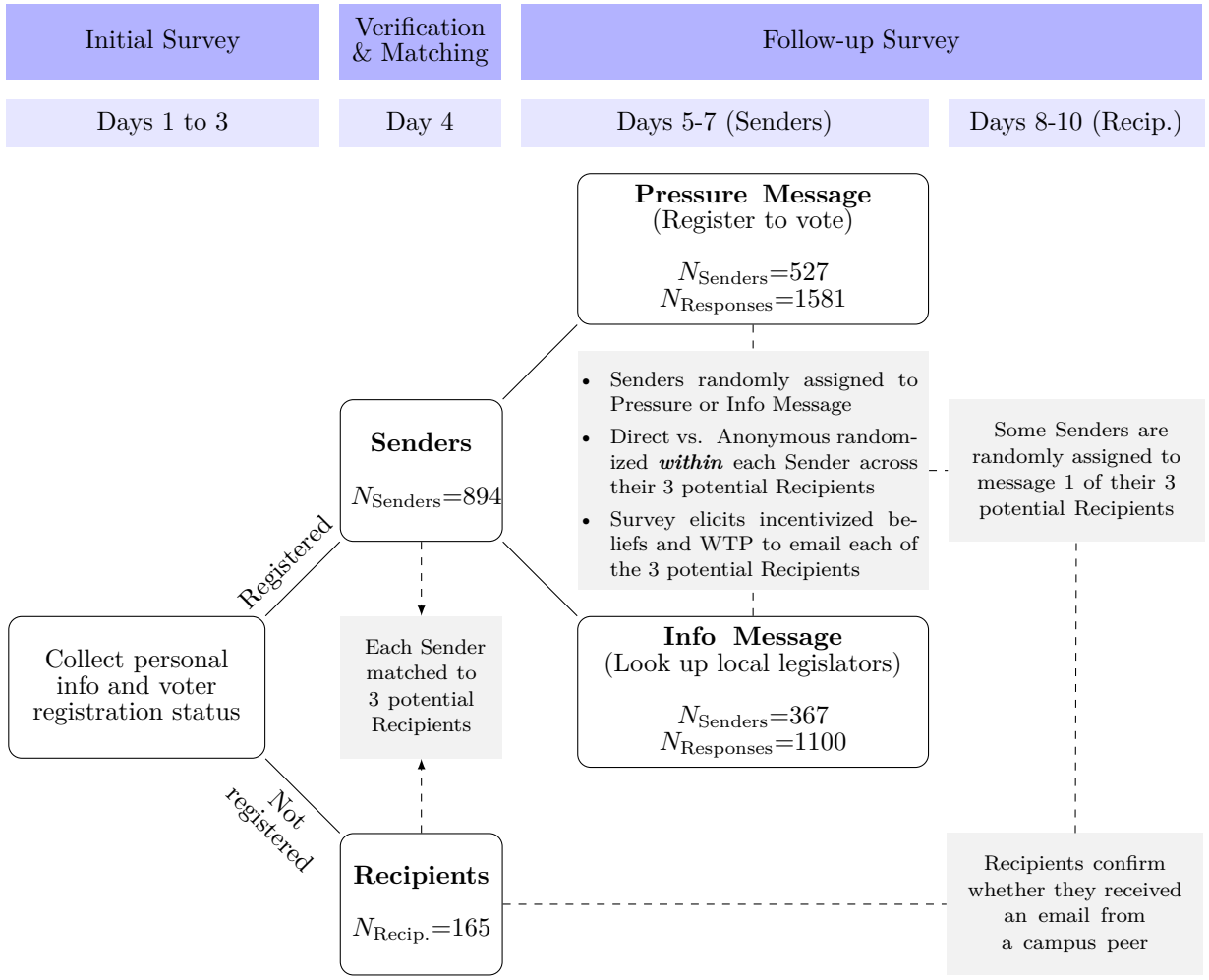
3.2.2 Initial Survey

The Initial Survey began by asking participants whether they were registered to vote under their current residence (as required by Californian law) prior to the study. The survey informed participants that the goal of the experiment was to study "peer-to-peer communication among students about the election" (without specifying voter registration), and for this purpose, their full name, voter registration status, and email address could be shared with other participants who may send them an email. The survey was carefully phrased to leave the aim of the study as ambiguous as possible to mitigate experimenter demand effects while providing adequate disclosure for informed consent.[32]

[31] To be eligible, participants (i) needed a valid California Driver License or State ID Card, and had to be (ii) U.S. citizens, (iii) Californian residents, and (iv) over the age of 18.

[32] For the non-registered participants, the survey also elicited their incentivized willingness-to-pay to not share their information with other participants. Further details are in Appendix Section B.1.1.

Figure 3.1: Experimental design



3.2.3 Verifying registration statuses and matching Senders to Recipients

Before the Follow-up Survey, the self-reported registration statuses were verified against the official California voter file. The registered participants (the Senders) were then randomly matched with three other participants on their campus (the potential Recipients). While a single Recipient could be matched to multiple Senders, the randomization procedure ensured that a Recipient could only receive at most one message from a Sender.

The randomization and matching procedure is described in Appendix Section B.1.2. The procedure was designed to balance the experimental arms and maintain incentive-compatibility for all preference and belief elicitations.

3.2.4 Follow-up Survey: Senders (Registered)

The Follow-up Survey was first sent to the registered participants all assigned to the role of Senders. Each Sender saw the names and registration statuses of three other participants on their campus. They characterized their relationship with each of them as either strangers,

acquaintances, friends, or close friends.

Between-senders condition: Message content (Pressure vs. Info). The surveys then informed the Senders that they could be randomly selected to email one of the three other participants. The content of the email was randomized between Senders from one of two templates. The first email template was labeled "Pressure Message" and was described as "a pressuring email message about registering to vote." The second email template was labeled "Info Message" and was described as "an informative email message about local legislative districts." As shown in Figure 3.2, the survey displayed the template and instructed that if the Sender was selected to send the message, they could edit it within narrow guidelines.

Within-sender condition: Message delivery (Direct vs. Anonymous). Next, the survey introduced two different modes of delivery—Direct and Anonymous—randomly varying within-Sender for each of the three potential Recipients. Each Sender was assigned at least one Recipient under the two different modes of delivery (i.e., either 2 Direct and 1 Anonymous or 1 Direct and 2 Anonymous). For a Direct Message, the Sender would have to send the email from their own campus account. For an Anonymous Message, the Sender would submit the email to the research team who would then send it to the Recipient without mentioning the Sender's name. The survey displayed whether the Sender would have to email each Recipient directly or anonymously (Figure B.2).

Belief elicitations. The survey then elicited two incentivized beliefs on the effects of the message for each of the three potential Recipients. The first belief elicitation asked Senders to predict the *affective* effect of the message on each Recipient. Senders indicated on a (strongly) agree to (strongly) disagree Likert scale how much they thought the Recipient would like to receive the message from them (Figure B.3).[33] This prediction was incentivized by informing the Senders that if they were selected to send the message, the Recipient would also be asked on the same Likert scale how much they actually liked receiving it, and if the answers were the same, the Senders could earn an extra $1.

The second belief elicitation asked Senders to predict the *instrumental* effect. Senders predicted the likelihood in percentages of each potential Recipient taking the recommended action in the message. In the Pressure Message, the recommended action was registering to vote by the election, and in the Info Message, it was knowing the names of the district representatives. The Senders made two predictions for each potential Recipient. They predicted the Recipient's chances of taking the action first without any message, and then with the message from the Sender (Figure B.4). The difference in the two predictions measures the Sender's beliefs on the persuasive instrumental effect of the message.

[33]Senders answered this prompt on a 7-point Likert scale (Strongly disagree, Disagree, Somewhat disagree, Neither agree nor disagree, Somewhat agree, Agree, Strongly agree). As arranged in the pre-registration plan, I merge responses for "Strongly disagree" and "Disagree" together under "(Strongly) disagree", and likewise for "Strongly agree" and "Agree" due to the limited number of responses in the extremities.

Figure 3.2: Message template and guidelines

(a) Pressure Message

If you're selected to send a message, you'll be provided the **template message** below (no need to copy it now), which you can **edit** within general guidelines. The message must:

- Be addressed to the participant by name
- Mention that the participant hasn't registered to vote yet and suggest that they register
- Include the voter registration link (registertovote.ca.gov) and deadline (Oct. 19)
- Have NO partisan content or personal attacks
- Be within 500 characters

Template:

Subject: You haven't registered to vote yet!

Hi <Participant>,

I'm also participating in the 2020 Election Study, and I saw that you haven't registered to vote yet as of this week.

Don't you plan on voting? Then you should register ASAP at registertovote.ca.gov.

The deadline to register online is October 19. There's no excuse not to—it only takes 5 minutes!

(b) Info Message

If you're selected to send a message, you'll be provided the **template message** below (no need to copy it now), which you can **edit** within general guidelines. The message must:

- Be addressed to the participant by name
- Inform about local legislative districts
- Include the link to look up their district (findyourrep.legislature.ca.gov)
- Have NO partisan content or personal attacks
- Be within 500 characters

Template:

Subject: Local legislative districts in California

Hi <Participant>,

I'm also participating in the 2020 Election Study, and have some info to pass along about your local legislative district. Many Californians don't actually know which district they're in, or who their local representatives are.

You can look yours up at findyourrep.legislature.ca.gov.

These local reps decide the laws that affect us daily, so it's good to know who they are!

The predictions on the instrumental effect were incentivized using the binary quadratic scoring rule explained in intuitive terms (Danz et al., 2022). If the Sender ultimately did not message the potential Recipient, then the Sender could be awarded $1 for a correct prediction in the first scenario (without any message), and if the Sender did eventually message the potential Recipient, then the Sender could be awarded for the prediction in the second scenario (with the message from the Sender).[34]

For both these beliefs, Senders could theoretically edit their messages to increase the chances of their predictions being correct. For example, Senders who predict the Recipient will dislike their message could edit the language to be more harsh. But in practice, less than 20% of Senders actually edited the template by more than 10 words (out of 56) beyond the greeting and sign-off. Senders chose not to edit the message at all 56% of the time. Even if they planned to edit the message, the belief elicitations remain incentive-compatible.

Willingness-to-pay (WTP) elicitation. After the predictions, the Senders stated their willingness-to-pay to message each of the potential Recipients. For each potential Recipient, the Senders went through a binary choice search procedure and answered four process-of-elimination questions that located their switching point on a 15-row multiple price list (Figure B.5a-B.5b). The multiple price list was two-sided, allowing positive and negative WTP amounts between -$7 and $7 in $1 increments.[35] A negative WTP corresponds to a Sender who would pay out of their compensation to *not* send the email. A positive WTP represents a Sender who would pay *to send* the email.

The survey explained that their WTP choices would be entered into the following incentive-compatible scheme. After they submitted their WTP choices for all three potential Recipients, the computer would either (A) randomly select one of the $3 \times 15 = 45$ rows across the three potential Recipients and implement the Sender's preferred option in that row, or (B) randomly decide whether the Sender must send a message, and if so, to which of the three potential Recipients. In Case B, the Sender would receive the full compensation regardless, since their WTP choices were not exercised. The survey gave no suggestion whether they should or should not send the message, and stressed that it was in their best interest to answer the questions truthfully. Since this was the main dependent variable in this study, an attention check immediately preceded their choices, and the final sample includes only those who passed.

Email sending. After submitting their WTP choices, Senders were told the outcome of the incentive-compatible scheme. Those who were selected to message a Recipient were given the template with the option to edit the language. Senders of a Direct Message were required to email their message to the Recipient from their campus email account with the research team in the blind carbon-copy field. All other Senders were required to send the research team a confirmation email. This confirmation email meant that all Senders, even those who

[34]To determine whether the predictions were correct, each non-registered participant's voter registration status was tracked after the election (for potential Senders in the Pressure Message condition), and the non-registered participants were asked to list the names of their local representatives in the Follow-up Survey (for potential Senders in the Info Message condition).

[35]If participants stated a WTP at either bound (-$7 or $7), the survey asked for their hypothetically unbounded WTP.

were not selected to message a Recipient, had to send some email to hold constant the cost of opening the email client, copy-and-pasting the necessary fields, and clicking send. Before closing the survey, Senders confirmed that they had sent the required email, and that if they did not follow all the steps, their participation would be deemed incomplete and they would not receive any compensation.

3.2.5 Follow-up Survey: Recipients (Non-registered)

After the Senders took the Follow-up Survey, a different Follow-up Survey was sent to the non-registered participants. The survey asked the non-registered participants whether they received an email about the election from another participant. If so, they were asked what the email was about in a brief phrase, how much they liked or disliked receiving it on a 7-point Likert scale.

3.2.6 Sample characteristics, balance, and attrition

Out of all 1,059 participants, the large majority (84%) had already registered to vote before the study.[36] These registered participants take the role of Senders in the experiment ($N_{\text{Senders}} = 894$), and the non-registered participants become the Recipients ($N_{\text{Recip.}} = 165$).

The registered and non-registered participants differ demographically in ways that reflect the general population (Table B.3). Registered participants are more likely to be female (69% vs 60%, $p = 0.03$) and are older on average (21.6 vs. 21.0 years old, $p = 0.03$). They also hold different beliefs about the registration rate. Compared to their non-registered counterparts, the registered participants predict a 5 percentage point (pp.) higher voter registration rate on their campuses ($p < 0.01$).

The final sample of 894 Senders consist of those who completed both surveys, passed the attention check, and sent the required email. The experimental conditions are balanced on all collected observable characteristics (Tables B.2a-B.2c) and show no differences in attrition or in passing attention checks (Table B.2d).

Since the Senders and the Recipients are from the same university campus, some Senders could personally know their Recipients, but this is rare in practice given the random matching. In this experiment, 96% of the Senders state that they do not personally know any of their three assigned Recipients (Table B.2a). The Senders who do personally know a Recipient are studied separately in the heterogeneity analysis as a notable (albeit small) sample. In general, the results represent a group of young American students who belong to the same college community within a few degrees of separation, but are not personal acquaintances or friends.

3.3 Results

The results are presented in three parts. Section 3.3.1 first describes the Senders' preferences for pressuring their peers and relates their preferences with their beliefs. The second part in Section 3.3.2 uses the the experimental conditions to interpret the motives underlying the

[36] 87% of all eligible students in the University of California system registered to vote for the 2020 elections (Source: NSLVE).

Figure 3.3: Willingness-to-pay (WTP) to send Pressure Message

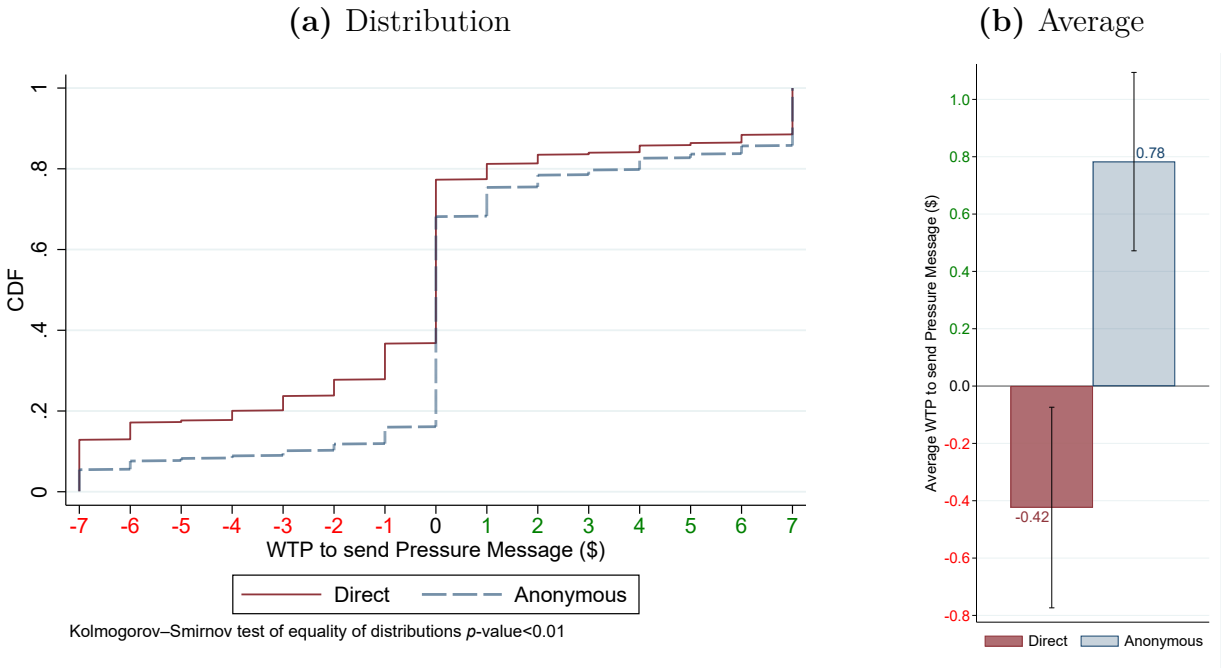


Figure 3.3b shows 95% confidence intervals shown with standard errors clustered by Sender.

Senders' preferences for pressuring. In Section 3.3.3, the final step analyzes the effect of peer pressure on the Recipients' participation in the election.

The analyses largely follow the specifications in the pre-analysis plan, though some pre-registered analyses could not be performed due to lack of sufficient variation in the sample (e.g., not enough Recipients who are close friends as opposed to acquaintances). Appendix Section B.2 maps the analyses in the paper to the specifications in the pre-analysis plan with further details.

3.3.1 Willingness to pressure and correlation with beliefs

From the Senders' willingness-to-pay (WTP) choices, we observe the full distribution in the Senders' willingness to pressure their peers and can also characterize the average Sender's preferences. We then correlate the Senders' preferences with their beliefs on the instrumental and affective effects of the message.

The WTP choices reveal that only a small minority of Senders would voluntarily send a pressuring email to an unregistered campus peer. The solid line in Figure 3.3a presents the CDF of the willingness to pressure. Most Senders either have a negative WTP (37%), meaning they would pay to avoid sending the email themselves, or are indifferent (41%). Less than a quarter have a strictly positive WTP.

On average, the WTP to send the pressuring email directly is -$0.42 (SE=0.18), as shown in Figure 3.3b. The typical Sender would prefer *not* to send the email and is willing to pay 42 cents instead. This statistically significant disutility is hard to explain with standard

effort or hassle costs alone, as Senders in this case are given a template to send.[37]

We begin to uncover the source of this cost of pressuring through the Senders' beliefs. Both have strong positive correlations with their willingness to pressure. From their beliefs on the instrumental effect, Senders in the lowest quartile who expect a persuasive effect of less than 4 pp. have an average WTP of -$1.91 to avoid sending the email. Meanwhile, Senders in the highest quartile who predict an effect greater than 20 pp. have a positive average WTP of $0.91 to send the email themselves.[38] This positive correlation is summarized in the bin scatter overlaying the square points in Figure 3.4a.

Likewise for the affective effect, the average WTP increases roughly linearly in Figure 3.4b across the Likert scale, where the squares represent the average within each bin. Senders who expect the Recipient to (strongly) dislike the message have an average WTP of -$2.34 to avoid sending the email, but Senders on the positive end are instead willing to pay +$1.89 to send the same email.

Given these strong correlations, the average levels of the beliefs reveal two counteracting motives at play. On the positive side, Senders believe that if they were to send the pressuring email, the Recipient would have a significantly higher chance of registering to vote by 14 pp. on average—far beyond the typical effect size in GOTV studies (Gerber and Green, 2017). Yet on the negative side, the majority (60%) of Senders also anticipate that the Recipient will dislike the message.

On net, whatever positive motive from the instrumental value of persuading the Recipient is outweighed by a negative motive from the detrimental affective effect on the Recipient, resulting in a negative average willingness to pressure. Senders could be demonstrating altruism toward the Recipient, or could be abstaining from pressuring for self-interested reasons. We next use experimental evidence to filter through the interpretations.

3.3.2 Interpreting the belief correlations through treatments

To refine our interpretation of the Senders' motives, we now introduce our two experimental treatments. We first examine how the Senders' preferences for pressuring, their beliefs, and the correlations change under anonymity, when the costs of revealing their identity to the Recipient are removed. We then investigate how the results differ among Senders assigned to the Info Message, which informs the Recipient about an online portal to look up state assembly representatives instead of pressuring them to register to vote. This comparison identifies the motives that are particularly relevant to peer voter mobilization.

Anonymity treatment.

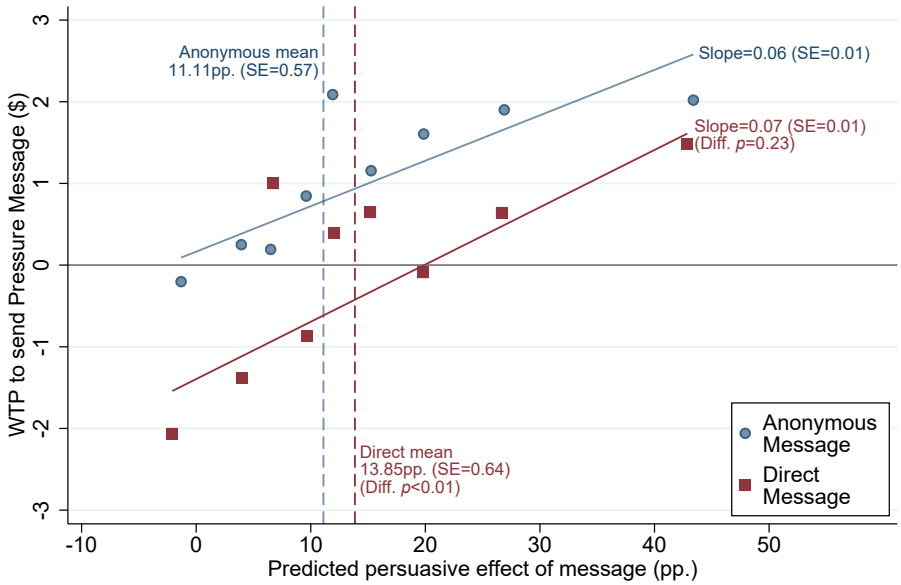
Effect of anonymity on the willingness to pressure. Under anonymity, there are significantly more Senders who are willing to send the pressuring email. The distribution of WTP under anonymity in Figure 3.3a demonstrates first-order stochastic dominance over the distribution without anonymity. On the "extensive margin", the rate of Senders with

[37] The majority of Senders who were assigned to send the email hardly edited the template as well.

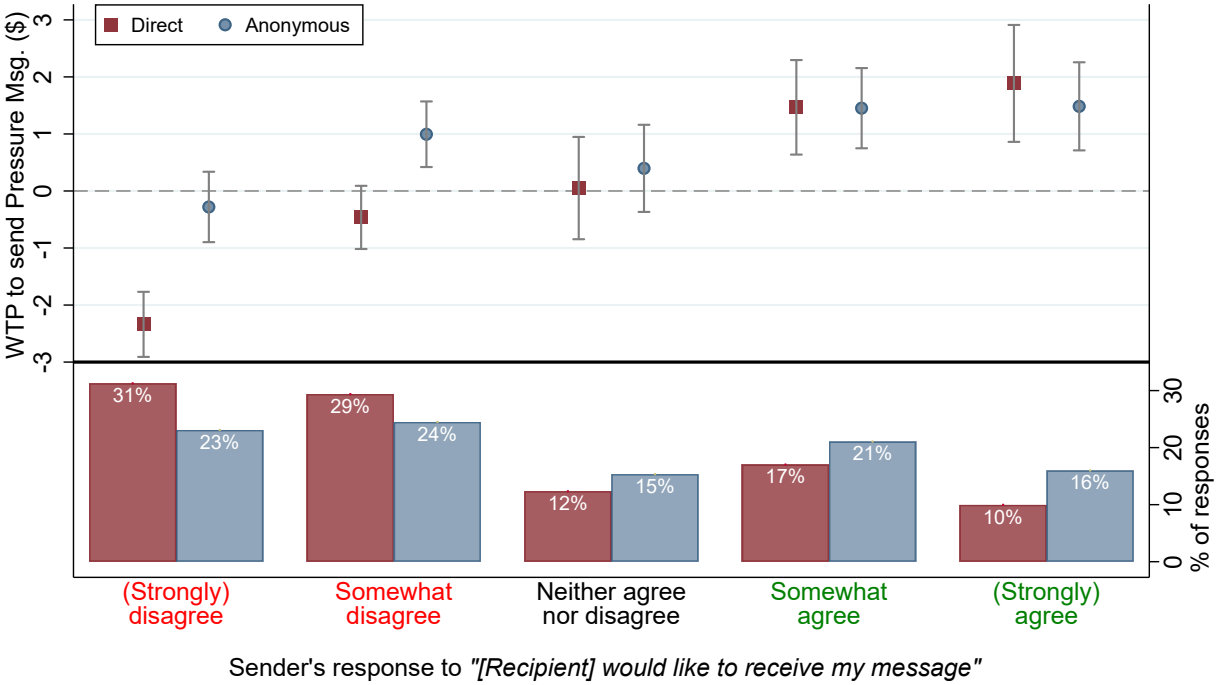
[38] The full distribution of beliefs on the instrumental effect is shown in Figure B.4.

Figure 3.4: Senders' WTP and beliefs for the Pressure Message

(a) Bin scatter of WTP on the instrumental effect



(b) Average WTP by beliefs on the affective effect



Bins in Figure 3.4a contain an equal number of observations. Figure 3.4b shows 95% confidence intervals with standard errors clustered by Sender.

a strictly positive WTP rises from 23% to 32%—an effect of +9 pp. (SE=2) or a relative increase of 41%.[39]

On the "intensive margin", the average WTP switches sign from -$0.42 when sending directly to +$0.78 when sending anonymously (Figure 3.3b). The effect of anonymity leads to an increase of +$1.21 (SE=0.16) in average WTP as reported in Column 1 of Table 5.3. The average Sender would pay to avoid emailing a peer directly, but then would pay almost twice as much to pressure the peer anonymously instead. We next aim to understand this drastic shift in the Senders' preferences through the changes in their beliefs.

Anonymity and the role of beliefs. Senders seem largely unchanged in their beliefs on the instrumental effect and their motives for persuading the Recipient, whether communicating directly or anonymously. Senders believe anonymous messages will be slightly less persuasive than direct messages (11 pp. vs 14 pp.),[40] but otherwise, the correlation between these beliefs and their WTP is fixed. In Figure 3.4a, there is little to explain the upward level shift in WTP for communicating anonymously.

The beliefs on the affective effect, however, divulge two reasons why Senders may prefer to message anonymously. The first, shown in the bottom panel of Figure 3.4b, is that Senders believe the Recipient is less likely to dislike a message delivered anonymously (47% vs. 60%, $p < 0.01$). Second, Senders become less sensitive to the negative affective effect on the Recipient when they can send anonymously, as evident from the flatter slope in the top panel of Figure 3.4b. Under anonymity, Senders not only expect the Recipient to react less negatively, but also seem to be less concerned about the negative affective effect itself.

To jointly quantify these relationships in a multivariate analysis controlling for the differences in levels and the cross-correlations in the beliefs, Column 2 in Table 5.3 estimates the equation:

$$\begin{aligned}\text{WTP}_{S,R} =& \alpha + \beta_1 \text{Instrumental belief}_{S,R} + \beta_2 \text{Affective belief}_{S,R} \\ &+ (\delta_1 \text{Instrumental belief}_{S,R} + \delta_2 \text{Affective belief}_{S,R} + \kappa) \times \mathbf{1}\{\text{Anonymous}\}_{S,R} + \varepsilon_{S,R}\end{aligned} \tag{4}$$

The dependent variable $WTP_{S,R}$ is Sender S's willingness-to-pay to send the message to Recipient R. The *Instrumental beliefs*$_{S,R}$ variable is the Sender's expected persuasive effect of the message on the Recipient's likelihood of taking the recommended action (i.e., registering to vote) in percentage points. The *Affective beliefs*$_{S,R}$ variable is the Sender's prediction for the Recipient's (dis)like of the message on a 5-point Likert scale, ranging from -2 for "(Strongly) disagree" to 2 for "(Strongly) agree" and centered at 0 for "Neither agree nor disagree". $\mathbf{1}\{Anonymous\}_{S,R}$ is an indicator for sending the message anonymously.

The estimates provide magnitudes on the relationships thus far. Even after controlling for the cross-correlation between the Senders' two beliefs, each still has its own sizable relationship with the willingness to pressure. Senders who predict a 10 pp. higher persuasive

[39]The richness of the data allows us to see the full joint distribution of WTP for sending directly and anonymously at the Sender-level in Figure B.6a. Most of the mass lies either on the 45-degree line (being indifferent) or above it (preferring anonymity). Figure B.6b summarizes the Sender-level joint distribution. The vast majority of Senders either prefer to communicate anonymously (36%) or are indifferent between sending directly and anonymously (57%).

[40]See Figure B.7a for the full distribution.

Table 3.1: Senders' WTP and beliefs across experimental conditions

	Pressure Msg.		Info Msg.			
Dep. Var.: WTP to send message ($)	(1)	(2)	(3)	(4)	(1)-(3)	(2)-(4)
Sender's beliefs on the effects of the message:						
Instrumental: Persuasive effect (10 pp.)		0.52		-0.03		0.55
		(0.11)		(0.07)		(0.13)
Affective: Recipient's (dis)like [-2, 2]		0.99		0.55		0.44
		(0.12)		(0.12)		(0.17)
Anonymous Message	1.21	0.90	0.93	0.93	0.27	-0.02
	(0.16)	(0.24)	(0.15)	(0.26)	(0.22)	(0.35)
Anonymous×Instrumental belief		-0.04		0.09		-0.13
		(0.12)		(0.09)		(0.15)
Anonymous×Affective belief		-0.68		-0.30		-0.38
		(0.14)		(0.14)		(0.20)
Constant	-0.42	-0.60	-0.74	-0.89	0.31	0.29
	(0.18)	(0.24)	(0.17)	(0.27)	(0.24)	(0.36)
Observations (Responses)	1581	1581	1100	1100	2681	2681
Senders	527	527	367	367	894	894
R^2	0.03	0.14	0.03	0.07	0.03	0.13

OLS estimates with standard errors clustered by Sender in parentheses. The dependent variable is the Sender's willingness-to-pay (WTP) to email the message. Two beliefs are elicited from the Sender. The first is the instrumental effect, or the predicted effect of the message on the Recipient's likelihood of registering to vote by the election (Pressure Message) or knowing the names of their local legislators (Info Message). The coefficient on this prediction corresponds to a 10 percentage point (pp.) increase in the predicted effectiveness. The second is the affective effect, or Sender's belief on a Likert scale for how much they agree with the statement that the Recipient would like to receive their message. The coefficient on this prediction corresponds to a 1-point increase on the Likert scale ranging from -2 (strongly disagree) to 2 (strongly agree).

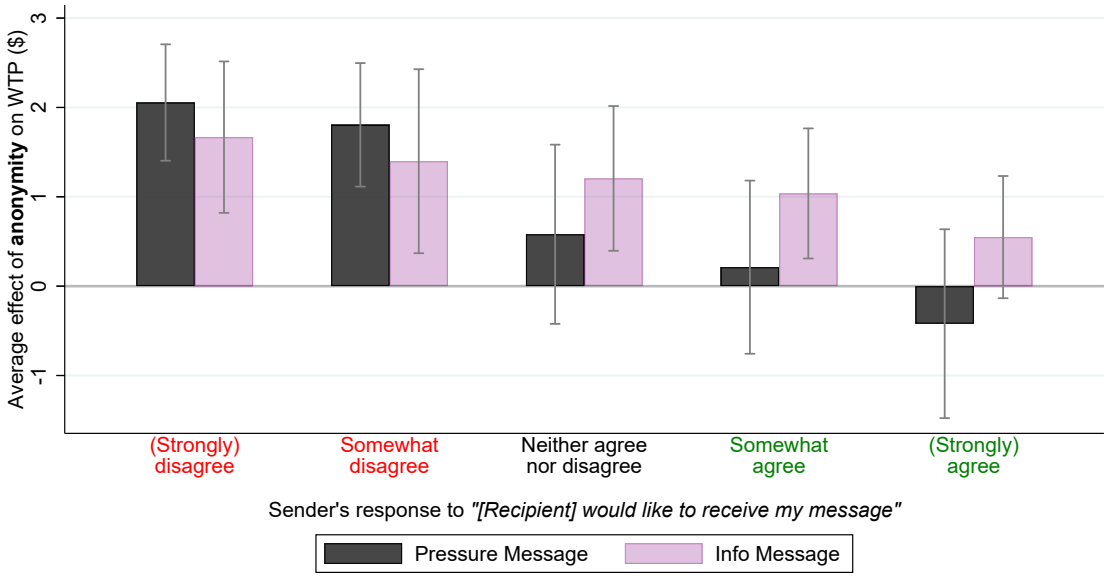
effect, or 0.7 in standard deviation units, have a $0.52 (SE=0.11) greater WTP on average, and those who expect the Recipient to like their message 1-point further up on the Likert scale (also 0.7 standard deviations) have a $0.99 (SE=0.12) higher WTP.

Under anonymity, the slope on the beliefs for the affective effect becomes significantly reduced by a remarkable two-thirds. The coefficient is still significantly positive ($0.99 - 0.68 = 0.31$, SE=0.11), which implies that this belief remains somewhat relevant for anonymous Senders. Yet Senders are three times more sensitive to how much they anticipate the Recipient will (dis)like the message when sending the email directly.

As the Senders' beliefs on the affective effect appear key to understanding the effect of anonymity, Figure 3.5 moves beyond the linear model in Equation 4 and toward a non-parametric approach by estimating the impact of anonymity within each bin of the Likert scale. Each bin contains Senders who hold the same belief on how much the Recipient will (dis)like their message. Within each bin, there are some Senders assigned to message anonymously. The figure displays the difference in the average WTP for those Senders communicating anonymously while also controlling for their beliefs on the instrumental persuasive effect.

The positive effect of anonymity on willingness to pressure is concentrated among Senders who believe the Recipient will *dis*like the message, which indicates that Senders avoid pres-

Figure 3.5: Effect of anonymity by beliefs on the affective effect



This figure presents the estimates for δ_k from the equation $\text{WTP}_{S,R} = \alpha + \beta \text{Instrumental belief}_{S,R} + \sum_k [(\beta_k + \delta_k \times \mathbf{1}\{\text{Anonymous}\}_{S,R}) \times \mathbf{1}\{\text{Affective belief}_{S,R} = k\}] + \varepsilon_{S,R}$ where k indexes each response on the 5-point Likert scale for the Senders' affective belief (i.e., (Strongly) disagree to (Strongly) agree). 95% confidence intervals are shown with standard errors clustered by Sender.

suring their peers mainly for self-interested concerns such as social costs from starting an unwelcome conversation. As shown in the darker bars of Figure 3.5, the effect of anonymity is around +$2 WTP for Senders who expect the Recipient to dislike their message. Meanwhile, the difference is statistically indistinguishable from zero among the Senders who believe the message will have a neutral or positive affective effect.[41] This pattern runs counter to an interpretation based on altruistic concerns for the Recipient's welfare, since the Senders should then be equally unwilling to send the message directly or anonymously when they expect the Recipient to dislike being pressured.

Heterogeneity. Since Senders see their Recipient's full name, they may also infer the Recipient's gender and ethnicity/race. Table 4.4 explores the heterogeneity in the estimates of the baseline specification (Equation 4) by the Recipients' gender, the Sender's gender, and whether the Sender shares the same racial/ethnic background as the Recipient.

When the Recipient is male, Senders are significantly more sensitive to whether the Recipient will (dis)like their message (Columns 1-2). This heightened sensitivity toward male Recipients disappears under anonymity. Senders could, for instance, expect a higher

[41] The Senders who predict the Recipient will (strongly) dislike their message still believe their email will have a robust persuasive effect of 8 pp. on average (SE=1). Furthermore, the correlation between the predicted instrumental effect and WTP *within* this most negative bin is just as high as in the rest of the bins. The results in Figure 3.5 are also comparable without controlling for the instrumental belief and even when restricting the sample to Senders who predict the same affective effect on the Likert scale for all three of their Recipients, whether messaging them directly or anonymously.

risk of retaliation from pressuring a male Recipient (as in Balafoutas et al., 2014). The results find no other distinction by the Sender's gender (Columns 3-4) or racial/ethnic homogeneity (Columns 5-6).

A small subset of Senders know their Recipients personally, as the Recipients are also students from the same university campus. Compared to Senders messaging a stranger (Column 8), Senders messaging their friends (Column 7) are significantly more sensitive to their beliefs on their friend's (dis)like of the message under anonymity. Senders seem more concerned about the affective effect when the Recipient is someone they know. This result, though intuitive, should be taken with caution given the sparse sample ($N = 22$) of Senders assigned friends as Recipients.

Info Message treatment. The second treatment, the Info Message, serves as a "placebo" comparison to identify the motives that are specific to peer voter mobilization, as opposed to peer communication for other purposes. Instead of pressuring the Recipient to register to vote, the Info Message informs the Recipient about an online search portal where they can look up their Californian State Senate and Assembly representatives. In general, the motives for peer communication can depend on many factors, such as the relationship (e.g., family vs. coworkers), the medium (e.g., face-to-face vs. online), and the presence of third parties. The Info Message condition holds all these other factors fixed while only changing the objective of the communication.

Senders of the Info Message are generally less willing to send the email. Figure 3.6a compares the WTP distributions of the Pressure and Info Messages (pooling the WTP from both the direct and anonymous conditions). Senders of the Info Message have a negative average WTP of -\$0.27 (SE=0.14) compared to the positive average of +\$0.18 (SE=0.15) for the Pressure Message (diff.=-0.44, SE=0.20). The distribution for the Info Message has less mass on the extremes, particularly on the positive side, and more mass on the neutral zero WTP. Overall, there is less variation in the preferences for sending the Info Message, with a slim majority (58%) being indifferent.

A key reason why Senders are more apathetic about the Info Message is that they seem to place no value on the instrumental effect of the message. Most of them do believe that their message will have a very high instrumental effect of persuading the Recipient to look up their local representatives (Figure B.7b). However, there is no relationship between the strength of this belief and their WTP to send the email (Figure 3.6b). This lack of correlation also indicates that the robust correlation observed in the Pressure Message condition is not due to experimenter demand effects, but instead reflects a real relationship between preferences and beliefs.

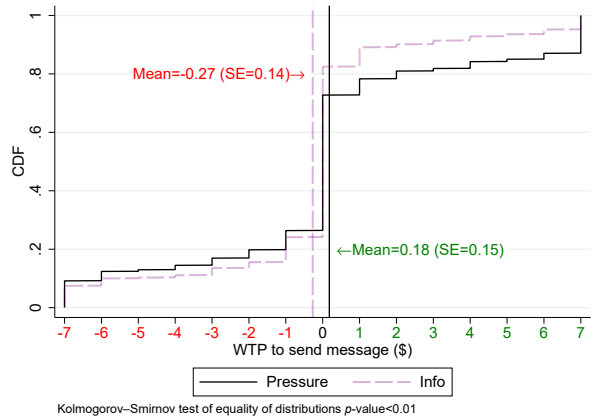
Table 3.2: Senders' WTP and beliefs for the Pressure Message: Heterogeneity

	Recipient's gender			Sender's gender			Recipient's race			Relationship		
Dep. Var.: WTP to send Pressure Msg. ($)	Fem./other (1)	Male (2)	(1)-(2)	Fem./other (3)	Male (4)	(3)-(4)	Same (5)	Different (6)	(5)-(6)	Friends (7)	Strangers (8)	(7)-(8)
Sender's beliefs on the effects of the message:												
Instrumental effect (10 pp.)	0.62	0.33	0.28	0.55	0.44	0.12	0.56	0.50	0.06	1.09	0.51	0.58
	(0.12)	(0.15)	(0.17)	(0.13)	(0.22)	(0.26)	(0.16)	(0.13)	(0.18)	(0.84)	(0.11)	(0.72)
Affective effect [-2, 2]	0.77	1.38	-0.61	0.90	1.24	-0.34	0.90	1.04	-0.15	1.02	0.98	0.03
	(0.16)	(0.16)	(0.21)	(0.14)	(0.25)	(0.29)	(0.17)	(0.15)	(0.19)	(0.79)	(0.13)	(0.69)
Anonymous Message	1.24	0.25	0.99	1.16	0.37	0.79	1.29	0.71	0.58	0.55	0.90	-0.35
	(0.29)	(0.41)	(0.50)	(0.29)	(0.45)	(0.53)	(0.39)	(0.31)	(0.51)	(2.02)	(0.25)	(1.75)
Anonymous×Instrumental belief	-0.09	0.07	-0.16	-0.05	-0.10	0.05	-0.18	0.04	-0.22	-1.57	-0.02	-1.55
	(0.14)	(0.21)	(0.25)	(0.14)	(0.27)	(0.30)	(0.21)	(0.15)	(0.26)	(1.19)	(0.12)	(1.02)
Anonymous×Affective belief	-0.43	-1.11	0.68	-0.70	-0.70	0.00	-0.61	-0.72	0.12	1.10	-0.68	1.78
	(0.18)	(0.21)	(0.27)	(0.17)	(0.27)	(0.32)	(0.22)	(0.18)	(0.27)	(1.11)	(0.15)	(0.96)
Constant	-0.94	0.06	-0.99	-0.74	-0.20	-0.54	-0.72	-0.54	-0.18	-0.96	-0.60	-0.36
	(0.29)	(0.34)	(0.41)	(0.29)	(0.48)	(0.56)	(0.35)	(0.29)	(0.40)	(1.23)	(0.25)	(1.08)
Observations (Responses)	972	609	1581	1086	495	1581	565	1016	1581	22	1559	1581
Senders	490	392	527	362	165	527	386	499	527	20	527	527
R^2	0.14	0.17	0.15	0.14	0.17	0.15	0.15	0.15	0.15	0.49	0.14	0.15

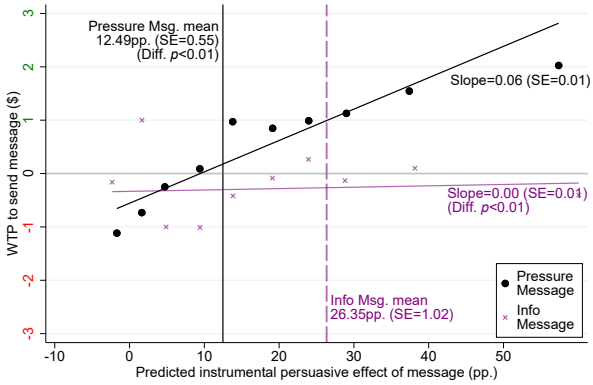
OLS estimates with standard errors clustered by Sender in parentheses. These regressions replicate the specification in Column 2 of Table 5.3 on the subsample indicated in the column headers.

Figure 3.6: Comparison between Pressure and Info Message conditions

(a) WTP distribution



(b) Bin scatter of WTP on instrumental belief



The contrasting importance of the instrumental effect is consistent with Senders holding stronger normative expectations for the Recipient to register to vote than to look up local representatives. In the absence of any messages, Senders predict that 34% of the Recipients will be able to identify their local legislators, whereas 42% will register to vote (diff. p-value<0.01). This difference in normative importance is also reflected by the prevalence of these behaviors in the literature. Participating in elections has been featured as a social norm in many studies (e.g., Gerber et al., 2008; Funk, 2010; DellaVigna et al., 2017), while looking up representatives in the state government has not.

The magnitude of the anonymity effect is comparable between the Pressure and Info Messages (Columns 1 and 3 of Table 5.3), but for the Info Message, the effect is less dependent on Senders' affective beliefs (Columns 2 and 4). Instead, the effect of anonymity for the Info Message is relatively more constant across their beliefs about the Recipient's affective response, illustrated in the lighter bars of Figure 3.5.

The comparison with the Info Message condition thus identifies two motives that distinguish the Pressure Message. First, Senders' beliefs about the instrumental effect predict their willingness to pressure, likely because the Senders value the objective of mobilizing their peers to register to vote. This correlation is not guaranteed in general. Second, the self-interested concerns for the Pressure Message are more contingent on the Recipient's dislike of the message, while the main costs for sending the Info Message seem to be more fixed in nature, such as the awkwardness of initiating an unsolicited conversation for a trivial purpose or the hassle of responding if the Recipient answers.

Robustness. Table B.4a performs robustness checks that control for observable Sender characteristics, add Sender fixed effects, allow quadratic terms on the Senders' beliefs, and account for the fact that Senders in the Info Message condition faced both registered and non-registered participants as potential Recipients. Columns 1 and 5 replicate the main results in Table 5.3. Columns 2 and 6 control for Sender characteristics (indicators for gender, racial/ethnic groups, undergraduate vs. post-graduate degree, and STEM major/field as well

as linear and quadratic terms in the standardized age) and find similar results. Columns 3 and 7 include Sender fixed effects, which also absorb any campus-specific effects. A notable change is that the coefficient on the instrumental belief for the Pressure Message becomes statistically insignificant. As there is limited within-Sender variability in the instrumental belief (i.e., Senders predicted a similar persuasive effect for all three of their potential Recipients), the fixed effects absorb most of the variation. Columns 4 and 8 estimate quadratic terms on the Senders' beliefs, which are all insignificant. Columns 9 and 10 account for the discrepancy that Senders of the Pressure Message only saw non-registered peers as potential Recipients, but Senders of the Info Message faced both registered and non-registered peers as potential Recipients. This distinction does not appear to matter. Lastly, Table B.4b uses the uncensored hypothetical WTP choices instead of -$7 or $7 for Senders who gave a WTP on one of these extremes. The results are similar.

3.3.3 Effect of peer pressure on voter participation

We now shift our focus from the Senders' motives to the effect of their peer pressure on the Recipients' participation in the election. 150 of the 165 non-registered participants in this study were randomly assigned either to not receive any message (control group) or to receive a Direct Pressure Message (treatment group) regardless of their WTP choices and their randomly assigned Sender's WTP choices (details in Appendix Section B.1). We analyze these 150 non-registered participants who exogenously received or did not receive a message to identify the causal effect of peer pressure in voter mobilization.[42]

In contrast to the Senders' beliefs, peer pressure has a marginally significant negative effect on voter registration in this experiment. Figure 3.7 shows the intend-to-treat (ITT) effects on the voter registration and turnout rates. The control group has a voter registration rate of 87% and a turnout rate of 75% in the election.[43] The rates in the treatment group are lower by 10 pp. for registration ($p < 0.1$) and by 6 pp. for turnout ($p = 0.5$).[44]

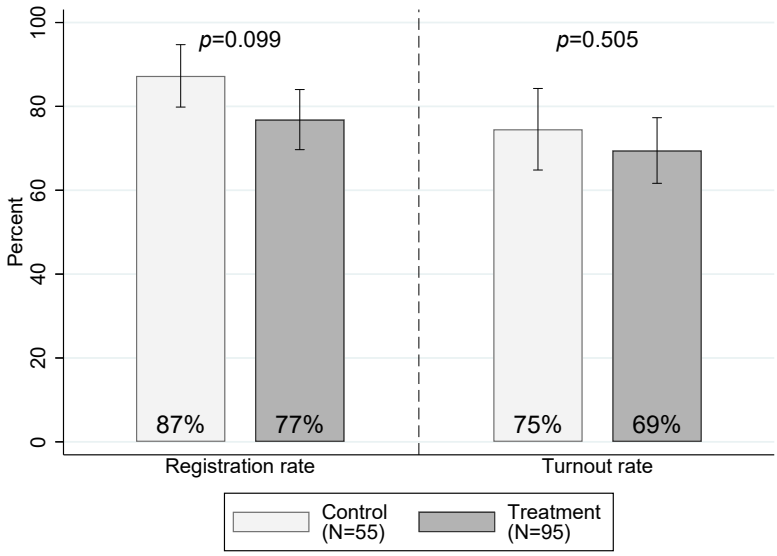
The treatment-on-treated (TOT) effects are also marginally negative for registration and null for turnout (Columns 2 and 5 of Table B.5), and the effects are uncorrelated with the Senders' beliefs about the effect of the message on the specific Recipient as well as their willingness to send the email to that Recipient (Columns 3 and 6). Treatment compliance is verified on the Follow-up Survey, which asks non-registered participants whether they had received an email from another participant, and if so, what the email was about. The

[42]The small sample of Recipients unfortunately precluded adding another treatment arm for receiving an *Anonymous* Pressure Message. The decision to only include the treatment of receiving a Direct Pressure Message was intended to maximize the power of detecting an effect for the common form of pressuring peers via social media.

[43]Among the Senders, the turnout rate is 94% (836 out of 894).

[44]For non-registered participants, there was no clear method to verify their self-reported status in the verification and matching process (Appendix Section B.1.2). Those who were missing from the voter database were clearly valid cases, but others could be registered under an old address and still technically be unregistered. There was no apparent reason for dishonesty, but some may have misunderstood the question or were unaware that they were already registered. For instance, California's Motor Voter program automatically registers those in the DMV system who become eligible. This should not bias the results, as treatment was randomized regardless. Nevertheless, Figure B.8 compares the treatment and control group excluding the "pre-registered" non-registered participants who were in the voter file before the study, yet voted without updating their address. The rates in the treatment group become even lower relative to the control group.

Figure 3.7: Intend-to-treat effects of receiving peer pressure on electoral participation



95% confidence intervals are shown with robust standard errors.

TOT analysis uses only the subset of 130 non-registered participants (87% of the sample) who completed the Follow-up Survey. Among those who completed, there are 48 all in the treatment group who answered that they had received a message about registering to vote (Table B.6).

A few months after the election, a Post-election Survey was offered to the 130 non-registered participants who completed the Follow-up Survey to determine whether the marginally significant negative effect on voter registration is more likely due to noise or a real impact. As some of them had experience receiving the email, their predictions may be more indicative of the true effect. 100 out of the 130 took the Post-election Survey. The survey explained that the participants who had not registered before the study either did not receive any email from another participant (labeled "Group A"), or received an email from another participant telling them to register to vote before the deadline (labeled "Group B"). The survey also displayed a template of the Pressure Message that the participants in Group B were assigned to receive from another participant. The survey disclosed that in Group A, 87% registered to vote and 75% voted in the election. The survey asked the non-registered participants to predict the corresponding rates for Group B. As in the main experiment, the survey incentivized the predictions by awarding a $1 bonus if either forecast was within 5 percentage points of the actual rates. The survey also asked them to briefly explain the reasoning behind their predictions.

Only 2 out of the 100 responses in the Post-election Survey predict a negative effect, which suggests that the more likely candidate for the marginally negative effect on voter registration is noise rather than a true effect. The median expected effect is +3 pp. for both registration and turnout. Compared to the overoptimistic Senders who forecast a median effect of +10 pp., the Recipients have a subdued perspective on the effectiveness of peer

pressure, perhaps because they are more familiar with the barriers to electoral participation. Eight of the respondents reflected on the experience of receiving the message themselves; 5 of them stated that the message had no bearing on them, and the remaining 3 said that it had a positive effect on their own decision to vote. Interestingly, although "pressure" was never mentioned in this Post-election Survey, 8 responses used the word to refer to the message.

3.4 Discussion

The young American college students in this study display two competing motives for and against pressuring their peers who have yet to register just before the 2020 Election. On the positive side, they believe that their act of pressure will hold significant sway over the chances of their fellow students registering to vote. They seem to consider this a worthwhile cause, in that those who believe they will be more persuasive are also more willing to send the email telling their peers to register, which is not the case for the placebo email encouraging their peers to look up their state representatives. In fact, a positive belief in their own persuasive power appears to be a necessary condition for any willingness to pressure at all. Given that this experiment is set in California where the electoral votes were essentially predetermined to swing left, these patterns do not match a "rational" voter who calculates the likelihood of the recipient voting for their preferred candidate and the marginal impact on election outcomes. Instead, these patterns align with preferences for civic duty (Feddersen, 2004), expressive voting (Brennan and Hamlin, 1998), or "altruistic" third-party enforcement of prosocial behavior (Fehr and Gachter, 2002).

On the negative side, they consider that receiving peer pressure is unpleasant. This regard for others is largely due to self-interested concerns rather than altruism. When they can message anonymously, they care much less. These self-interested concerns could stem from social image costs, for example, from the embarrassment of sending a somewhat obnoxious email, or from hassle costs, such as potentially having to respond if the recipient answers. Such costs seem relevant even when sending the placebo email.

When sending the pressuring email, however, the costs are especially concentrated among those who believe the recipient will dislike the message. This pattern hints at retaliatory concerns, which people state as a key drawback for enforcing prosocial behavior in other survey evidence (Balafoutas and Nikiforakis, 2012). In this context, students may be apprehensive about a derisive reply, gossip, or soured future interactions.

Overall, the negative motives prevail, leaving them significantly unwilling to pressure their peers on average. The costs of giving social pressure are consequential, even in this environment where they are given a template to minimize effort costs and have some degree of "social cover" (Burszytn et al., 2023) to pressure their peers as part of the study. The costs may be even larger in practice.

Despite optimistic beliefs about their own influence, their pressuring emails had no effect on the other students' participation in the election, which may be due to several reasons. The 2020 U.S. General Election was an extremely charged election. With expanded mail-in voting in the midst of the COVID-19 pandemic, voter turnout was the highest since 1900.[45] Those who did not participate even in this election may be facing barriers too high for a peer

[45]Source: United States Election Project.

"nudge" to overcome. In addition, before non-registered participants began the study, they did not know that their lack of compliance with the norm (i.e., their unregistered status) would be shared with others. Their actions were made public *before* they had time to adjust their status, whereas experiments such as the seminal Gerber et al. (2008) study revealed subjects' voting records to their neighbors *after* they had been given prior notice. This anticipation of public observation, as Fudenberg and Pathak (2010) find in the lab, is a crucial component in the effectiveness of peer monitoring. Email as a medium of communication also tends to less persuasive in general (Bennion and Nickerson, 2011; Bronchetti et al., 2015).

This experiment focuses on this sample of young Americans and the context of peer online voter mobilization for practical relevance, but this setting shares similarities with a host of daily interactions. The students in this sample interact with others from their campus within a few degrees of separation, like residents in the same housing community, workers in different teams of a company, or friends of friends on social media. The beliefs and the motives studied in this paper may also hold implications for giving social pressure in everyday situations as well.

Chapter 4

Policy diffusion and polarization across U.S. states

Economists have studied the impact of numerous state laws, from welfare rules to voting ID requirements. Yet for all this policy evaluation, what do we know about policy diffusion—how these policies spread from state to state? This chapter presents a series of facts based on a data set of over 700 U.S. state policies spanning the past 7 decades. First, for the introduction of new laws, state capacity seems to have a small role, in that larger and richer states are only slightly more likely to innovate policy. Second, the diffusion of policies from 1950 to 2000 is best predicted by proximity—a state is more likely to adopt a policy if nearby states have already done so—as well as similarity in demographics and voter policy preferences. Third, since 2000, political alignment is the strongest predictor of diffusion. Models of learning and correlated preferences can account for the earlier patterns, but the findings for the last two decades indicate a sharply increasing role of party control. We conclude that party polarization has emerged as a key factor recently for policy adoption, plausibly leading to a worse match between state policies and voter preferences.

This chapter is co-authored with Stefano DellaVigna.

4.1 Introduction

Economists have long studied the diffusion of innovations going back at least to the pioneering analysis of Griliches (1957) of agricultural innovations, followed by an extensive literature in the context especially of developing countries (e.g., Conley and Udry, 2010).

They have paid much less attention to the diffusion of policy innovations across government units, with the notable exceptions of the study of tax competition across U.S. states (Case, Rosen, and Hines Jr., 1993; Besley and Case, 1995; de Paula, Rasul, and Souza, 2020), the theoretical literature on states as laboratories of democracy, (Callander and Harstad, 2015) and learning across countries (Buera, Monge-Naranjo, and Primiceri, 2011). This limited attention is surprising given that numerous studies across nearly each subfield of economics have examined the impact of policy innovations. A few recent examples are the impact of Medicaid adoption on health (Goodman-Bacon, 2021), voter ID laws on turnout (Cantoni and Pons, 2021), and minimum-wage laws on worker earnings (Cengiz et al., 2019). Better understanding the diffusion of such policies is not just of interest on its own, but could also inform our understanding of studies such as these.

In this paper, we study the innovation and diffusion of policies at the U.S. state level. While one could also consider the diffusion across countries or at other decision-making levels, the analysis of U.S. states has several advantages. The U.S. federalist system allows states to serve as "laboratories of democracy". At the same time, the states are still comparable, given similar political institutions. We also have a rich political science literature to build upon.[46] Further, a crucial benefit is the abundance of state-by-state data on policy adoptions.

Our main data source is the State Policy Innovation and Diffusion (SPID) Database (Boehmke et al., 2020) which includes information on over 700 state law policies adopted in the last century. For each state law—for example on "Kinship Care Program" or on "Voter Registration by Mail"—the data set reports the year of adoption by state (if ever). This recent data set, which to our knowledge has not been previously used in economics, provides a fairly representative coverage by topic of state laws, but only a limited coverage of the last decade. We thus extended its coverage through the 2010s for a subset of the policies.

While this data provides broad coverage, it may not necessarily cover the state-level policies of interest to economists. We thus constructed a second sample from economics papers. From the 11,316 NBER working papers from April 2012 to September 2021, we identify 169 papers with U.S. state-level policy variation. Out of this set, 91 papers meet our criteria, for a total of 57 policies (given that some policies are in multiple papers).

The combined data set covers 705 policies adopted from the 1950s onward, 648 from the SPID data set and 57 from the NBER data set. The laws are most often about the provision of public services, law and crime, economics, and civil rights. Figure 4.1 presents three examples. Anti-bullying laws (Figure 4.1a) spread from the initial adoptions in Louisiana, West Virginia, and Colorado in 2001 in a fairly idiosyncratic way. In comparison, the Medicaid expansion from the Affordable Care Act (Figure 4.1b) followed political lines. Finally, the adoption of the initial prescription drug monitoring policy (Figure 4.1c) appears geographically clustered.

[46] Political scientists have studied the innovation and diffusion of policies across U.S. states as early as Walker (1969). See Graham, Shipan, and Volden (2012) for a review article and Mallinson (2020) for a meta-analysis.

We consider first a case study on Medicaid. As mentioned, the ACA Medicaid expansion spread largely to Democratic states (McCarty, 2019). A possible explanation is the higher need in Democratic states, but in fact the share of population that would benefit from the policy is larger in the Republican states. Since the costs of the policy are heavily subsidized by the federal government (Gruber and Sommers, 2020), this suggests that the state-level adoption was more a function of political considerations than of match to local needs. Has this always been the case? Interestingly, the initial Medicaid introduction from 1966 at the state level was essentially orthogonal to state-level voting, as was the introduction of the food stamp program in the 1960-70s. This case study thus suggests a recent increase in the role of partisan politics in the diffusion of state-level policies, but we cannot tell whether this is a general feature, or when this change occurred. We thus turn to the full data set.

We consider three main questions. First, are some states more likely to introduce new policies? Second, what predicts the diffusion of a policy across states? Third, are there patterns that allow us to tease out different models of policy adoption?

We point out some caveats. First, the findings mostly describe the patterns of policy diffusion and do not reflect causal inferences (Manski, 1993). Second, while the data set has broad coverage, it lacks details such as the text of the law or the likely medium of diffusion. Third, we do not observe the effectiveness of each policy, and thus cannot evaluate the role of effectiveness in the diffusion process. We nonetheless think that this descriptive evidence is valuable to cast light on different models and for predictive purposes, e.g., predicting which states are likely to adopt a particular policy in a difference-in-differences study.

Which states originate new laws? One theory is that states with more resources, capacity, or "legislative professionalism" innovate more (Walker, 1969; Besley and Persson, 2009). If innovative policies require a fixed cost, then larger and richer states should be more likely to generate new policies (Mulligan and Shleifer, 2005). Nevertheless, population and income per capita are not reliable predictors of originating more laws. Overall, while there are specific states that consistently produce new policies (e.g., California) and those that do not (e.g., Mississippi), innovation appears to be mostly orthogonal to observable state characteristics.

How do policies diffuse? The diffusion may depend on competition, e.g., states raising expenditures when neighboring states do (Case, Rosen, and Hines Jr., 1993; de Paula, Rasul, and Souza, 2020), learning (Wang and Yang, 2021), common preferences across states, and ideological alignment (Volden, Ting, and Carpenter, 2008). We measure this both "statically" and "dynamically". For the static measure, we take the states that have adopted the policy at a particular cross-section (say, after the first 10 adoptions), and assess their degree of similarity in the different dimensions (e.g., geographic or political similarity). In the dynamic method, we use a logit hazard model outlining the dimensions along which policies tend to diffuse, given the observed adoption up to that period. The dimension of diffusion is informative about the underlying models. For example, diffusion along politically similar states would suggest the importance of ideological alignment.

We document that the patterns of policy diffusion have changed substantially over the last seven decades. Policy adoption from the 1950s to the 1990s is best predicted by geographic proximity. Another important predictor is demographic similarity: a state is more likely to adopt a policy if other states with similar demographics (such as income or urban percentage) have already done so. The adoption by politically aligned states is a weaker predictor.

In the 2000s and 2010s, geographic and demographic proximity remain similarly pre-

dictive, but by far the strongest predictor becomes adoption by politically aligned states. Specifically, similarity in the Republican vote-share in recent elections becomes an important predictor in the last two decades, and even more predictive is the similarity in state party control. The latter factor implies a role of party influence.

Next, we relate these findings to leading models of policy diffusion. A set of explanations stresses *correlated preferences and environments*, *learning*, or *competition* among states. These (distinct) explanations all capture the importance of geographic and demographic proximity in the earlier decades, whether due to similar contexts, local spread of information, or competition at the borders. The recent patterns are a less obvious fit, but it is plausible that recently information flows, the extent of competition, and the correlation in preferences may have shifted from mostly geographic to largely political. To control for preferences, we measure the similarity in policy views across states among voters surveyed in the American National Election Studies (ANES) and in the General Social Survey (GSS), as well as using other measures of voter preferences in the literature. To capture information flows and to an extent competition, we use migration across states. These variables do predict policy diffusion, and they reduce the coefficient on geography and demographics by nearly half and the coefficient on vote-share by nearly a third. However, they do not affect at all the importance of state government control, which remains the most predictive variable.

As a further test of the growing importance of *party control*, we estimate an event study of switches from divided state governments to unified state governments (that is, the governor and the majority in both state houses belong to the same party). We detect no impact in the earlier decades, but in the last two decades, this transition indeed raises the probability of passing laws associated with the governing state party, with no impact on bipartisan laws.

A final explanation is that different types of laws, for instance in more politically controversial topics, have become more common. While we find similar results even after reweighting for changes in the composition of policy areas over time, we also address this concern by focusing on one specific category of laws: public health policies for preventing infectious diseases. While COVID-related state laws and rules are strongly driven by political factors, there is no such pattern for state vaccination laws passed since 1980.

Our findings indicate an important change in the match of state policy to voter preferences. The patterns for the earlier years are consistent with the findings of Erikson, Wright, and McIver (1989), that state policy used to be largely driven by voter preferences, not state party control. A contribution of our dyadic diffusion model is that we do not need to assign a partisan value to each law, as we use the *similarity* in voter preferences and in state party control to predict the diffusion of laws; this approach allows us to use a larger sample of laws. Our findings for the last two decades indicate instead a decreased role of voter preferences and a comparatively stronger role for party control. We thus add to the literature on polarization (Poole and Rosenthal, 1985; Fiorina and Abrams, 2008; Caughey, Warshaw, and Xu, 2017; McCarty, 2019; Canen, Kendall, and Trebbi, 2020; Boxell, Gentzkow, and Shapiro, forthcoming), documenting a sharp uptick at the state level since the 2000s that mimics, with a delay, the trend for politicians in Congress since the 1950s. These findings imply likely a worse match of policies to local voter preferences (e.g., Strumpf and Oberholzer-Gee, 2002).

The paper is related to the literature on policy experimentation (e.g., Callander and Harstad, 2015, Hjort et al., 2021, and Wang and Yang, 2021). While we do not observe the

policy effectiveness, the increased impact of party politics suggests that factors other than policy impact are playing a growing role in policy adoption.

The paper is related to the literature on policy diffusion. Relative to the small number of papers in economics, we examine a wide range of policies, complementing the detailed evidence on specific policies, e.g., taxation in the pioneering contribution of Besley and Case (1995), state-level fair employment laws (Collins, 2003), and welfare reform (Bernecker, Boyer, and Gathmann, 2021). In political science, in line with our findings, Caughey, Warshaw, and Xu (2017), Grumbach (2018), and Mallinson (2021) also detect evidence of widening polarization in the adoption of state laws. Relative to these papers, summarized in Table C.1, our unique contribution is that we estimate models that compare quantitatively the determinants of diffusion, allowing us to evaluate the role of different models. We also document that the recent polarization is even stronger for the policies studied by economists.

4.2 Case study: Medicaid and Food Stamp Program

Before we present the full analysis, we consider a case study. An important component of the Affordable Care Act was the expansion of the Medicaid health insurance to cover adults earning up to 138% of the Federal Poverty Line. The expansion comes at nearly no cost to the states, as the federal government pays 100% for newly eligible enrollees until 2016, and 90% thereafter (Gruber and Sommers, 2020). Despite this generous subsidy, the adoption at the state level has followed partisan lines, as Figure 4.1b shows. Indeed, Figure 4.2a shows that the Republican vote-share of the state predicts very accurately the year of adoption.

This suggests a large partisan impact on policy adoption, but it could be that the political preferences align with the underlying demand for the policy: the Republican states that delay adoption may have fewer people who would benefit from it. In fact, the opposite is the case: the states with higher Republican vote-share—the non-adopters—have a higher share of population that would benefit from the expansion (Figure 4.2b). The political preference thus appears to come at the expense of the match quality between the policy and the state.

A possible explanation is that major benefit expansions have always had this partisan structure. We thus revisit the initial Medicare roll-out enacted in July 1965. Voluntarily participating states received federal funds from January 1966, with an initial match of 50-83% across states, though the states had to cover certain groups and provide required benefits. This subsidy structure is thus not too dissimilar from the one for the ACA Medicaid expansion (though not as generous). Overall, 26 states enacted the Medicaid program within the first year, 37 within two, and nearly all within four years. Strikingly, the state political leaning does not predict the timing of adoption, as Figure 4.2c shows.

Another major public benefit expansion in the 1960s is the food stamp program. After county-level food stamp programs piloted in 1961, the Food Stamp Act was passed in 1964 and counties set up their own food stamp programs, with the federal government paying for the benefits and the states setting their own eligibility criteria. As the bin scatter in Figure 4.2d shows, the county voting patterns have no predictive power for the timing of approval. Demographics are predictive for the timing of adoption (i.e., counties with more vulnerable population) as Hoynes and Schanzenbach (2009) show, but not politics.

These case studies suggest that polarization may be playing a role in the current adoption of state laws in a way that was not the case in earlier years. Is this a general lesson? We

address this question and others in the next sections.

4.3 Data and summary statistics

SPID data set. The main source of data is the State Policy Innovation and Diffusion (SPID) Database (Boehmke et al., 2020). The data set includes information on over 700 state law policies adopted in the last century and combines existing data sets on state-level adoptions with the purpose of providing a representative sample of state policy topics. The main datasets aggregated in the SPID data set are (i) Boehmke and Skinner (2012) with 79 policies, itself building on the pioneering work of Walker (1969); (ii) Caughey and Warshaw (2016) with 104 policies mostly related to certification requirements for professions; (iii) the Uniform Law Commission (which focuses on nonpartisan legislation) with 187 policies, (iv) the National Center for Interstate Compacts with 52 policies, and (v) other smaller sources. We present 50 randomly sampled examples of these laws in Table C.2a.

For each state law—for example on "Kinship Care Program" or on "Voter Registration by Mail"—the data set reports the name of the law, the source, its policy area, and the year of adoption in each state (if ever). The data set does not record if a law is rescinded, since it is a fairly rare event. Furthermore, the data set records only binary adoption, and not continuous variables such as the level of the minimum wage across states.

As with any data set, one wonders about the representativeness and reliability. While there was certainly selection by topic in some of the meta-analyses used to build the data set, the final product is representative of the policy areas in typical state laws (Boehmke et al., 2020). We also validated the adoption dates for a sample of laws with rare corrections.

A significant limitation of the data set is the limited coverage of the most recent decade, as Figure 4.3 shows. We thus extended its coverage especially from 2015 to 2020 for a subset of the policies using publicly available data sources.

NBER data set. While the SPID data set is extensive, there is no guarantee that it covers the type of state laws of interests to economists. We thus collected a similar, though smaller, sample from economics papers. From the 11,316 NBER working papers from April 2012 to September 2021, we manually checked and identified 169 papers with U.S. state-level policy variation, covering especially labor, public, and health economics (Column 2 in Table 4.1a). We then apply our sample restrictions, including the restriction to binary policy adoption, yielding 91 papers (Column 3). For 81 out of these 91 papers we can extract the timing of state-level policy adoption, typically from a table in the paper, covering 57 policies (given that, for example, multiple papers analyze the same policy of Medicaid expansion). Health economics is the most common field, followed by labor and public economics, and the share of published papers, 46 percent, is similar to the overall share for NBER papers of 48 percent (Column 1), and similarly for the share published in "Tier A" journals (following the categorization in Heckman and Moktan, 2020). The full list of papers is in Table C.2b.[47]

Sample. We apply a set of restrictions to the pooled SPID and NBER data. First, we keep policies with the last adoption after 1950 since we do not have enough coverage

[47] 17 of the NBER policies are represented in the SPID data set to some extent. For example, one NBER policy is on the concealed carry of handguns, whereas a similar policy in SPID is on concealed carry in general. Some policies are present in both data sets, such as the state EITC, though the adoption dates tend to be more recently updated for the NBER source.

of historical patterns. Second, we consider only adoption in the contiguous 48 states, since coverage of Alaska, Hawaii, and Washington DC is spotty. Third, the data set does not include repeals and includes only binary measures of adoption (as opposed to, say, the minimum wage level).

As Table 4.1b shows, the data set includes 648 policies from the SPID data set, with an average of 24 states ultimately adopting each policy, and 57 policies from the NBER data set, with an average of 28 states ultimately adopting. As Table 4.1c documents, the most common topics, broadly grouped, are public services such as health and education, law and crime (especially in the SPID data set), economics (especially in the NBER data set), and civil rights (especially in the SPID data set). Over time, the topics covered have not changed much (Figure C.1a), and neither has the speed of adoption (Figure C.1b).

Outcome variables. For 20 of the 57 policies in the NBER sample, we reconstruct the dependent variable studied in the papers, either through the replication files or public data sources. The 10 state-level outcome variables (given that there are repetitions across the papers), such as the private insurance coverage rate and BMI, are summarized in Table C.3a. We supplement these variables with 18 other state-level variables typically used in policy evaluations from the Correlates of State Policy Project (CSPP), such as the state-level poverty rate or per capita welfare expenditure. We use these variables in Section 4.5.1.

COVID and vaccination samples. We collect 76 state policies enacted from October 2019 to August 2021 to deal with the COVID pandemic, such as the requirement to wear masks or school closures, from the COVID-19 U.S. State Policy database (CUSP) (Table C.3b). We record the policy adoption at the weekly level. We also collect information on the introduction of 28 state policies regarding vaccination mandates enacted since 1980 from sources such as the CDC and the Immunization Action Coalition (Table C.3c).

4.4 Evidence on innovation and diffusion

4.4.1 Innovation

We first consider whether some states are more likely to be early adopters. One theory is that states with more resources, capacity, or "legislative professionalism" tend to innovate policies (Walker, 1969; Besley and Persson, 2009). If there is a substantial fixed cost, larger and richer states should be more likely to generate new policies (Mulligan and Shleifer, 2005).

We define states that adopt a policy in its first year to be innovators, and sum the number of innovations by state. In Figure 4.4a-b we present a color-coded map of the U.S. displaying how often a state was an innovator in 1950-89 (Figure 4.4a) and in 1990-2020 (Figure 4.4b).[48] The map does not show an obvious pattern. California, the largest U.S. state by population, tops the list of innovators, but other large states such as Florida and Texas are in the middle of the pack and a smaller state such as Connecticut is among the top innovators.

Table 4.2 presents a statistical comparison between states in the top 20% of this innovation measure, versus states in the bottom 20%.[49] We find little evidence that states larger in population are more likely to be innovators, but some evidence only in the earlier period that states with higher per-capita income or higher "legislative professionalism" (Bowen and

[48] Figure C.2 presents similar plots splitting by the data source, SPID or NBER.

[49] In Table C.4 we present parallel evidence for the policies from the NBER papers.

Greene, 2014) are more likely to be in the top-innovators group. Furthermore, innovations are not predicted by the pattern of voting in the state, and are not more likely to come from unified Republican or Democratic governments, compared to divided state governments. Innovative states do have a larger share of population in urban areas. Overall, while some states consistently produce new policies (e.g., California) and others less so (e.g., Mississippi), innovation appears to be mostly idiosyncratic on observable state characteristics.

4.4.2 Policy diffusion

Following innovations, we examine the dimensions of similarity across states—geographic, demographic, and political—that predict the diffusion of policies. We consider first a static analysis of the first 10 states adopting a given policy, comparing their similarity along a particular dimension, relative to a benchmark of random diffusion. This static comparison provides non-parametric evidence but it does not use all the information on the path of diffusion, and it does not lend itself to multivariate comparisons of various determinants. We thus analyze the dynamics of adoption with a logistic hazard model.

Static evidence. For each law, we compute the proximity of the first 10 adopters (provided that this threshold of adoption was reached) with respect to the relevant dimension—e.g., geography and politics. As a measure of clustering along a dimension, we use the Geary's C statistic, which is typically used to measure geographic correlation (Geary, 1954; Barrios et al., 2012). The denominator is an unweighted average of the squared differences between all pairs, and the numerator is a weighted average where the weight for each pair increases in their proximity along the specified dimension:

$$C = \frac{\frac{1}{W}\sum_{i=1}^{n}\sum_{j\neq i} w_{ij}(x_i - x_j)^2}{\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}(x_i - x_j)^2}$$

where $x_i \in \{0, 1\}$ is an indicator for whether state i has adopted the policy, n is the number of states in the sample, w_{ij} is the weight for the pair ij, and W is the sum of weights.[50] If the states that are closer in the dimension are similar in policy adoptions, the weighted average of the differences in the numerator should be smaller than the unweighted average in the denominator. Consequently, values of this measure below 1 indicate clustering, values above 1 suggest the opposite, and a value of 1 is the null hypothesis.

To gain intuition, consider 5 states on a line, A, B, C, D, E, with each state contiguous to the nearby ones, that is, A is contiguous to B, B is contiguous to A and C, and compute Geary's C with respect to contiguity. Consider first the case in which the adoption of a policy is (1,1,1,0,0), that is, A, B, and C adopted, but D and E did not. The contiguous pairs are (1,1), (1,1), (1,0), and (0,0), each repeated. We average the squared difference between these pairs, yielding a numerator of 1/4. The denominator is the average of squared differences between all pairs, 12/20=3/5. This results in a C of $\frac{1/4}{3/5} = 5/12 < 1$, indicating substantial correlation among contiguous neighbors. Consider instead the case in which adoption is (1,0,1,0,1), with the same number of adoptions, but none contiguous. The numerator is 1

[50]The weight for pair ij may not equal the weight for the pair ji. For example, Michigan is in the closest third of states for Maine, but Maine is not in the closest third of states for Michigan.

given that all contiguous pairs are of the type (0,1), while the denominator is unchanged; the C is $1/(3/5) = 5/3 > 1$, indicating a negative degree of contiguous clustering.

In our case, in the numerator we assign equal weight to the third of other states most similar in the dimension of interest—geography or politics—and put zero weight on other states. We display 1-C, so higher values correspond to higher similarity, and 0 corresponds to no clustering. We compare the observed clustering after 10 adoptions to a counterfactual of adoption by 10 random states, from 1000 simulations.

In Figure 4.5a we display the geographic clustering of policies in the 1950s-70s (108 policies), 1980s-90s (222 policies), and 2000-10s (168 policies), indicating a degree of geographic clustering that is both substantial and persistent over time. For example, in the 1950-70s the Geary's C for the median policy corresponds to the 80th percentile of random policies.

In Figure 4.5b, we consider the extent of political clustering measured by the vote-share for the Republican presidential candidate, averaged over the two most recent elections. For the 1950s-70s and 1980s-90s, the median policy has a 1-C statistic close to 0, implying no measurable political clustering. In the 2000-10s, instead, we observe a clear rightward shift at all quantiles, including in the right tail. At the 90th percentile, the average $1 - C$ for the 2000-10s is 0.2, indicating substantial correlation, compared to 0.1 for the earlier decades.

Thus we detect both geographic and, increasingly, political clustering in policy diffusion. This finding is robust to measuring the clustering at the 16th adoption (a third of the contiguous states) and at the 24th adoption (a half) (Figure C.3).

A limitation of this analysis is that geography and politics are correlated, which this analysis does not separate. We thus turn to a hazard-type multivariate model.

Hazard model of diffusion. For all states i that have not yet adopted policy q in year t, we model the discrete-choice decision to adopt ($Y_{iqt} = 1$) with a logit specification:

$$\log\left(\frac{P(Y_{iqt}=1)}{1-P(Y_{iqt}=1)}\right) = \eta_q + \Pi X_{it} + \sum_k \beta_k p\left(A^k_{-iqt}, A_{-iqt}\right) + \varepsilon_{iqt}. \tag{5}$$

This specification, with the log odds on the left-hand side, has three right-hand-side variables. The first one, η_q, is a policy-specific baseline hazard rate for each decade, allowing for differences across policies in the overall probability of adoption. The second term, ΠX_{it} captures the overall impact of state-level features, such as state capacity, on adoption.

The third, key term, $\sum_k \beta_k p\left(A^k_{-iqt}, A_{-iqt}\right)$, captures the influence of adoption by other states that are similar along a particular factor k, such as geography, demographics, or politics. We adopt a functional form that measures how likely, or unlikely, the pattern of adoption by similar states (A^k_{-iqt}) is, relative to the adoption by all states (A_{-iqt}), with respect to a particular dimension k. Considering the case of geography ($k = g$), we first compute the probability of $a^g_{-iqt} \in \{0, ..., 15\}$ adopters within the closest third of states, given the total number of adopters $A_{-iqt} \in \{1, ..., 47\}$, under the null of uniform adoption:

$$P(a^g_{-iqt}|A_{-iqt}) = \begin{pmatrix} A_{-iqt} \\ a^g_{-iqt} \end{pmatrix} \frac{\left(\frac{15!}{(15-a^g_{-iqt})!}\right)\left(\frac{32!}{(32-(A_{-iqt}-a^g_{-iqt}))!}\right)}{\left(\frac{47!}{(47-A_{-iqt})!}\right)}$$

The measure is then the probability of having fewer adopters in the closest set of states

minus the probability of having more adopters in the closest set of states:

$$p\left(a^g_{-iqt}, A_{-iqt},\right) \equiv P(A^g_{-iqt} < a^g_{-iqt}|A_{-iqt}) - P(A^g_{-iqt} > a^g_{-iqt}|A_{-iqt}) \tag{6}$$

Consider a state i that has yet to adopt a policy that has been adopted by $A_{-iqt} = 16$ states, of which $a^g_{-iqt} = 5$ are in the closest third geographically. Under the null, the probability of seeing fewer adoptions in the closest third of 15 states is 0.38, and the probability of more adoptions in the closest third is 0.37. Hence, $p\left(a^g_{-iqt}, A_{-iqt},\right)$=0.38 − 0.37 = 0.01: the adoption by nearby states is in line with the adoption nationwide. Suppose instead that 10 of the 16 adoptions had been in the closest third of states. In this case, the probability of seeing fewer adoptions in the closest third is 0.998, and the probability of seeing more is 0.0002, and $p\left(a^g_{-iqt}, A_{-iqt}\right) = 0.998 - 0.0002 = 0.998$, indicating high diffusion in the neighboring states.

This measure ranges from -1 (states similar to state i statistically have been unlikely to adopt a policy) to +1 (states similar to state i have proven quite likely to adopt). This functional form captures the strength of clustering along a particular dimension, with a cap; that is, if hypothetically 14 out of the 16 adoptions had been in the contiguous states, instead of 10 out of 16, the measure $p\left(a^g_{-iqt}, A_{-iqt}\right)$ would have been essentially the same, as the evidence was already statistically very strong. Later, we consider alternative measures, such as the proportion of the states in the closest third that have adopted.

We build analogous measures of demographic and political similarity, except that the set of similar states is time-varying. To capture demographic similarity, we take the average state-level log population, share of urban residents, and log income per capita over the last two years, standardize each variable within each year, calculate the absolute difference in each dimension, average to create the index, and then identify the closest third of states.

We create two measures of political similarity, one for voter preferences and one for party control. For voter preferences, we take the third of states with the smallest absolute difference in the average Republican vote-share from the two most recent Presidential elections. For similarity in state party, we categorize three types of state governments—-unified Democratic (i.e., the governor is Democratic and both state houses have a Democratic majority), unified Republican, and divided state control (all other cases)—and define the "closest" states to be those with the same partisan control. We consider separately the case of unified control (Republican or Democrat) and the case of divided split-party governments.

Table C.5a shows for each decade pairs of states that are especially close along that dimension, and Figure C.4 displays how often a pair of states that are close along a dimension in year t are still close in that dimension in year t+4. The stability is of course 1 for geography, above 0.9 for demographics, between 0.6 and 0.9 for vote-share, and between 0.5 and 0.8 for party control of state government.

The four similarity parameters—β_g for geography, β_d for demographics, β_v for vote-share and β_p for party control—are scaled to be comparable allowing for a quantitative comparison across determinants, which is unique in the literature (Table C.1). Hence if β_g is larger than β_d, adoption by geographically similar states is more predictive on average for future adoption by state i than adoption by demographically similar states.

We estimate specification (5) separately by decade, though we pool the 1950s and 1960s given the more limited coverage for the earliest years. In each year t, only states that have

not yet adopted policy q are in the sample. For each policy, we start the model in the first year of adoption and end it in the last year of adoption in the sample, and exclude policies that end with fewer than 5 adopters or span less than 3 years. We cluster the standard errors at the state level to capture autocorrelation, as well as correlations across policies.

We stress that we do not place a causal interpretation on the estimates in (5) (Manski, 1993). For example, the adoption of a policy by a state may be predicted by the adoption of geographic neighbors because of learning and diffusion of information (Banerjee, 1992; Bikhchandani, Hirshleifer, and Welch, 1992), or alternatively because of common demand for a policy or a common shock (e.g., a shared lobbyist). With this in mind, it is still useful to examine which dimensions predict adoption, as they inform us about the most likely nature of common shocks and circulation of ideas. Furthermore, even viewing the results as purely descriptive, they enable one to make predictions about future adoptions, which can be useful, for example, in the econometric evaluation of a difference-in-differences design. In Section 4.5.3, we provide estimates with a causal interpretation from an event study design for a specific variable, the change in state government control.

Hazard estimates. As Table 4.3 shows, we do not find any reliable pattern that state-level demographics X_{it}, including state income or population, predict faster adoption.

We thus turn to the similarity predictors β_k. Demographic similarity is predictive of adoption: in the 1980s we estimate a coefficient of 0.21 (s.e.=0.06), which remains about constant up to the most recent decade, at 0.24 (s.e.=0.06). These estimates are certainly consistent with the impact of similar context and preferences, but can also be interpreted in light of models of competition and learning, if demographic similarity reflects these margins.

Next, we consider the impact of geographic closeness, which we expect to capture the impact of competition across neighboring states, learning about policies, and similarity in contexts and preferences. Geographic similarity is highly predictive, with a larger impact than demographic similarity, and with consistent importance over time, with a coefficient of 0.38 (s.e.=0.07) in the 1970s and of 0.36 (s.e.=0.07) in the most recent decade.

Third, we consider the role of similarity in the state-level Republican vote share. For the first five decades, political similarity is a modest predictor, with an effect size mostly between a third to a half of the one for geographic similarity: 0.13 (s.e.=0.06) in the 1970s, 0.06 (s.e.=0.05) in the 1980s, and 0.18 (s.e.=0.05) in the 1990s. In the last two decades, however, the impact *triples*, at 0.38 (s.e.=0.05) in the 2000s and 0.38 (s.e.=0.06) in the 2010s.

The impact of similarity in political voting could capture similarity in voter political preferences, or the impact of state parties. To capture the latter component, we include party control of the state government. In all the decades up to the 1990s, similarity in state party control matters relatively little and is not statistically significant. Yet in the 2000-10s period, previous adoption by other governments with the same state party control becomes the strongest predictor of adoption (estimate of 0.52, s.e.=0.07) for states under a unified state government. For states with split governments, there is no predictive power of adoption by other states with split governments, which further underscores the role of party control.

Figure 4.6 summarizes the evidence from the hazard regressions. Geographic and demographic similarity between states has consistently predicted the likelihood of passing the same laws. Meanwhile, similarities in the state-level voting and party control explained little in the past, but in the last two decades, have become the most important predictors. We interpret this change as evidence of a shift in the model of state policy-making, with party

discipline taking on a newfound key role in the 21st century.

We note that the pseudo R-squared has generally increased over time from 0.13 in the 1970s to 0.20 in the 2010s. Thus, the process of adoption has become more predictable.

Simulated diffusion. In Figure 4.7a we present counterfactuals for the 1990s (Figure 4.7a) versus for the 2010s (Figure 4.7b). We take a hypothetical policy introduced by California in 2000 and simulate its diffusion over 20 years or until 10 adopters. For every state that has yet to adopt, we calculate its probability of adopting, and based on that probability, we randomly draw whether it adopts in that year. We assume the same political and demographic variables from the relevant years (2000 onward) across the two plots, and only vary the estimated diffusion coefficients. We color-code the states as function of the probability that a state is among the first ten adopters across 1,000 simulations.

The policy with the estimated 1990s coefficients (Figure 4.7a) diffuses geographically in the West, as well as in some demographically similar states such as Florida and politically aligned states in the Northeast. Meanwhile, with the estimated 2010s coefficients (Figure 4.7b), the spread of the policy becomes highly concentrated in the states with similar political leaning in the Northeast and along the West Coast in Oregon and Washington, while geographically close but politically distanced states such as Nevada, Utah, and Arizona become less likely to adopt.

In Figures A.5a-f, we document a similar increase in the role of political leaning following an innovation in: (i) Connecticut, a state that is reliably Democratic like California but is smaller and on the other coast (Figure C.5a-b); (ii) Texas, a large, Republican state (Figure C.5c-d); and (iii) Ohio, a Republican-leaning Midwestern state (Figure C.5e-f).

Robustness. In Table C.6 we explore the robustness of the results in Table 4.3. We run the models for the decades 1950-70s, 1980-90s and 2000-10s and report the coefficients on geographic, political, and state party similarity.

We estimate (i) a linear probability model instead of a logit; (ii) a reweighted version that holds the composition of policy areas fixed over decades, (iii) a model with an expanded set of controls;[51] (iv) a parsimonious specification which drops the state characteristics X_{it} (e.g., the level of urban %), which are typically not significant. The results are similar across these specifications. We adopt the parsimonious specification in the panels to follow.

Next, we adopt alternative measures of adoptions among similar states: (i) using thresholds of the closest fifth, fourth, third, or half (Figure C.6) instead of the closest third in Equation 6; (ii) adoption by other states up to year $t-1$, instead of up to year t; (iii) a weighted average of the adoption status of all other 47 states, with weights proportional to the other state's rank in similarity; e.g., the most distal state carries 1/47th of the weight of the most similar state. These results are very similar to the benchmark.

In rows 8-10, we present simpler parametrizations compared to Equation 6, such as the proportion of adoption among states in the closest third. These measures, which suffer from

[51]The additional set of controls include the non-white percentage, the unemployment rate, indicators for unified Democratic and Republican state governments; quadratic terms for the proportion of other states adopted, Republican vote-share, log population, income per capita, urban percentage, non-white percentage, and the unemployment rate; adoption measures among the closest third of states in state government partisanship, migration flows, non-white percentage, and the unemployment rate; a flexible policy-specific baseline hazard parametrized as a step function that varies every five years; and state fixed-effects. Table C.7a also shows the estimates for each demographic variable separately.

mis-specification issues (Appendix Section C.1), all point to the increasing role of politics.

Heterogeneity. In Table 4.4 we estimate the parsimonious specification in row 5 of Table C.6 for different subsamples. In the NBER sample of policies studied by economists, the recent rise in importance of vote-share is even larger, with a coefficient of 0.50 (s.e.=0.08), compared to 0.35 (s.e.=0.04) in the SPID sample. For Interstate Compacts on which states cooperate to address a common problem, such as the Interstate Wildlife Violator Compact, we see less evidence of party polarization, as expected.

In Panel B we vary the policy area. For economic policies, we find a fairly constant role of politics and a decrease in the role of geography over time. The latter result seems to run counter to strong competition across neighboring states. For non-economic policies, the role of geography is about constant, and the impact of political polarization is especially strong, as one would expect given the polarizing nature of social issues.

In Panel C we split the states into thirds based on their vote-share as Republican-voting, Democratic-voting, and "battleground" states. The increased importance of politics is mostly driven by the Republican-voting states and especially the Democratic-voting states, fitting with a party-driven model for the polarization. In the battleground states, the party in control varies from Democratic to Republican, so battleground states do not specifically adopt policies from one another, as opposed to states at the polar opposites that do.

Finally, returning to the "state capacity" model, we may expect a smaller impact of geographical closeness for larger states if state capacity enables them to learn from a broader range of other states. We find only suggestive evidence of such heterogeneity.

Comparison to results in the literature. The diffusion of policy along geographical and demographic lines is consistent with the results on tax legislation and competition across U.S. states in Besley and Case (1995) and de Paula, Rasul, and Souza (2020), for example, and with findings in the political science literature as early as Walker (1969) and in Mallinson (2020) who reviews the papers since then. More recently, Caughey, Warshaw, and Xu (2017), Grumbach (2018), and Mallinson (2021) find evidence, as we do, for the increasing importance of political alignment for policy diffusion. Relative to these papers, we compare quantitatively the impact of polarization to the impact of geographic and demographic similarity, we present results for the most recent years, and we document even stronger patterns for the policies studied by economists.

4.5 Evidence relating to models of policy diffusion

We now relate findings in the previous section to leading models of policy diffusion.

4.5.1 Correlated environments, learning, and competition

A set of explanations stresses the *role of correlated preferences and environments*, *learning* across states, or *competition* among states. While these explanations are distinct, they share the prediction about the importance of demographic and geographic proximity for policy diffusion, whether due to similar contexts, local spread of information, or competition at the borders. The evidence for the 1950s to the 1990s thus fits neatly with these models.

These explanations are a less obvious fit for the patterns from the 2000-10s, though it could be that the diffusion of information, the extent of competition, and the correlation

in preferences or environments across states have recently followed less geographic lines and more political lines. We present three pieces of evidence to assess these explanations.

Voter policy preferences. The first test for *correlated preferences* uses survey measures of voters' policy preferences from both the ANES and the GSS beginning in the 1960s. Specifically, we find the average response to policy preference questions (e.g., whether abortion should be legal) in each state, standardize the ordinal responses across questions, and then calculate the average absolute difference across questions to measure the similarity in voter preferences between each pair of states. Since 15 states, such as Delaware, Vermont, and Wyoming, have irregular representation in these data sets (Figure C.7a), we use the remaining 33 contiguous states (with the closest third now including only 10 of the other 32 states) throughout the analysis. Further, we use an index of existing measures of voter policy preferences in the literature to provide an alternative measure. We provide more detail on the policy preference questions and the sample coverage for ANES and GSS, as well as on the measures from the literature, in Appendix Section C.2.

Migration flows. The second test uses cross-state migration. If unobserved interstate flow variables such as information and competition are responsible for the diffusion of policies and have recently followed more political lines, the observed interstate flow of migration likely would exhibit similar patterns and predict policy diffusion. We thus identify the top third of other states with the highest volume of inflow-outflow migration.

Estimates. In Table 4.5 we first replicate the result of Table 4.3 pooling decades in Column 1-3 and including only the 33 states consistently represented in ANES and GSS. Then in Columns 4-6 we add controls for similarity in voter preferences as well as in migration flows. The measure of migration flows has modest explanatory power, while the two measures of similarity in voter preferences are stronger predictors, though somewhat less so recently. The measure based on the ANES and GSS has coefficients of 0.26 (s.e.=0.09) and 0.28 (s.e.=0.09) in the earlier decades, but only 0.15 (s.e.=0.08) for the most recent 2000-10s decades. The coefficients on the index of public opinion measures in the literature are fairly constant over time ranging from 0.16 (s.e.=0.04) to 0.20 (s.e.=0.04).

What is the impact of controlling for voter preferences and migration flows on the predictive power of the other variables? The addition of these variables reduces by nearly half the explanatory power of geography and demographics, especially in the most time periods. The predictive power of Republican vote-share in the most recent decades also falls, from 0.38 (s.e.=0.05) to 0.27 (s.e.=0.05). Strikingly, these variables leave the coefficient on the similarity in state government party control essentially unaffected, from 0.44 (s.e.=0.08) to 0.42 (s.e.=0.08). The lack of movement in the coefficient even after including measures of voter preferences suggests that its rise in recent decades likely reflects top-down partisanship rather than bottom-up demand from the voters.[52]

Evidence from outcome variables. As a final piece of evidence, we consider variables that are typical policy outcomes, such as the state-level opioid mortality rate, income, and poverty rate. If changes in local preferences or environments are driving the increased impact

[52]In Table C.7a we examine separately the impact of each policy opinion measure used in the index. In Table C.7b we analyze the impact of the GSS-ANES similarity variable considering separately questions that either match or do not match the policy area of the law. For example, we match voter responses to ANES questions on the economy to policies in the Economics policy area. Appendix Section C.2 discusses these results.

of politics in policy adoption, we would expect these outcomes to have become more correlated among politically similar states. If, instead, other factors are at play, the correlation between the outcomes and politics may not have changed over time.

We compute the Geary's C statistic using the closest third of states by vote-share for these variables, first for the period 1980-85 and then for the period 2005-10. Figure C.8a provides no evidence that these variables have become more politically correlated.[53]

These findings suggest that the increased weight of political variables on policy adoption is not due to patterns of interstate correlation in voter policy preferences, information flows, or competition, but mostly to other factors.

4.5.2 Evidence within area: Vaccination policies

A possible confound for the findings thus far is that the composition of policies in the sample has changed over time, for example, to include more politically controversial laws. Reassuringly, Figure C.1a shows that the composition of policy areas has remained fairly stable, except for the last 5 years, and re-weighting to hold the composition of policy areas fixed over the decades does not affect the estimates (Row 3 in Table C.6). Nonetheless, it would be useful to consider a narrower class of policies. We thus focus on public health policies for preventing infectious diseases, comparing COVID-related state policies adopted since October 2019, such as masking policies and school closures, with earlier vaccination policies adopted since 1980, such as immunizations requirements for schools and hospitals.

For the COVID policies, given the shorter time frame, we estimate the model (5) at the weekly level in Columns 1 and 2 of Table 4.6. We estimate a significant impact of demographic and geographic similarity, but especially of state party control.[54] For comparison, in Columns 3 and 4 we estimate (at the yearly level) the adoption of vaccination policies beginning in earlier decades. In this sample, demographic and geographic similarity are the strongest predictors, with no impact of political similarity in vote-share or state party control.

Even in this narrower topic of infectious disease prevention laws, we reproduce our key result: while geographic and demographic similarity have historically been the most important predictors of policy diffusion, recently party politics has become the foremost driver.

4.5.3 Event study on party discipline

The hazard estimates so far provide descriptive evidence on predictors of adoption. We now use an event study to provide causal evidence on the impact of party political control. We focus on the switch to unified party control at the state level, a critical threshold according to the political science literature. We estimate the model

$$Y_{iqt} = \sum_{d=-4}^{4} \delta_d 1\left\{t - e_i = d\right\} \left(1\{q \text{ aligned}\} + 1\{q \text{ opposing}\} + 1\{q \text{ neutral}\}\right) + \Pi X_{it} + \boldsymbol{\alpha}_i + \gamma_{qt} + \varepsilon_{iqt}$$

where Y_{iqt} is an indicator for whether state i adopts policy q in year t, e_i is the year of switch to unified party control (with the state elections typically occurring late in the prior year,

[53] Figure C.8b documents that the outcomes have become less geographically correlated in recent times.

[54] Cui et al. (2021) also provides consistent evidence of partisan spread of COVID policies.

$t = e_i - 1$), and the key parameter δ_d is allowed to depend on the ideology of the policy q. We categorize the ideology of policies using the vote-share of the states that have adopted the law so far.[55] We control for each state's baseline probability of adopting left-leaning, right-leaning, and neutral policies with $\boldsymbol{\alpha}_i$, for state government election years with X_{it}, and for the different levels of adoption with policy-year fixed effects γ_{qt}. We include all state-year-policy observations for states that have yet to adopt around the event window if at least one state has a switch during that window to identify the baseline parameters, such as the policy-year fixed effects γ_{qt}.

Figure 4.8a displays the event study coefficients for the period 1990-2020. A switch to a unified state government does not lead to any increase in the passage of neutral-leaning state laws; it does not appear that unified government reduces gridlock. Next, we consider the impact on the probability of adopting a policy that aligns ideologically with the inaugurated unified state government, compared to the adoption of policies leaning in the opposite direction. We detect a statistically significant increase of about 2 percentage points in the 4 years following the switch, compared to the year before the switch. The increase arises already in year t, as one would expect, and appears to be persistent. In contrast, in the earlier 1950-1989 time period (Figure 4.8b) we do not uncover any partisan impact of a switch in party control.[56] We find similar results using the event study estimator from Chaisemartin and D'Haultfoeuille (2020) (Figure C.10c-d). Thus, this event study confirms that partisan support of laws is a recent phenomenon at the level of U.S. states (Caughey, Warshaw, and Xu, 2017).

4.6 Discussion and conclusion

We documented a series of facts about the diffusion of state-level policies in the U.S., and related them to different models of policy diffusion. The estimated impact of similarity in geography, demographics, and voter preferences resonates with models of competition across states, learning from state to state, and underlying similarity of voter preferences. It is

[55] We take the average 2-party Republican vote-share (demeaned by year) in the latest Presidential election as of the year of adoption, among the states that have adopted the policy by year $t-1$. If a policy has been adopted on average by states with a 1 percentage point or higher advantage in the Republican vote-share, we define the policy as Republican-leaning, and conversely for Democratic-leaning policies. If the average vote-share of states adopting a policy is within the 2 percentage-point buffer, we code the policy as neutral-leaning. Policies can be classified as neutral in one year but ideologically aligned with one party in another year when new adoptions occur, but we drop a small fraction of policies that switch from left- to right-leaning or vice versa at some point in their life-cycle. Figure C.9a shows the distribution of the average demeaned Republican vote-share among adopters over the last 30 years. Figure C.9b follows the ideological evolution of the three most left-, right-, and neutral-leaning policies in 1990 until 2020. Figure C.9c displays the classification of policies for thresholds other than 1 pp.

[56] In Figure C.10a-b, we also show the event study estimates with the most plausible confound path (Freyaldenhoven et al., forthcoming). In Table C.8 we estimate the separate components of the event study: the switch to a Republican unified government on the passage of Republican-leaning policies (as per the coding above, Column 2) and of Democratic-leaning policies (Column 3), with the difference in Column 4; the impact on neutral policies (Column 5); and the same specifications, but for switches to unified Democratic state government (Columns 6-9). The findings generally follow the expected patterns, with the largest impacts from switches to Democratic state governments for Democratic-leaning policies. In Column 10 we consider switches away from unified state governments, which yield smaller impacts.

difficult to tell these models apart, given that they share several key predictions.

The pattern for the most recent two decades—a significant increase in the importance of political similarity, and especially of state party control—points to the increasing role of another factor: party influence. Thus, policy adoption at the state level increasingly appears to have a top-down influence, beyond a simple match to bottom-up voter preferences.

This result runs parallel with other studies on polarization. Politicians in the U.S. Congress have shown polarizing voting patterns since the 1950s, as reproduced in Figure 4.9 using DW-NOMINATE data. Our results indicate that the polarization of state-level policies did not start until later, in the 2000s. Still, its role is rapidly rising and it has affected even topics such as vaccinations which in previous years were not polarizing.[57]

One of the most touted advantages of the U.S. federalist system is the ability of independent states to tailor their policies swiftly and optimally to voter preferences and state-specific needs. Yet the current trends suggest that the adoption of state policies is becoming if anything less responsive to local economic demands, and instead bending more to partisan forces. While measuring the welfare implications of such top-down policy choices is beyond the scope of the paper, we note the implications about the quality of the match between policies and state voter preferences, as well as welfare externalities on other states (e.g., Knight, 2013).

Our findings raise a number of questions for future work. For one, it would be meaningful to disentangle the sources behind the increasing role of political factors, whether it be lobbyists, party rules, or organizations that provide "copy-and-paste" legislation, such as the American Legislative Exchange Council (Angelucci, Ash, and Longuet Marx, 2022). It would also be useful to know whether this trend of polarization has reached even lower levels of governments, such as city policy-making, or other decisions in the public interest, such as the content of textbook or medical rules.

Our findings also suggest that researchers can assess the extent to which any particular law diffuses more geographically or politically. As a first approximation, in Figure 4.10 we plot a scatter plot of our measure of clustering, 1 - Geary's C, computed for every policy along both the geographic and the political dimension. The shaded regions show the 5th to 95th percentile of the $1 - C$ statistic under the null of random diffusion. Generally, the actual policies fall into three categories. One group has a pattern of diffusion that is largely predicted by politics, such as the Medicaid expansion. A second group has diffusion that is predicted by both geography and politics, such as the ban on employers asking about a prospective employee's past salary. Finally, a third group, which includes Anti-Bullying Laws, appears to be fairly idiosyncratic, at least based on these parsimonious measures.

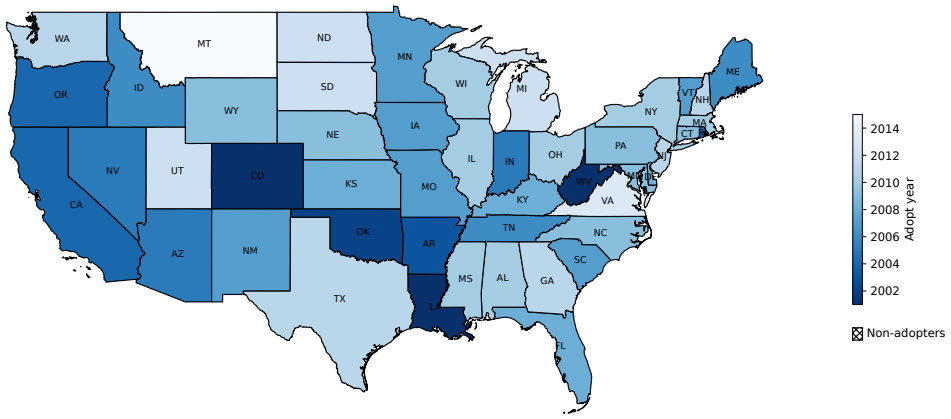
We envision that this categorization can guide researchers studying a policy change to identify the degree of correlation in the diffusion process of their policy, relative to the average paper of this kind. For example, the presence of geographic versus political diffusion suggests different concerns for identification, a topic which we leave for future work.

[57] This evidence is consistent with the roll-call state data patterns in Shor and McCarty (2011).

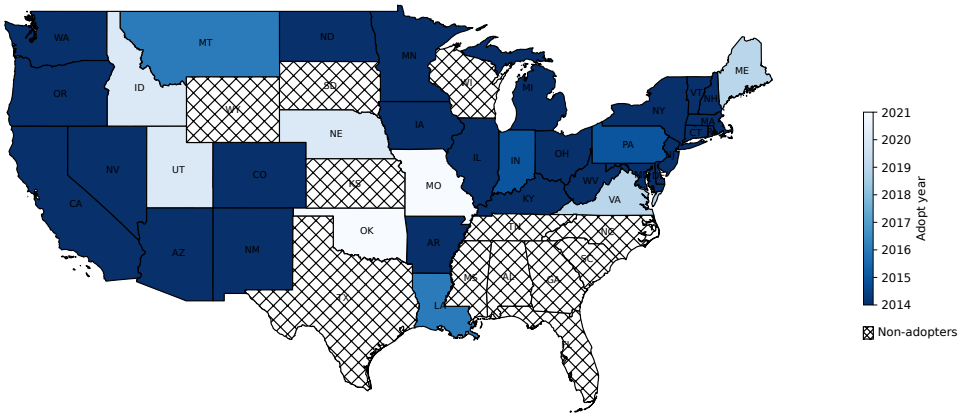
4.7 Figures and tables

Figure 4.1: Three policy examples

(a) Anti-bullying laws



(b) 2014 Medicaid expansion (Affordable Care Act)



(c) Initial prescription drug monitoring program

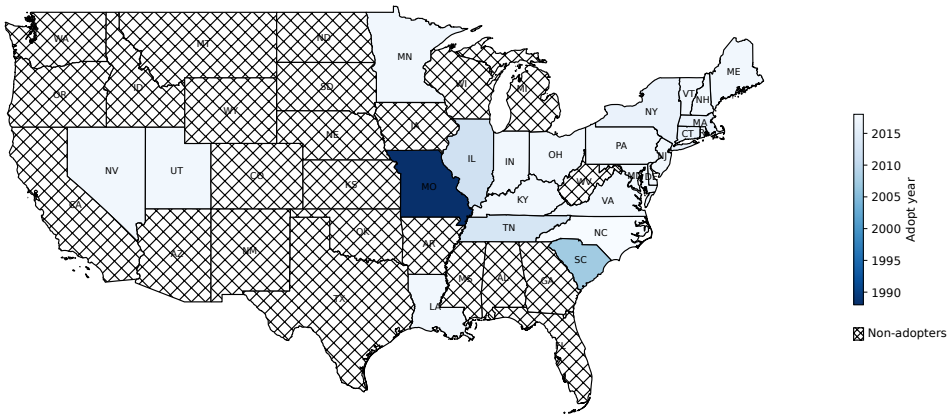
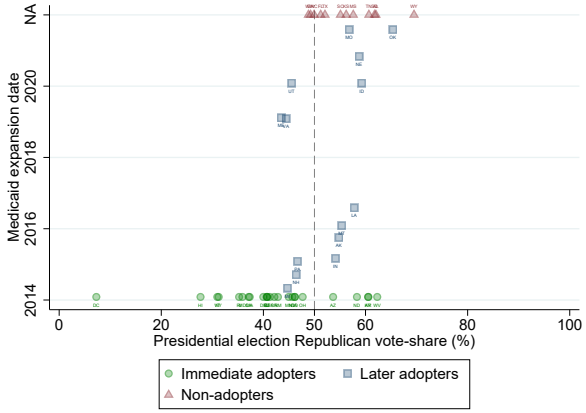
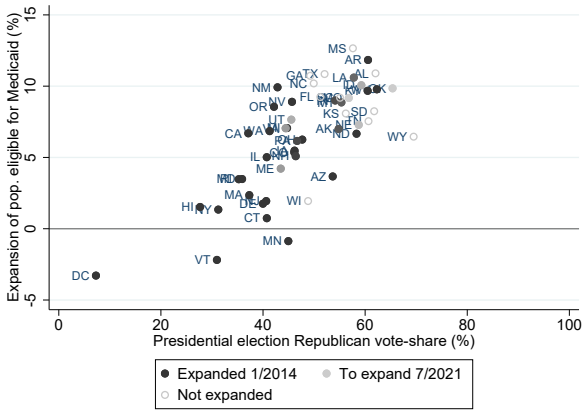


Figure 4.2: Case studies of welfare programs

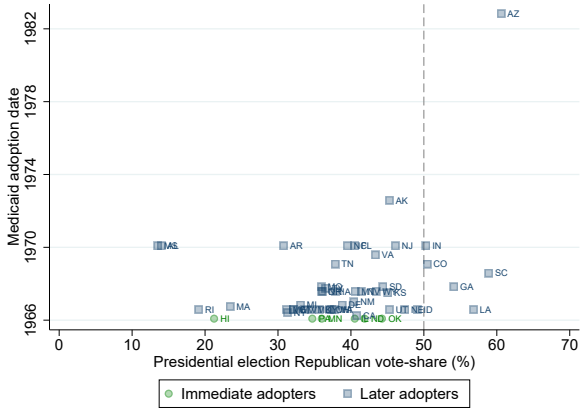
(a) 2014 Medicaid expansion (ACA)



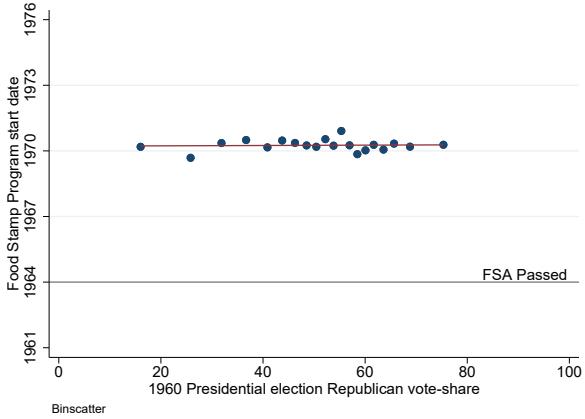
(b) 2014 Medicaid expansion coverage



(c) 1960s Initial Medicaid adoption



(d) 1960-70s Food Stamp Program adoption (county-level)



For Figures 4.2a-4.2c, the Presidential vote-share is from the most recent election to the year of adoption, and for non-adopters in Figures 4.2a-4.2b, the vote-share is from the 2020 election.

Figure 4.3: Policy sources

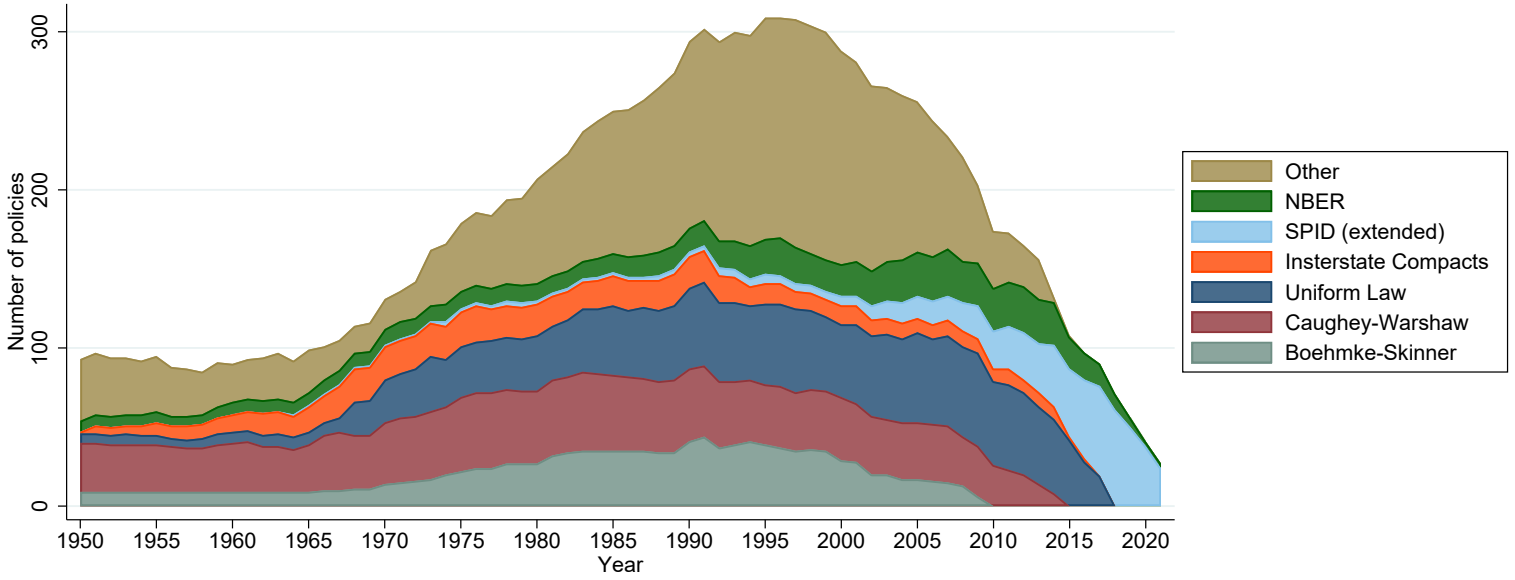
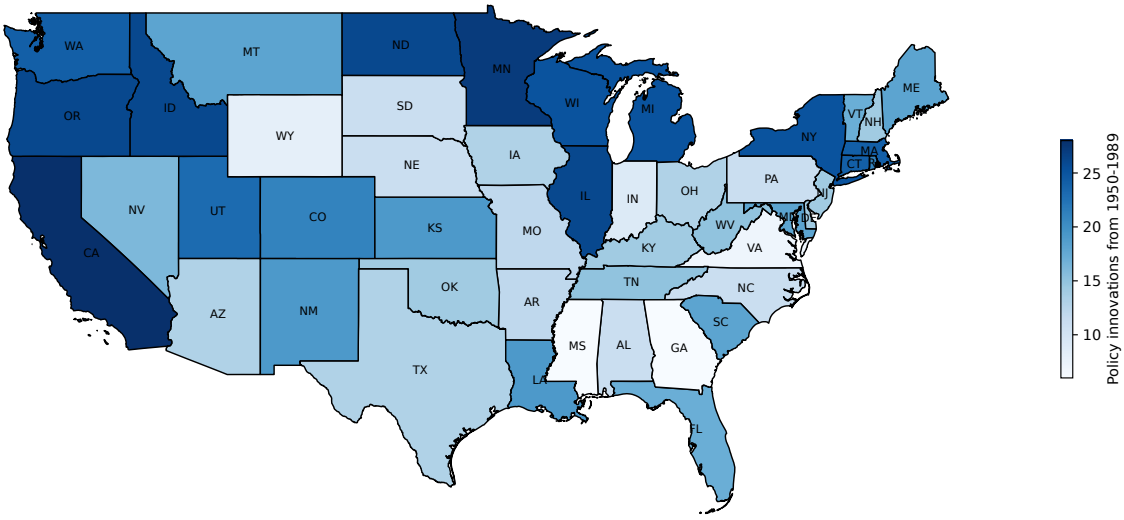


Figure 4.3 shows the number of active policies with ongoing adoptions for each year by the source of the policy. All sources are from the SPID dataset, except for the NBER policies. The "SPID (extended)" subgroup refers to policies from SPID that this paper extended for further coverage of adoption in recent decades.

Figure 4.4: Innovating states

(a) Policies innovated 1950-1989



(b) Policies innovated 1990-2020

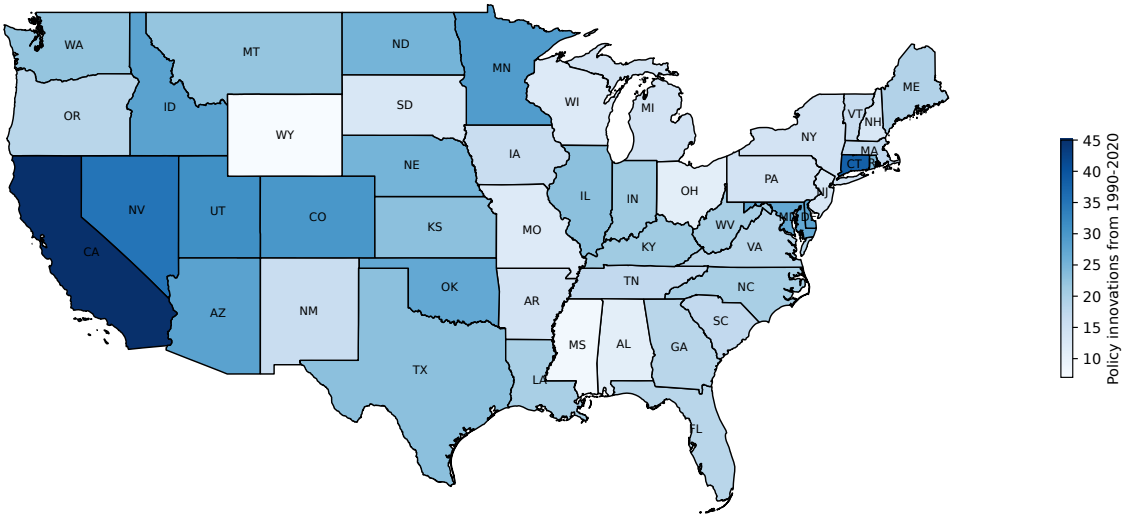
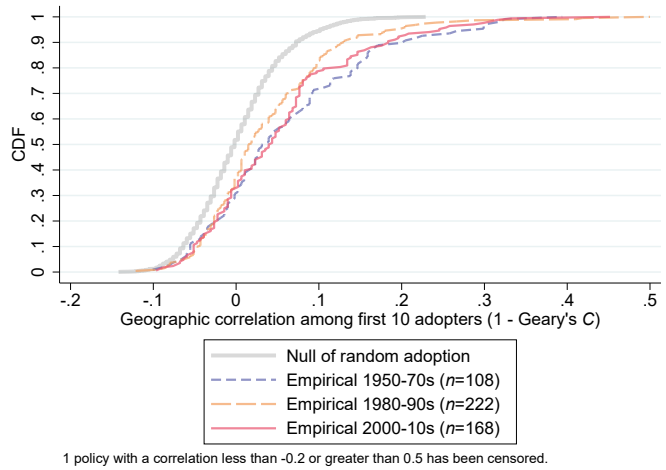
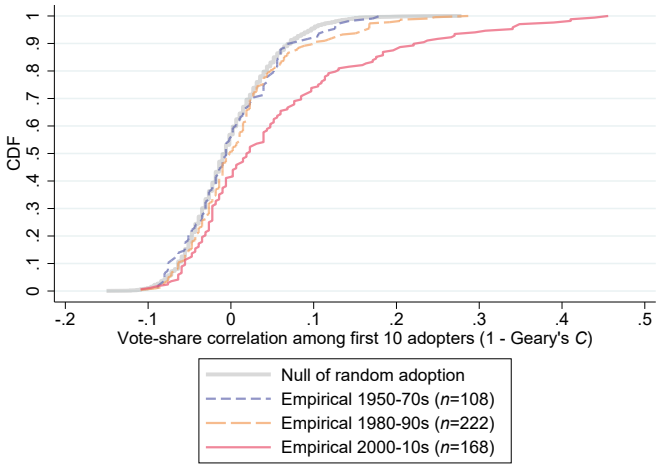


Figure 4.5: Correlation in geography and politics among adopters (random and observed)

(a) Correlation in geographic distance (first 10 adopters)

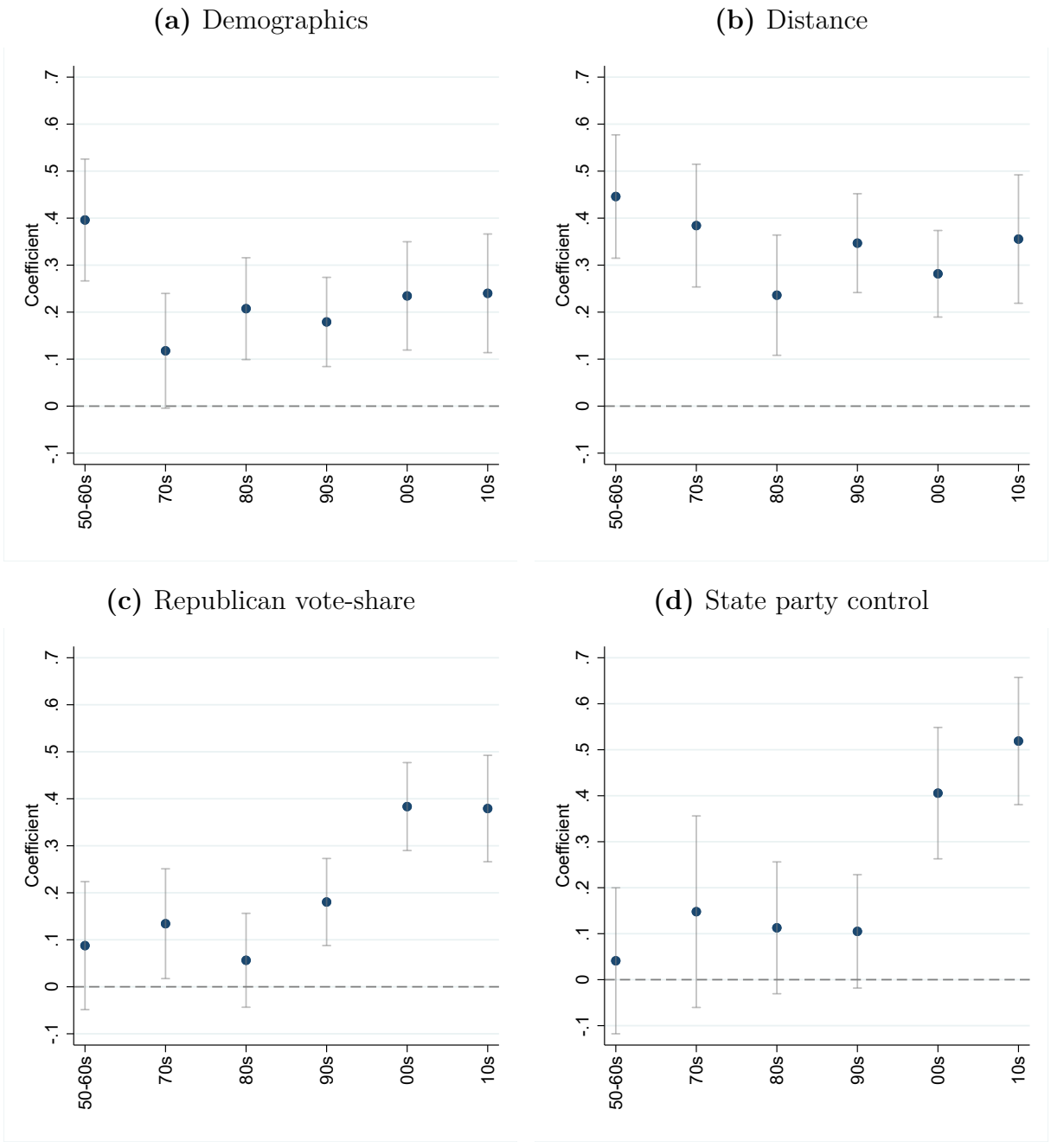


(b) Correlation in Republican vote-share (first 10 adopters)



This figure plots the CDF of the 1−Geary's C statistic for policy adoptions, which measures the correlation of adoptions within a specified dimension. Geary's C is calculated by taking the weighted average of the pairwise squared differences in adoptions, where the weights are increasing in the similarity between the pair of states along the specified dimension. The weighted average is then divided by the unweighted average of the pairwise squared differences across all pairs of states. This figure uses a simple weighting scheme, in which for each state, the other states in the closest third by geographic distance (Figure 4.5a) or by Republican vote-share (Figure 4.5b) are given equal weight, and the remaining states outside the closest third are assigned zero weight. The measure is calculated in year that the policy reaches 10 adopters with ties are broken randomly. Under the null of uniformly random adoptions, the expected value of 1 - Geary's C is 0.

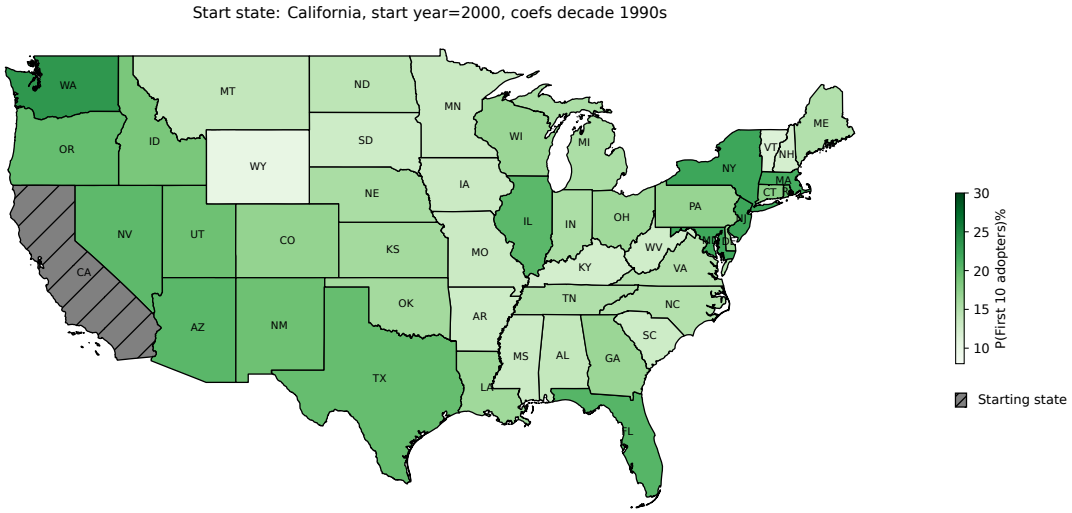
Figure 4.6: Dynamics of policy diffusion dimensions



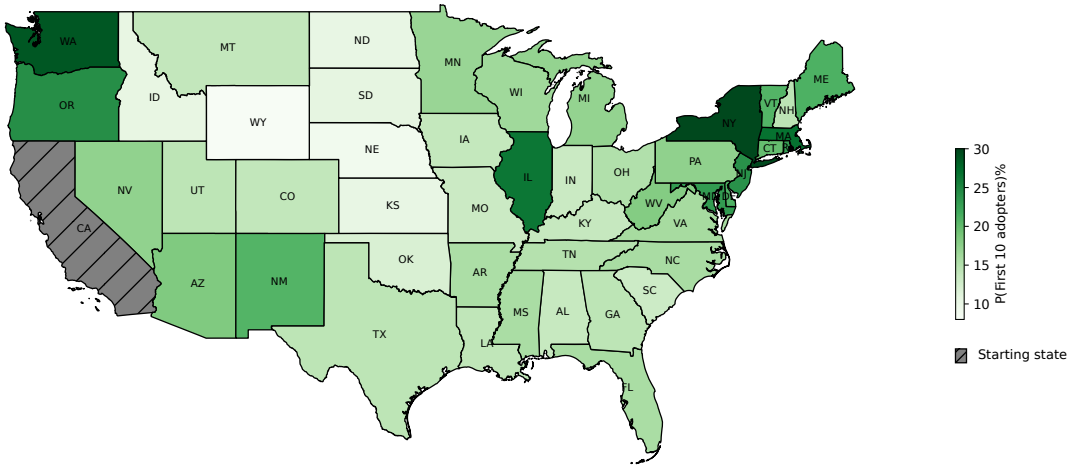
This figure plots the decade-by-decade estimates from the model in Table 4.3 for the coefficients on the measure of adoption among the closest states in each dimension. 95% confidence intervals are shown with standard errors clustered by state.

Figure 4.7: Simulated policy diffusion

(a) Coefficients from 1990s



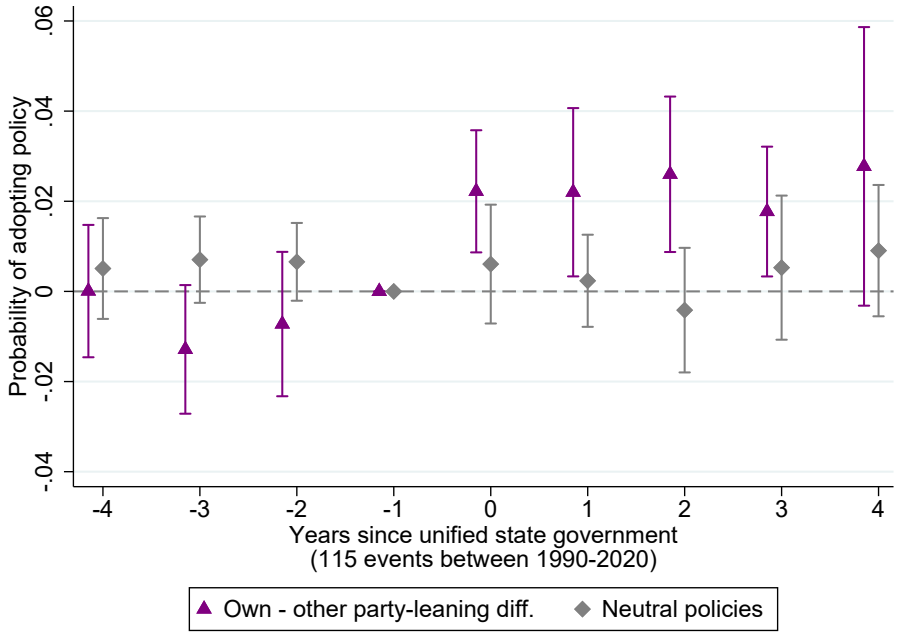
(b) Coefficients from 2010s



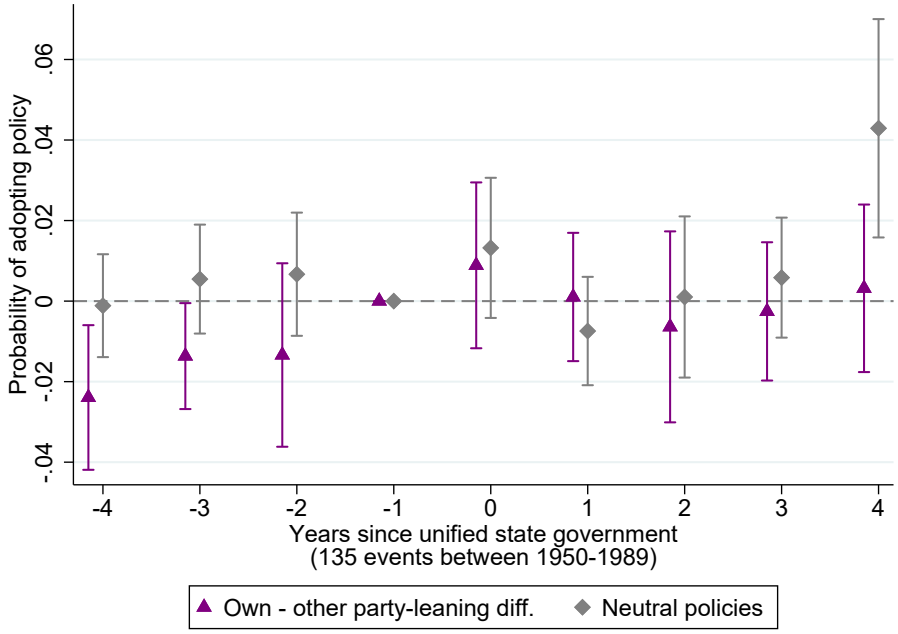
These maps show the probability of the diffusion of a policy innovated by California in 2009 for each of the other states based on the model estimated in Table 4.3. Figure 4.7a uses estimated coefficients from the 1990s decade, and Figure 4.7b from the 2010s decade.

Figure 4.8: Event study from switches in state government party control

(a) Events during 1990-2020



(b) Events during 1950-1989



Policies are included after 5 adoptions. Policies that ever switch ideological categorization from one party to the other (e.g., from left to right) are excluded. 95% confidence intervals are shown with standard errors clustered by state.

Figure 4.9: Comparison to polarization in DW-NOMINATE

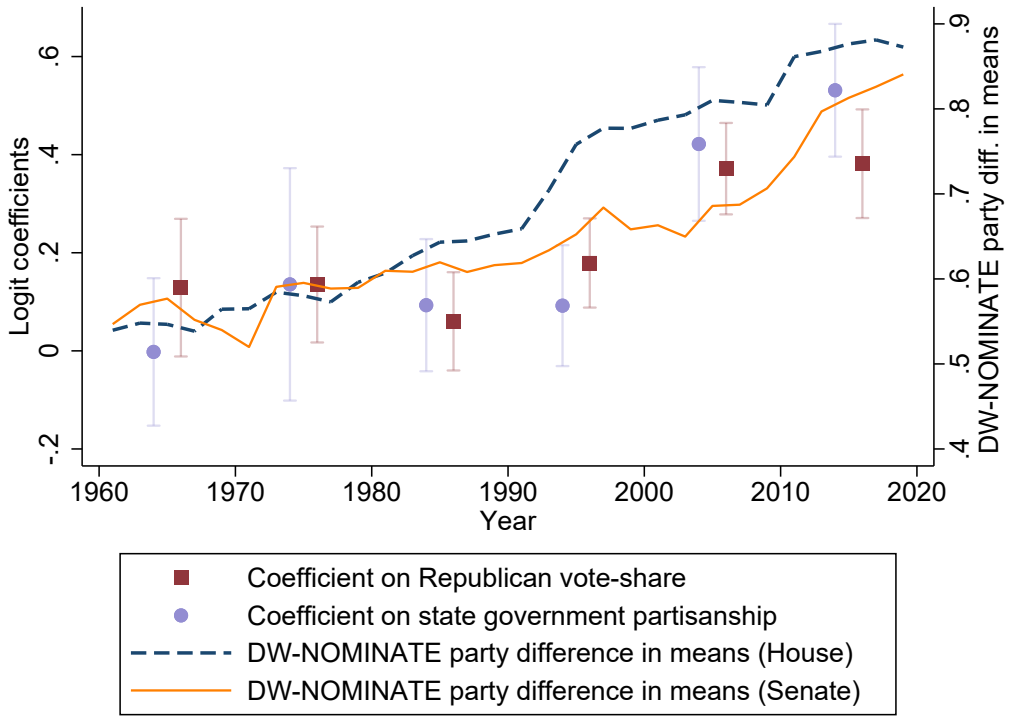
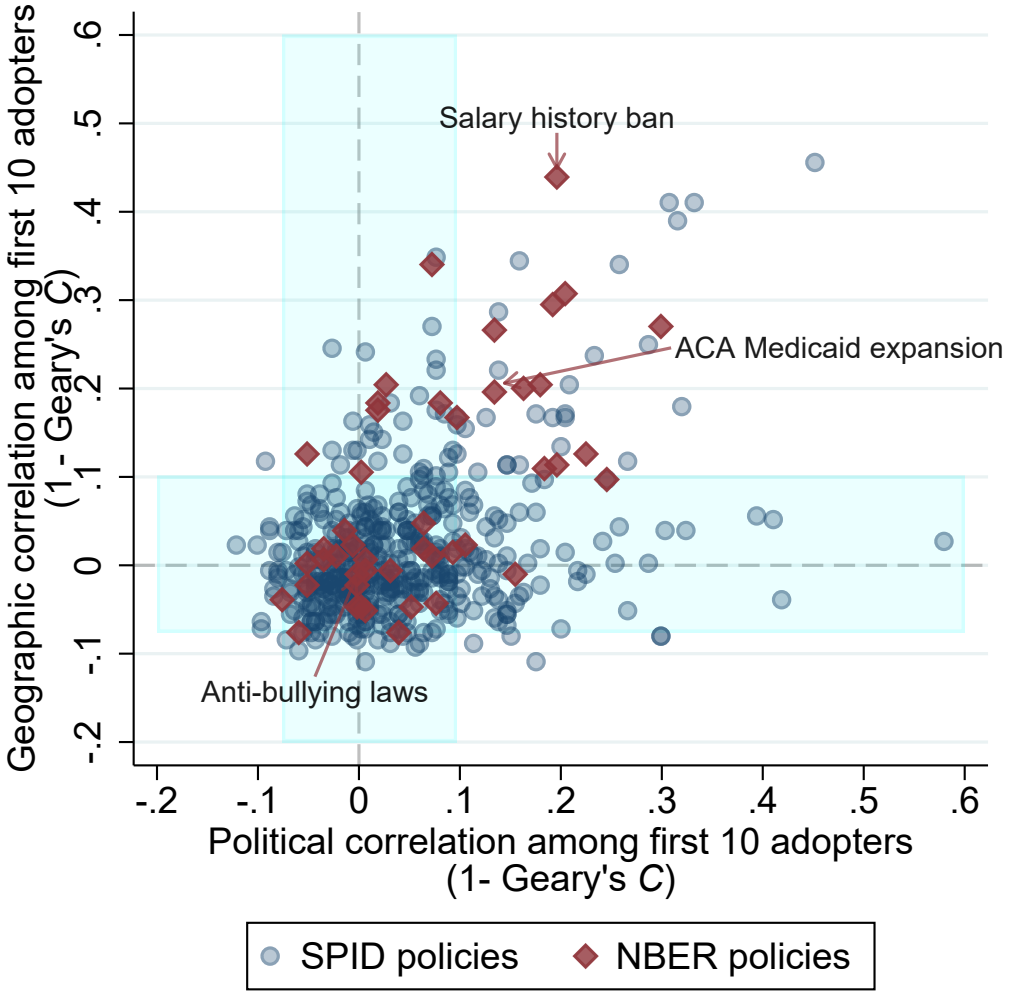


Figure 4.10: Policy-by-policy diffusion patterns



Geographic correlation mean: SPID=0.02, NBER=0.08, diff. p=0.00
Political correlation mean: SPID=0.05, NBER=0.07, diff. p=0.18
Shaded region indicate 5-95th percentiles for placebo policies

Table 4.1a: Summary of NBER data set

	(1) All (4/12 - 9/21)	(2) Cross-state policy	(3) Meets criteria*	(4) Sample
Total	11316	169	91	81
Issue date	2017.3 [2.7]	2017.6 [2.8]	2017.2 [2.8]	2017.5 [2.7]
Field				
% in Labor Studies	23	32	30	28
% in Public Economics	23	40	32	31
% in Economic Fluctuations and Growth	22	7	1	1
% in Health Economics	12	52	62	67
Other	41	15	11	10
Publication				
% Published	48	46	49	46
% Published in "Top General Interest"	9	4	1	0
% Published in "Tier A"	14	15	19	20
Year published	2017.3 [2.4]	2016.9 [2.3]	2016.6 [2.5]	2016.8 [2.6]
% Policy adoption data available	–	–	89	100
% Replication data available	–	–	–	9

Working papers numbered 18000-29318 are included. Means are reported with standard deviations in brackets for dates. Working papers can be listed under multiple fields. Papers on the same policy are all included in the sample. *Criteria: Policy must be binary and active after the 1950s. Covid-19 policies are also excluded.

Table 4.1b: Summary statistics of policy data sets

	SPID				NBER			
	Mean (SD)	Min	Median	Max	Mean (SD)	Min	Median	Max
Number of policies	648	–	–	–	57	–	–	–
First year of adoption	1977.76 (28.55)	1811	1983	2017	1987.81 (25.34)	1911	1995	2017
Last year of adoption	1998.89 (17.10)	1950	2002.5	2022	2007.30 (13.82)	1955	2014	2021
Number of states adopted	23.70 (15.23)	1	22	48	28.40 (15.08)	5	27	48

Policies with the last adoption before 1950 are dropped. Alaska, Hawaii, and Washington D.C. are excluded.

Table 4.1c: Policy areas

			Number of policies (freq.)	
Policy area	Main subgroups	Example	SPID	NBER
Public Services	Health, Education	Medical savings accounts	174 (27%)	28 (49%)
Law & Crime	Law & Crime	Gun open carry laws	184 (28%)	4 (7%)
Economics	Domestic Commerce, Labor	Bankrupcy laws	117 (18%)	20 (35%)
Civil Rights	Civil Rights, Immigration	Gender discrimination laws	105 (16%)	2 (4%)
Environment & Energy	Energy, Environment	Renewable energy standards	35 (5%)	2 (4%)
Gvnt. Operations & Foreign Affairs	Government Operations, Defense	Direct democracy	33 (5%)	1 (2%)

Table 4.2: Highest and lowest innovators (20%)

	1950-1990		1991-2020		Difference (SE)	
	(1) Top 20%	(2) Bottom 20%	(3) Top 20%	(4) Bottom 20%	(1)-(2)	(3)-(4)
Rep. two-party vote-share %	52.64	54.83	50.81	50.51	-2.19	0.30
	[8.67]	[13.42]	[10.09]	[9.80]	(2.30)	(3.94)
|Demeaned two-party vote-share|	5.06	8.48	8.23	7.57	-3.42	0.66
	[4.25]	[9.55]	[5.40]	[5.56]	(1.81)	(2.00)
Unified Dem. state gvt.	0.22	0.52	0.21	0.15	-0.30	0.06
	[0.42]	[0.50]	[0.41]	[0.36]	(0.14)	(0.08)
Unified Rep. state gvt.	0.23	0.25	0.33	0.37	-0.02	-0.05
	[0.42]	[0.43]	[0.47]	[0.48]	(0.10)	(0.15)
Legislative professionalism	0.81	-0.68	0.35	0.42	1.48	-0.07
	[1.90]	[0.84]	[2.23]	[1.95]	(0.63)	(0.90)
Log(population)	15.17	14.71	14.99	15.16	0.46	-0.17
	[1.07]	[0.98]	[1.05]	[1.08]	(0.45)	(0.46)
Income per capita	0.70	0.58	3.79	3.77	0.11	0.02
	[0.57]	[0.49]	[1.32]	[1.36]	(0.04)	(0.27)
Log(income per cap.)	8.53	8.31	10.48	10.47	0.21	0.01
	[0.80]	[0.86]	[0.35]	[0.37]	(0.06)	(0.07)
Urban pop. %	69.43	54.17	80.92	69.59	15.25	11.33
	[15.55]	[11.03]	[11.60]	[14.69]	(5.51)	(5.66)
Minority %	10.30	16.60	27.55	22.09	-6.30	5.46
	[8.81]	[10.51]	[14.46]	[11.98]	(4.03)	(5.50)
Unemployed %	6.72	6.56	5.27	5.41	0.16	-0.14
	[1.99]	[2.33]	[2.07]	[1.82]	(0.65)	(0.45)
States	11	10	11	11		

This table compares characteristics of the states in the highest and lowest 20% for first innovations. Averages are taken over the entire time period. Standard deviations are in brackets and standard errors in parentheses. Standard errors for the difference are clustered by state. Hawaii, Washington D.C., and Alaska are excluded.

Table 4.3: Policy diffusion predictors by decade

Dep. var.: Policy adoption (logit)	(1) 50-60s	(2) 70s	(3) 80s	(4) 90s	(5) 00s	(6) 10s
Proportion of states adopted	2.93	-0.03	1.72	2.99	2.20	2.84
	(0.22)	(0.23)	(0.17)	(0.13)	(0.19)	(0.24)
Standardized log(pop)	0.05	0.03	-0.01	0.04	-0.02	0.06
	(0.07)	(0.06)	(0.03)	(0.05)	(0.04)	(0.06)
Standardized income per cap.	0.11	0.23	-0.06	-0.04	-0.13	-0.07
	(0.10)	(0.06)	(0.04)	(0.06)	(0.04)	(0.08)
Standardized urban %	0.08	-0.12	0.13	0.12	0.16	-0.01
	(0.09)	(0.07)	(0.03)	(0.03)	(0.06)	(0.07)
Standardized Republican vote-share	-0.08	-0.11	-0.06	-0.00	-0.03	-0.10
	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.07)
Divided state government	0.10	-0.15	0.09	-0.03	-0.12	-0.04
	(0.10)	(0.08)	(0.06)	(0.05)	(0.08)	(0.09)
Measure of adoption among other states closest in:						
Demographic index (pop., income per cap., urban %)	0.40	0.12	0.21	0.18	0.23	0.24
	(0.07)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Distance	0.45	0.38	0.24	0.35	0.28	0.36
	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)	(0.07)
Republican vote-share	0.09	0.13	0.06	0.18	0.38	0.38
	(0.07)	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)
State gvnt. partisanship (unified Dem./unified Rep./divided)	0.04	0.15	0.11	0.11	0.41	0.52
	(0.08)	(0.11)	(0.07)	(0.06)	(0.07)	(0.07)
State gvnt. partisanship×Divided gvnt.	0.15	-0.41	-0.02	0.02	-0.45	-0.82
	(0.17)	(0.24)	(0.15)	(0.14)	(0.13)	(0.13)
Baseline P(Adopt)	0.03	0.03	0.03	0.05	0.05	0.06
Observations	58814	53268	75259	90165	69305	32602
Policies	162	196	272	380	330	194
Pseudo R^2	0.22	0.13	0.14	0.20	0.18	0.20

This table shows the coefficients from a logit regression. Standard errors are clustered by state. The baseline hazard for each policy is parametrized by policy fixed effects for each decade. The closest states are defined as the third of all the states with the smallest absolute value difference in each characteristic. The difference in the demographic index is calculated by first standardizing the two-year moving averages of log population, urban %, and log income per capita across all states in each year, then taking the absolute difference in each of the three standardized demographic variables, and finally averaging the three absolute standardized differences. The closest states in terms of distance are the third of states that have the smallest distance calculated using the centroid of the states. For Republic vote-share, the closest states are defined as the third with the smallest absolute difference in the vote-share for the Republican presidential candidate averaged over the most recent two elections. For state government partisanship, the closest states are defined as those with the same party control of state government (unified Republican, unified Democratic, or divided). We assign Nebraska, which has a unicameral nonpartisan state legislature, to the party of its governor. Alaska, Hawaii, and Washington D.C. are excluded from the analyses. The last year in the dataset is 2020, which is included in the 2010s decade. Only policies spanning at least 3 years with at least 5 adopters are included.

Table 4.4: Heterogeneity in policy diffusion

Distance			Republican vote-share			State gvnt. party control		
1950-70s	1980-90s	2000-10s	1950-70s	1980-90s	2000-10s	1950-70s	1980-90s	2000-10s
Dep. var.: Policy adoption (logit)								
Panel A. Source of policy								
NBER (R^2: 0.23, 0.25, 0.20; $N_{\text{pol.}}$: 14, 30, 43)								
0.63	0.36	0.46	0.09	0.27	0.50	-0.71	0.07	0.51
(0.15)	(0.11)	(0.10)	(0.15)	(0.11)	(0.08)	(0.21)	(0.14)	(0.08)
SPID (R^2: 0.16, 0.16, 0.17; $N_{\text{pol.}}$: 253, 390, 332)								
0.42	0.30	0.27	0.15	0.12	0.35	0.07	0.05	0.49
(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.06)
Interstate Compacts (within SPID) (R^2: 0.14, 0.13, 0.22; $N_{\text{pol.}}$: 22, 26, 15)								
0.47	0.70	0.09	0.13	-0.07	0.04	0.03	-0.12	0.28
(0.11)	(0.13)	(0.13)	(0.09)	(0.18)	(0.13)	(0.11)	(0.26)	(0.20)
Panel B. Policy area								
Economics (R^2: 0.10, 0.21, 0.19; $N_{\text{pol.}}$: 48, 63, 71)								
0.64	0.36	0.22	0.14	0.14	0.22	0.12	0.12	0.29
(0.09)	(0.07)	(0.07)	(0.12)	(0.07)	(0.08)	(0.09)	(0.10)	(0.10)
Non-Economics (R^2: 0.17, 0.16, 0.17; $N_{\text{pol.}}$: 219, 357, 304)								
0.38	0.29	0.33	0.15	0.13	0.41	0.02	0.05	0.53
(0.06)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.05)	(0.06)
Panel C. By state characteristics								
Third of states with highest Republican vote-share (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.32	0.32	0.28	0.10	0.12	0.34	0.08	0.07	0.43
(0.09)	(0.05)	(0.07)	(0.10)	(0.07)	(0.08)	(0.08)	(0.06)	(0.09)
Third of states with most neutral vote-share (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.38	0.30	0.27	0.11	0.05	0.16	0.03	0.07	0.43
(0.08)	(0.06)	(0.08)	(0.07)	(0.05)	(0.06)	(0.11)	(0.06)	(0.10)
Third of states with highest Democratic vote-share (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.58	0.30	0.34	0.16	0.19	0.57	0.05	0.04	0.59
(0.08)	(0.10)	(0.08)	(0.10)	(0.08)	(0.09)	(0.10)	(0.08)	(0.09)
Third of states with highest population (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.39	0.22	0.27	0.27	0.04	0.33	0.11	0.02	0.59
(0.09)	(0.09)	(0.09)	(0.06)	(0.05)	(0.06)	(0.12)	(0.08)	(0.05)
Third of states with lowest population (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.50	0.37	0.39	0.07	0.17	0.38	-0.05	0.02	0.39
(0.09)	(0.07)	(0.09)	(0.08)	(0.04)	(0.07)	(0.08)	(0.06)	(0.09)

This table predicts the diffusion of policies along geographic and political lines in several subsets of the data set. For each subset and time period (1950-70s, 1980-90s, and 2000-10s), a parsimonious diffusion model is estimated, which includes only (i) policy fixed effects, (ii) the proportion of adopters in all states, and the measure of adoption among the closest third of states in (iii) a demographic index combining population, income per capita, and urban % (see notes in Table 4.3 for details), (iv) geography, (v) Republican vote-share in the most recent presidential election, and (vi) state government party control (unified Democratic, unified Republican, or divided). The table shows coefficients on (iv), (v), and (vi) from the logit regression with standard errors clustered by state below in parentheses. The pseudo-R^2 and number of policies are reported in parentheses in chronological order corresponding to the three time periods.

In Panel A, the model is estimated separately for policies in NBER working papers, the SPID data set, and the Interstate Compacts source from the SPID data set. The Interstate Compacts are policies on which states cooperate to address a common problem.

In Panel B, the results are reported separately for policies in the "Economics" policy area and all other policies.

In Panel C, the states are first partitioned into thirds each year based on a characteristic (e.g., Republican vote-share in the most recent presidential election). The coefficients are then allowed to differ and reported separately for each third. The exercise is implemented for two characteristics: Republican vote-share and population.

Table 4.5: Models of policy diffusion: Role of migration and voter preferences

Dep. var.: Policy adoption (logit)	(1) 60-70s	(2) 80-90s	(3) 00-10s	(4) 60-70s	(5) 80-90s	(6) 00-10s
Proportion of states adopted	0.95	2.27	2.38	0.92	2.28	2.41
	(0.20)	(0.14)	(0.20)	(0.20)	(0.15)	(0.21)
Measure of adoption among other states closest in:						
Demographic index	0.46	0.18	0.21	0.32	0.09	0.13
	(0.08)	(0.05)	(0.06)	(0.08)	(0.05)	(0.06)
Distance	0.38	0.29	0.25	0.21	0.19	0.16
	(0.06)	(0.07)	(0.06)	(0.10)	(0.08)	(0.07)
Republican vote-share	0.19	0.13	0.38	0.14	0.09	0.27
	(0.07)	(0.04)	(0.05)	(0.06)	(0.04)	(0.05)
State gvnt. partisanship	0.09	0.10	0.44	0.08	0.10	0.42
	(0.13)	(0.07)	(0.08)	(0.12)	(0.07)	(0.08)
State gvnt. partisanship×Divided gvnt.	-0.15	-0.01	-0.55	-0.18	-0.04	-0.50
	(0.23)	(0.15)	(0.15)	(0.22)	(0.15)	(0.14)
Migration flows				0.14	0.01	0.12
				(0.13)	(0.10)	(0.09)
Voter preferences (ANES & GSS)				0.26	0.28	0.15
				(0.09)	(0.09)	(0.08)
Index of public opinion measures				0.19	0.16	0.20
				(0.06)	(0.04)	(0.04)
Observations	53757	107999	67504	53757	107999	67504
Policies	227	414	359	227	414	359
Pseudo R^2	0.16	0.17	0.19	0.16	0.17	0.19

This table shows the correlation in policy adoption among states that are closer in demographics, distance, Republican vote-share, state government partisanship, migration flows, voter preferences stated on ANES and GSS surveys, and an index of public opinion measures from political science. See Table 4.3 for the definition of the states closest in demographics, distance, Republican vote-share, and state government partisanship. All regressions include an indicator for divided state governments as a control (coefficient not reported). For migration flows, the closest states are defined as the third with the highest sum of in- and out-migration. For voter preferences, the closest states are those with the smallest average difference in standardized responses on ANES and GSS questions regarding policy preferences. 15 states are excluded as they do not have sufficient representation to measure voter preferences in the ANES and GSS surveys (see Online Appendix Section B). For the index of public opinion measures, we standardize the Berry et al. (1998) revised 1960-2016 citizen ideology series, the Lagodny et al. (2022) state-level public policy mood measure, and the Caughey and Warshaw (2018) mass social and economic liberalism scores in each year, and average the absolute differences between each pair of states. The closest states are defined as the third with the smallest average difference. Each column reports a separate logit regression within the time period indicated in the header. Policy-by-decade fixed effects are included as the baseline hazard rate for each policy. Standard errors clustered by states are in parentheses.

Table 4.6: Vaccine regulations and COVID-19 policies

	COVID		Vaccine laws	
Dep. var.: Policy adoption (logit)	(1)	(2)	(3)	(4)
Proportion of states adopted	3.31	3.04	1.37	1.46
	(0.24)	(0.27)	(0.46)	(0.47)
Measure of adoption among other states closest in:				
Demographic index	0.21	0.22	0.37	0.18
	(0.08)	(0.13)	(0.11)	(0.14)
Distance	0.31	0.37	0.20	-0.10
	(0.06)	(0.14)	(0.09)	(0.13)
Republican vote-share	0.16	-0.08	0.04	-0.07
	(0.09)	(0.10)	(0.10)	(0.11)
State gvnt. partisanship	0.47	0.58	-0.19	-0.26
	(0.10)	(0.14)	(0.12)	(0.16)
State gvnt. partisanship×Divided gvnt.	-0.30	-0.48	0.30	0.59
	(0.19)	(0.29)	(0.28)	(0.30)
Migration flows		0.21		0.45
		(0.16)		(0.11)
Voter preferences (ANES & GSS)		0.08		-0.18
		(0.18)		(0.22)
Index of public opinion measures		-0.15		0.01
		(0.11)		(0.12)
Observations	27751	10944	22646	15174
Policies	76	64	28	28
Pseudo R^2	0.33	0.33	0.17	0.19
Time unit	Weeks (Mo-Su)	Weeks (Mo-Su)	Years	Years
Time range	10/2019-8/2021	10/2019-12/2020	1980-2020	1980-2020

This table shows the coefficients from a logit regression. Standard errors are clustered by state. The baseline hazard is parametrized by policy-decade fixed effects for vaccine laws and policy-month fixed effects for COVID policies. See Tables 4.3 and 4.5 for the definition of closest states in each characteristic. Alaska, Hawaii, and Washington D.C. are excluded from the analyses. In Columns 2 and 4, only the 33 states with measures of voter preferences from the ANES and GSS surveys are included (see Table 4.5 notes). Only policies spanning at least 3 time periods with at least 5 adopters are included.

Chapter 5

Bottlenecks for evidence adoption

Governments increasingly use RCTs to test innovations, yet we know little about how they incorporate results into policy-making. This chapter studies 30 U.S. cities that ran 73 RCTs with a national Nudge Unit. Cities adopt a nudge treatment into their communications in 27% of the cases. We find that the strength of the evidence and key city features do not strongly predict adoption; instead, the largest predictor is whether the RCT was implemented using pre-existing communication, as opposed to new communication. We identify organizational inertia as a leading explanation: changes to pre-existing infrastructure are more naturally folded into subsequent processes.

This chapter is co-authored with Stefano DellaVigna and Elizabeth Linos.

5.1 Introduction

In a drive to incorporate evidence into their policy-making, governments at all levels have increasingly rolled out RCTs to test policy innovations before scale up (e.g., Baron, 2018; Foundations for Evidence-based Policymaking Act, 2018; DIME, 2019).

This experimentation has the potential to improve public policy. But how often are the innovations tested in RCTs actually adopted? To what extent do factors other than the strength of the evidence moderate this adoption, such as state capacity, turnover of personnel, or organizational inertia?

Table 5.1 summarizes the limited evidence. A first set of papers, e.g., Vivalt and Coville (2023), Mehmood et al. (2022), Nakajima (2021), and Toma and Bell (2022), examine policymakers' interest in adopting policies in mostly hypothetical scenarios. A second set examines the adoption of one intervention; e.g., Hjort et al. (2021) show that Brazilian mayors who received information on a successful taxpayer reminder letter from RCT evidence are more likely to adopt the communication. A third group, to which our study belongs, examines how multiple institutions incorporate the results of different experiments, e.g., Kremer et al. (2019) documenting the scaling of 41 USAid-funded RCTs and Wang and Yang (2022) examining policy experimentation by cities in China. Studies in the third group have the advantage of allowing comparison of variation in both institutions and in features of the interventions—such as effect size—on adoption.[58]

In this paper, we bring new evidence to bear from the Behavioural Insights Team-North America (BIT-NA). During the period under study, BIT-NA supported primarily North American cities to develop or revise light-touch government communications (e.g., a letter or an email) aimed at improving policy outcomes of interest to the city, such as the timely payment of bills or the recruitment of a diverse police force. The behavioral scientists at BIT-NA and the staff members in the relevant city department co-designed different versions of a given communication and tested what works using an RCT. Compared to most settings, these RCTs have relatively low barriers to adoption, as the innovations are light-touch and low-cost, the evidence is developed in the relevant context, key stakeholders are involved in designing and approving the innovation, and political or other feasibility barriers are cleared before the implementation of the RCT.

BIT-NA shared all the records on their RCTs conducted between 2015 and 2019. As documented in DellaVigna and Linos (2022), the average nudge intervention in these 73 trials over 30 cities increases the outcome of interest by 1.9 percentage points, a 13% increase relative to the baseline average of 15%, with substantial heterogeneity in the effect size. However, this data set does not indicate whether the nudge innovation is adopted in subsequent communication by the city. This is not surprising, as datasets tracking adoption, as in Kremer et al. (2019), are rare.

Thus, over the course of a year, starting in March 2021, we contacted each city department involved and asked about the adoption of the featured communication, as well as additional information about the implementation such as staff retention. We are able to assess the adoption for *all* 73 RCTs and can thus estimate the rate of evidence adoption and its determinants. We compare these results to predictions by researchers and Nudge Unit

[58] Table 5.1 also includes some studies examining adoption of the results of experimentation in firms, where the evidence is similarly mixed and limited. See also Athey and Luca (2019); List (2022).

staff members, along the lines of DellaVigna, Pope, and Vivalt (2019).

Before turning to results, we emphasize some features of our setting that make it a good fit to evaluate the adoption of the treatment innovations. For one, we observe the entirety of RCTs run by this unit and their adoption, not just the successful cases. Also, the sample of RCTs is large enough to grant statistical power, and yet the RCTs are comparable enough to enable inference. Furthermore, there is sufficient variation in the effectiveness of the interventions, the characteristics of the policy partner (the city), and the design of the trials, to provide evidence on a range of adoption predictors.

We first document the overall level of adoption. Out of 73 trials, the nudge innovation is adopted in post-trial communications by the city 27% of the time. This level is comparable to the average prediction of forecasters (32%) and the average adoption among comparable studies (Table 5.1).

We then consider three determinants of adoption: (i) the strength of the evidence—statistical significance and effect size—which is the normative benchmark, provided that the effect sizes after adoption are related to the RCT estimates; (ii) features of the organization (city), such as the "state capacity" of the city and whether the city staff member working on the RCT is still involved; and (iii) the experimental design, namely the type of nudge treatment, and whether the communication was pre-existing or new.

We find surprisingly limited support for the role of evidence in adoption. There is no difference in adoption for results with negative point estimates (25% adoption), results with positive but not statistically significant estimates (25%), and estimates that are positive and statistically significant (30%). The likelihood of adoption increases with effect size (measured in percentage points), from 17% in the bottom third to 38% in the top third, though this difference is not statistically significant at conventional levels. Along both dimensions, the impact of the evidence is less than what forecasters expect.

Next, we find modest predictive power of organizational capacity, proxied by city population (32% for larger cities versus 22% for smaller ones) and the certification by What Work Cities as a "data-driven" city (30% versus 24%). Adoption is more likely when the city contact for the RCT is still employed by the city (33% versus 19%).

We thus turn to the last set of factors, the experimental design. The adoption rate is somewhat higher for interventions involving simplification (33%), as opposed to personal information and social cues (19% and 24% respectively).

The strongest predictor by far is whether the communication in the trial was pre-existing or new. In the 21 trials for which a pre-existing city communication had been modified to contain the nudge, the adoption rate is 67% (14 out of 21). Conversely, in the 52 trials for which no similar communication had been sent prior to the collaboration with BIT-NA, the adoption rate is only 12% (6 out of 52). This 55 pp. difference, which is highly statistically significant (t=4), is far beyond the expectation of academics and members of Nudge Units, who expect only a 11 pp. difference. This impact is not only large but also robust, at 60 pp. (s.e.=0.15) when including all controls.

How do we interpret these findings, and especially the key impact of pre-existing communication? We discuss four potential mechanisms: (i) *cost of materials*, (ii) *state capacity*, (iii) *unobservable features*, and (iv) *organizational inertia*.

First, the *cost of materials* is already included in the city budget for pre-existing communications, but are not for new communications in the years to come. However, our findings

are similar for online communications, which have near zero marginal cost, as for paper communications, which require some financing for material costs.

Second, cities or departments with pre-existing communications may have better staffing and infrastructure, which is why they were already sending the communications (*state capacity*). However, we find the same adoption gap between pre-existing and new communications even within city, after controlling for city fixed effects. Still, there could be within-city differences in communication infrastructure at the department level. To make progress, we collect a sample of online forms from these city department websites as a proxy for communication infrastructure, comparing to the same city departments in nearby cities with comparable population. We find that the availability and rate of change in online forms for the departments in our BIT-NA sample are very similar to that in the comparison cities. We conclude that the extent to which departments vary in their pre-existing communication infrastructure is similar to the patterns in other cities and does not reflect some unique state capacity of the departments in our sample.

Third, as we outline in a simple model, *unobservable variables*, such as prior beliefs of the policymakers or political concerns, may be correlated with pre-existing communication in a way that explains the results. While we cannot control for unobservables, controlling for a number of features of the interventions does not reduce the estimated impact of pre-existing communication at all.

Thus, we argue that the primary interpretation is *organizational inertia*. Consider a two-step adoption process, with a first decision—whether to send any communication—and a second decision—designing the content. Setting up a new communication can imply substantial organizational costs, while content changes are low cost. In cases with pre-existing communication, there is a routine process and staffing in place to send the communication, so the first step is not a hurdle, and altering the wording to adopt an effective innovation is relatively straightforward, leading to high adoption. In cases with a new communication set up for the experiment, instead, there is no automatic pathway to send it again, leading to low adoption. Indeed, the low adoption for experiments with new communication is entirely driven by cities sending no communication following the RCT, as opposed to sending something other than the nudge versions in the experiment.

Organizational inertia can be caused by a broad set of factors, including low prioritization or insufficient staffing. To make progress on narrowing the potential pathways, we survey two samples—city employees responding for 25 trials that did not adopt a nudge treatment after finding positive effects, and 45 city policymakers not involved in the trials. We ask which among seven factors would help them most to adopt successful nudges in communications. City employees in both samples indicate that prioritization from decision-makers is a key factor, above logistical support, funding for communication, staffing costs, and staff training. Stronger evidence and the provision of simple reminders are rated as the least helpful. These results suggest that a key bottleneck is likely the *allocation* of resources by leadership to prioritize adoption.

The limited adoption of evidence has a large economic impact. If all the effective nudges had been adopted, the RCTs would have increased the targeted outcome on average by 2.70 pp. (assuming the effect sizes are stable over time). In contrast, the actual improvement is estimated to be 0.89 pp., just one third of the potential gains. This gap is almost entirely due to the RCTs with new communication.

An important question is how our findings compare to other settings, such as non-behavioral interventions and RCTs in lower-income countries. The level of adoption in our study, 27 percent, is in the range of the (few) estimates in the literature (Table 5.1). Regarding the key role of pre-existing communication, Kremer et al. (2019) also reports that scaling is higher for RCTs using established channels of distribution. Further, we re-analyze the data from Hjort et al. (2021) and estimate a larger persuasive impact from providing evidence to Brazilian cities that already were sending a communication than to cities that were not (with the caveat that an alternative model can also rationalize these effects). We hope that future papers will also compare the effect size in an RCT to other determinants of adoption, such as organizational inertia rooted in pre-existing communication. As far as we know, ours is the only paper that does this comparison.[59]

The paper relates to the literature on nudges (e.g., Thaler and Sunstein, 2008; Benartzi et al., 2017; Milkman et al., 2021) and on experimental design (Kasy and Sautmann, 2021). Our findings suggest that anticipating the bottlenecks to adoption may change the experimental design to prioritize treatments that are likely to be adopted if effective as well as the allocation of resources to target adoption.

The paper also relates to the literature on scaling RCT evidence (Banerjee and Duflo, 2009; Allcott, 2015; Muralidharan and Niehaus, 2017; Meager, 2019; Vivalt, 2020). The Nudge Unit interventions were already "at scale" in terms of sample size, since they applied nudges in the literature to a large population in the policy setting, as documented in DellaVigna and Linos (2022). We focus on the temporal dimension of scaling: the translation and adoption of RCT results into continuing government practice.

Finally, the paper is related to the literature on organizational inertia and learning (Levitt and March, 1988; Simon, 1997; Argote and Miron-Spektor, 2011). The fact that the key mediating variable for adoption was not foreseen suggests that more emphasis on organizational processes will be important in future studies.

5.2 Setting and data

5.2.1 Trials by Nudge Unit BIT-NA

Nudge Units In 2015, the UK-based Behavioural Insights Team (BIT) opened its North American office, BIT-North America (BIT-NA), partially in support of the "What Works Cities" initiative to provide technical assistance to mid-sized cities across the U.S. This team, like other "Nudge Units," aims to use behavioral science to improve the delivery of government services through rigorous RCTs, and to build the capacity of government agencies to use RCTs independently. BIT-NA has collaborated with over 50 U.S. cities to support the implementation of behavioral experiments within local government agencies. In interviews, the leadership noted that the primary goal of these experiments is to measure "what works" in moving key policy outcomes.

[59] In Table 5.1, papers in the second group cannot study the role of different effect sizes as they provide evidence from only one RCT. Among papers in the third group, Kremer et al. (2019) computes the benefit-cost ratio for four interventions that scaled, but does not compare the effect size across RCTs, and Wang and Yang (2022) documents that the city-level impacts of the innovations are likely biased by site selection and politicians' extra efforts and thus should not be interpreted as RCT effect sizes.

The vast majority of their projects during the period under study are RCTs, with randomization at the individual level, involving a low-cost nudge delivered as a letter or online communication (such as email), targeting a behavioral variable, such as reducing late utility bill payments. Figure D.1a-b shows an intervention aimed to increase the payment of delinquent fines from traffic violations, with a status-quo letter in the control group (Figure D.1a) and a simplified letter in the treatment group (Figure D.1b). The outcome is the share of recipients making a payment within three months.

Process of experimentation As the left panel of Figure 5.1a shows, trials are developed in collaboration with a city department that is interested in working with BIT-NA on a policy area of interest. In most cases, scoping calls between a city staff member and a BIT-NA behavioral scientist help determine if an RCT is feasible, by defining a behavioral outcome of interest, estimating the potential sample size, and confirming the possibility for a scalable light-touch intervention. Unlike purely academic research, most trials are explicitly designed with feasibility of adoption in mind.

Once BIT-NA confirms that a well-powered trial is possible, department staff and other city stakeholders (e.g., legal and communications teams) collaborate with behavioral scientists at BIT-NA to co-design the specific intervention and evaluation plan. This stage is relevant for potential adoption—many of the hurdles for scaling up evidence such as legal or political barriers have already been overcome at the RCT design stage. Moreover, in selecting the intervention, the team aims to only test interventions that the city could plausibly adopt, should they work. The city staff involved in designing and implementing the trial are also the ones who would be involved in adoption, assuming no major changes in department leadership or key players. Before running the trial, the intervention and evaluation design as well as the related hypotheses are recorded. While the technical assistance that covers the behavioral and evaluation design is free from the perspective of the given department, the city bears any labor or material cost related to actually implementing the intervention.

Following the RCT, the BIT-NA staff analyze the results and produce a non-technical report typically a few pages long that is shared with the city alongside a presentation to the relevant stakeholders, including city leadership (e.g., an example in Appendix D.1). This, as well as the ongoing collaboration for the purposes of RCT implementation, should ensure that the relevant players can understand and act on the evidence. Indeed, even years post-implementation, several of the staff contacts in the cities reported remembering the results, and in 14 cases out of 15 cases, they recalled them correctly.

Sample of trials We select our sample similarly to DellaVigna and Linos (2022), which analyzed the average treatment effects of the RCTs run by BIT-NA, as well as by the Office of Evaluation Sciences (OES). As Figure 5.1b shows, from the universe of 93 trials conducted between 2015 and 2019 by BIT-NA, we remove 2 trials that are not RCTs in the field, 8 trials without a clear "control" group, 3 trials with monetary incentives, and 2 trials without a binary primary outcome. Compared to the sample in DellaVigna and Linos (2022), we exclude 8 trials run with partners other than U.S. cities (charities and cities in Canada and Africa), in order to focus on a more comparable set of policymakers. Finally, while contacting cities, we identified and added 3 additional trials run by the same cities in collaboration with BIT-NA in subsequent years. This yields the final sample of 73 trials run in collaboration with 67 city departments in 30 cities (given that BIT-NA often works with multiple departments within a city).

An important question that may impact adoption is how the trials and cities are selected. While a full examination is beyond the scope of this paper, we present two pieces of evidence. First, in Table D.1 we compare the 73 trials in our sample to 27 trials that BIT-NA began with a partnering city and listed in their internal records, but abandoned before completing the RCT due to logistical or bureaucratic obstacles. The cities in the two samples have similar features, except in the median age of their residents. Second, we compare city departments in the BIT-NA sample to departments from cities with similar population size in the same census region, with respect to two measures of bureaucratic effectiveness: the availability of forms online, as a proxy for broader communication capacity, and the extent to which such forms are updated over time, as detailed in Appendix D.2. As Figure D.2 shows, the city departments in the BIT-NA sample are comparable to other city departments on these measures.

Impact of nudges DellaVigna and Linos (2022) estimate the average impact of nudges in terms of percentage point on the policy outcome, relative to the control group. We reproduce the regression in Column 1 of Table D.2, and in Column 2, we present the average for the city sample used in this paper. For BIT-NA trials, we estimate an impact of 1.9 percentage points (s.e.=0.6), a 13% increase relative to the control group average outcome level of 15.1%. In Figure 5.2 we present the trial-by-trial evidence for the BIT-NA sample, plotting the effect size for the most effective nudge arm compared against the take-up of the targeted outcome in the control group. The figure also denotes the adoption and the pre-existence of the trials, two key aspects we revisit later.

Features of trials In Column 1 of Table 5.2 we describe the characteristics of the 73 trials, starting with the effect size: 45% of the trials have at least one arm with a positive and statistically significant effect size, and 47% have at least one arm with an effect size larger than 1 percentage point. Next, we consider organizational features of the city: whether the city has been certified by What Works Cities, which uses a set of criteria to validate that a city is a "data-driven, well-managed local government", and whether the city contact for the trial is still employed by the same city department. We also measure the seniority of the city staff working on the trial (i.e., whether one of the city staff is the department director or chief) and whether the partnering city department delivers the communication directly (e.g., a Codes Enforcement department sends the notice for code violations), as opposed to it collaborating with multiple departments (e.g., an Innovation Team or a Mayor's Office team).

We then categorize the experimental design: whether the communication was pre-existing before the trial, and the behavioral mechanisms used. There are typically multiple mechanisms within a treatment, including simplification with clear instructions and plain language (53% of trials); personalizing the communication or using loss aversion to motivate action (58% of trials); and exploiting social cues or norms (56% of trials).

Next, we consider the policy area. A typical "revenue & debt" trial nudges people to pay fines after being delinquent on a utility payment, while an example of a "registration & regulation" nudge asks business owners to register their business online as opposed to in-person. The "workforce and education" category includes prompting police applicants to show up for their in-person examination. One "benefits & programs" trial encourages households to apply for a homeowners tax deduction. A "community engagement" intervention motivates community members to attend a town hall meeting and a "health" intervention urges people

to take up a free annual physical exam. The most common categories are revenue & debt, registration & regulation, and workforce & education.

Finally, the communication is delivered via a physical medium in the majority of cases, physical letter (38%) or postcard (22%), as opposed to online or digital delivery.

Columns 2 to 7 characterize subsamples split by the median effect size (Columns 2 and 3), by whether the original city collaborator has been retained (Columns 4 and 5), and by whether the communication is pre-existing or new (Columns 6 and 7). There are some differences in the characteristics of trials, e.g., pre-existing communications are more likely to be physical letters and to feature simplification. These correlations highlight the importance of controlling for these characteristics. In Table D.3 we expand this comparison to other city features, finding very limited evidence of differences.

5.2.2 Adoption of nudge treatments

The BIT-NA record for each trial, as comprehensive as it is, does not include whether the city adopted the nudge treatments in their communications following the RCTs.

We emailed each city department involved in the RCTs and followed up with emails and occasionally phone calls. Collecting the full data set took one year and an average of four interactions with each department (Figure D.3). In our conversations with the city staff, we first described the past collaboration with BIT-NA, provided the templates of the communications sent out in the trial, and asked whether the city was still sending the communication. If so, we asked them to send us the current version. If they were not sending the communication, we confirmed whether they had sent the communication anytime after the trial, even if they were no longer doing so (e.g., due to COVID). In addition, we asked whether the communication had been used before the trial or was sent for the first time in the trial itself (i.e., whether it was pre-existing or new). We also checked whether the city staff members who worked on the trial were still employed by the city. We took note when they referenced the results of the trial (which we did not reveal) and recorded any barriers to adoption that they mentioned.

Ultimately, we were able to obtain responses about the adoption for all 73 RCTs. We define adoption as the case in which "*one nudge treatment arm has been used in communications from the city department after the RCT*". Given that the nudge arm was never the status-quo communication, adoption corresponds to a policy change. In the large majority of cases, whether a nudge treatment arm was adopted was straightforward to code. For the example in Figure D.1, the most recent communication (Figure D.1c) is clearly based on the nudge treatment letter (Figure D.1b), and is thus a case of adoption. In other cases, the recent communication resembles the communication in the control group, or there is simply no communication sent out in the years following the RCT; we code these cases as instances of non-adoption. We validate our coding with a machine-based measure of similarity in content between the current version (when available) versus the control and treatment forms in the RCT, as documented in Appendix D.3.

In a small number of cases, documented in Appendix D.4, the coding of adoption is not obvious. Where there are multiple components to the intervention, we define a case as adopted if at least 50% of the nudge components pre-specified in the BIT-NA trial protocol are present in the post-trial communication. We also count cases as adopted if the city is no

longer sending the communication in 2021 or 2022 (e.g., due to COVID), but had used the nudge communication at some point after the RCT.

5.2.3 Other forms of adoption

While we focus on the adoption of the nudges for an objective criterion of adoption linked to the RCTs, the city contacts occasionally noted that the trials had motivated the city to (a) use nudges in other contexts, or (b) run their own RCTs for other city communications or services. We consider both as cases of "broad adoption", as described in Appendix D.5. The former case occurs at the trial level when the city uses a communication that is distinct from, but inspired by, a nudge tested. For example, a city department sent text reminders for show-cause hearings as part of a trial, but did not continue these reminders; instead, the department now sends similarly worded texts for citations. The latter case of broad adoption occurs at the city level, when a city notes that they conducted additional RCTs after learning the process of experimentation from their collaboration with BIT-NA. It does not include cases where a different city implemented a communication, based on evidence from a city in our sample.

5.2.4 Forecasts of results

Forecast Survey Along the lines of DellaVigna, Pope, and Vivalt (2019), we compare the results to the predictions of forecasters, to capture the direction of updating. We posted on the Social Science Prediction Platform a 10-minute Qualtrics survey (reported in the Appendix D.6) before the results were posted publicly.

Specifically, after presenting the setting and the question, we asked for (i) a prediction of the average rate of adoption for the 73 nudge RCTs; (ii) an open-ended question on possible reasons for non-adoption: "*When cities do not adopt the nudges from the trials, what do you think are the main reasons?*"; (iii) the prediction of how adoption would vary as a function of 7 determinants, 2 about strength of evidence (1 on effect size, 1 in statistical significance); 3 about city characteristics (1 about staff retention, 1 about state capacity, 1 about certification as an evidence-based city); 2 about experimentation conditions (1 about nudge content and 1 about pre-existing communication); (iv) a qualitative assessment of how the likely adoption of evidence in this context would differ from the adoption of evidence in firms and in RCTs run in low-income countries.

We obtain 118 responses, as detailed in Table D.4, with 19 response from individuals affiliated with Nudge Units, 67 researchers (university faculty, post-docs, and graduate students), and 14 government workers, among others.

5.3 Framework

To motivate the analysis, consider a policymaker that collects evidence (a signal) about the effectiveness of the nudge treatment, compared to a control. The policymaker has a prior $\pi_0 \sim N(\mu_0, \sigma_0^2)$ about the relative effectiveness of the treatment; the mean prior μ_0 is positive if for example the policymaker believes that the nudge wording is likely more effective. The prior is likely to be more positive for experiments that were more costly to run, to justify

running the experiment itself. While we do not model this preliminary stage of experimental design, we return to this observation when discussing the results.

The experimental results come in the form of a Normal signal $s_i \sim N(\mu_{s,i}, \sigma^2_{s,i})$, where the variance depends on the statistical power of the experiment i. Combining the prior with the signal, the policymaker has a posterior $\pi_{1,i}$ about the effectiveness, with mean $\mu_{1,i} = \frac{\sigma^2_{s,i}}{\sigma^2_0+\sigma^2_{s,i}}\mu_0 + \frac{\sigma^2_0}{\sigma^2_0+\sigma^2_{s,i}}s_i$. The decision maker will adopt the innovation ($D_i = 1$) in trial i if the expected utility is better than the alternative ($D_i = 0$). We model this as

$$\frac{\sigma^2_{s,i}}{\sigma^2_0 + \sigma^2_{s,i}}\mu_0 + \frac{\sigma^2_0}{\sigma^2_0 + \sigma^2_{s,i}}s_i + \beta X_i - \gamma C_i + \epsilon_i \geq 0.$$

We observe the signal s_i (the effect size for nudge i) and its variance ($\sigma^2_{s,i}$) as implied by the statistical power. We also observe other characteristics X_i of the treatment that may affect the adoption, and, in particular, proxies for the cost of implementing the nudge C_i, such as the organizational capacity of the city and the retention of staff members involved in the experiment. At the same time, we do not observe the priors of the policymaker. Under the assumption of a logistic distribution for the error term, the specification can be estimated as a logit. We also estimate a simple OLS model.

We estimate the model under the assumption that the parameters for the prior, μ_0 and σ^2_0, are independent of trial i. In this model, some nudge treatments with negative effect sizes could still be adopted both because of the error term and if the policymakers have stronger positive priors. Larger effect sizes should, however, increase the likelihood of adoption.[60] Other determinants, X_i and C_i, will mediate the adoption.

More generally, though, the priors can vary across treatments in ways the researcher cannot observe. In principle, this can reconcile any pattern of results: a feature X_i may be correlated with adoption not because it has a direct effect, but because it is correlated with the unobservable priors. We discuss below the plausibility of this confound.

5.4 Results

5.4.1 Average adoption

In Figure 5.3 we display three plausible benchmarks for the rate of adoption. As the first columns show, 78% of the trials have at least one nudge arm with a positive effect size, and 45% of the trials have a nudge arm with a positive and statistically significant increase. Compared to these two benchmarks (which were shown in the survey), forecasters predict a lower adoption rate, at 32%, with forecasters working in nudge units being slightly more optimistic, with a forecast of 37% (Table D.4).

As the final column shows, the average rate of adoption is 27%, that is, adoption in 20 out of 73 trials. The result is not statistically significantly different from the average forecast, though it is significantly lower than the initial two benchmarks.

[60]The policymakers may also display non-Bayesian updating and be more responsive to positive results (Vivalt and Coville, 2023), leading to a higher impact of positive effect sizes on adoption.

5.4.2 Determinants of adoption and survey predictions

The forecasters gave their open-ended opinions in the survey on the bottlenecks for evidence adoption before the survey highlighted the specific channels we focus on. As the word cloud in Figure 5.4 shows, they stress the potential importance of effect size ("small", "lack" and "effect"), organizational inertia ("inertia" and "status quo"), cost of implementation ("cost" and "budget"), and the staff ("staff", "people", and "turnover"). Thus, the survey respondents highlight some of the key channels we now turn to.

5.4.3 Adoption: Evidence-based determinants

To the extent that the long-term expected impact of a communication is monotonically related to the results in the RCTs, the rate of adoption should be related to the effect size and statistical significance of the nudge arms in the RCT, as implied by the framework.

In Figure 5.5a we split the RCTs into thirds by the percentage point effect of the most effective nudge arm in each trial. The first three bars show that, on average, the forecasters expect an adoption rate of just 13% in the lowest third, and of 49% in the top third. In reality, the adoption is increasing in effect size—17% in the bottom third for effect size, 28% in the middle third, and 38% in the top third—but the impact is smaller than forecasted, and is not statistically significant at conventional levels. Considering the evidence in 10 bins in Figure 5.5b, the responsiveness to effect size is quite tentative.

It is possible though that cities are responding even more to statistical significance than to effect size. The two measures differ because the arms are not equally powered (though they are generally well powered, compared to a typical academic paper on nudges, as documented in DellaVigna and Linos, 2022). On average forecasters expect a strong response by statistical significance (Figure 5.5c). In reality, the rate of adoption is nearly the same for results that are negative or zero (25%), positive but not statistically significant (25%), or positive and statistically significant (30%). Thus, statistical significance does not seem to play a role in adoption.

A possible explanation for this lack of response is that BIT-NA may lean on factors other than evidence in their recommendations to cities to either adopt or not adopt a treatment arm. As Figure D.4 and Table D.5 show, though, statistical significance is the major determinant of BIT-NA's recommendations in the 28 trial reports (starting in mid-2017) that record explicit recommendations for or against adoption.

We consider one final component: for RCTs with multiple nudge treatment arms, one of which is adopted, is the treatment with the highest effect size adopted? Indeed, this is the case in 5 out of 6 such trials (Figure 5.5d). Thus, when adoption takes place, effect size *does* play a key role.

The framework in Section 5.3 suggests two explanations for this limited response to effect size. First, the city officials may have strong priors and are therefore only partially moved by the evidence. Second, other factors, such as those related to the cost of implementing the treatments, predict adoption. We turn to some of these factors next.

5.4.4 Adoption: Organizational features

Some organizations may have more "organizational slack" or state capacity to enact reforms (Besley and Persson, 2009). Organizational features that may drive or hinder adoption of evidence are size, wealth, and personnel (Naranjo-Gil, 2009; Fernandez and Wise, 2010; see de Vries, Bekkers, and Tummers, 2015, for a systematic review). In our framework, these determinants could lower the costs of adoption. Many studies also point to the impact of political constraints, external pressures, or outside networks. In our setting, such factors are likely less important since the innovations tested have been vetted for political, legal, and communications feasibility.

We measure "state capacity" with two proxies, starting with city population. As Figure 5.6a shows, there is a moderate difference in adoption by city size, with 22% adoption in the smaller cities, and 32% adoption in the larger cities. As a second proxy, we consider the certification from What Works Cities described in Section 5.2.1. As Figure 5.6b shows, there is a more modest difference along this line, 24% versus 30%.

A different dimension is the personnel. We measure if at least one of the original city staff members who helped to design and implement the experiment is still working in the same city department at the time of contact.[61] If so, it is more likely that the city has an internal "champion" with the expertise and the institutional memory to continue the nudge innovation.[62] As Figure 5.6c shows, there is a positive impact of this staff retention, with adoption rates of 19% in cases when the original staff left, versus 33% when they were retained, a difference barely short of statistical significance (p=0.12).

5.4.5 Adoption: Experimental design

Turning to the experimental design, we examine first whether policymakers have a preference for particular behavioral mechanisms. We distinguish between simplification, which seems uncontroversial, versus social comparisons or personal motivation which can be seen as more aggressive interventions. Figure 5.7a shows that forecasters on average expect trials with simplification to be more often adopted than trials using other behavioral mechanisms. Indeed, the adoption rate is 33% of trials adopted for simplification versus 19% for personal motivation and 24% for social cues (though the differences are not statistically significant at conventional levels).

Next, we turn to a second aspect of the experimental design, whether the communication in the trial was pre-existing. To clarify, suppose that in a trial, BIT-NA and the city sent reminder letters for timely utility bill payment. We label such letters *new communication* if the city had not been sending such letters before the trial. We label them as *pre-existing communication* if the city had been sending the letters before the trial, and the trial incorporated new nudge features in the treatment arms, compared to the status-quo control communication. As Figure 5.7b shows, in the 21 trials in which there was a pre-existing communication and the city tested variations using nudges, the adoption is 67% (14 out of

[61] Most trials have only one (42% of trials) or two (34%) city staff members listed on the trial protocol. In two trials, the staff member was still working for the city, but in a different department. We do not count these two trials as cases of staff retention, but including them does not change the results.

[62] The persistence of key staff may be endogenous to the RCT results, or to organizational features, though we do not detect differences by staff retention (Table 5.2, Columns 4 and 5).

21). Conversely, in the 52 trials in which the communication was new, the adoption rate is only 12% (6 out of 52).[63]

This 55 pp. difference, which is highly statistically significant ($p<0.01$), is five times larger than the expectation of forecasters who predict only an 11 pp. difference on average. Government workers, who may have more experience with such matters, are more accurate than nudge unit staff or researchers, but their average predicted difference of 22 pp. is still less than half the actual impact (Table D.4).

To appreciate how predictive this one variable is, we revisit Figure 5.2, which reports all the nudge treatment effects and also labels whether the nudges were adopted and whether the communication was pre-existing (diamond) versus new (circle). The large majority of adoptions are for pre-existing communication. Conversely, almost no new communication is adopted, including two treatment effects of over 20 pp.

5.4.6 Adoption: Multivariate evidence

So far, we have considered each determinant on its own, but there could be a correlation between the different factors. What if, for example, the impact of pre-existing communication is partly due to different effect sizes, or different city features?

In Table 5.3 we present the estimates from a linear probability model predicting adoption, considering first only evidence-based determinants (Column 1), only organizational features (Column 2), then only experimental design features (Column 3), and finally all three conditions together (Column 4). There is essentially no predictive power from the measures of strength of evidence (Column 1) and only some impact from city staff retention (0.13 pp., s.e.=0.08) and the other city features (Column 2). Focusing on the experimental design (Column 3) we detect a modest impact of simplification, compared to personal motivation and social cues (both of which are compared to other mechanisms) and most importantly a very large and statistically significant impact (t=4) of pre-existing communication, 0.53 pp. (s.e.=0.13). The high predictive power of this factor yields a 0.34 R-squared, compared to 0.01 in Column 1 or 0.03 in Column 2.

Considering all the factors together (Column 4), the standard errors for the various estimates do not generally increase and in fact decrease in some cases. The key determinant remains the pre-existence of communication, unaltered at 0.52 pp. (s.e.=0.13), while none of the other determinants is statistically significant.

We then add city fixed effects (Column 5), controlling for any city-level features and identifying adoptions only comparing across different trials within a city.[64] This extra set of controls does not meaningfully alter the results.

In Column 6 we include the most comprehensive set of controls: (i) fixed effects for the policy areas (e.g., revenue collection versus environment), proxying for different outcomes

[63]The *new communication* category includes both cases in which the nudge treatment arm is compared to a control arm which also receives a (new) communication, and cases in which the nudge arm is compared to a no-communication group. As Figure D.5a shows, the adoption rate is very low in both groups and thus we pool them. There are also 6 trials in which a new insert was sent in addition to a pre-existing mailer. We discuss these cases in Appendix D.7.

[64]In the sample, 11 cities have only one trial each, and 19 cities have at least two trials. The coefficient on pre-existing communication is identified by 10 cities with at least one trial with pre-existing communication and one without, covering 36 trials.

and city departments, (ii) the level of take-up in the control group of the targeted policy outcome, which could proxy for how malleable the outcome is (e.g., a control-group take-up of 1% indicates a rare behavior that may be hard to affect), (iii) an indicator for online (as opposed to in-print) communication, (iv) the number of years since the trial was conducted, to control earlier versus later trials (e.g., from institutional learning in BIT-NA) or the decay of adoption over time, (v) whether the partnering city department is directly responsible for implementing the nudge, and (vi) the seniority of the city staff partner. Some of these controls are motivated by evidence (Table 5.2) that the trials with new communication differ, for instance, in certain policy areas.

Adding all these controls raises the R-squared up to 0.79 while leaving the impact of pre-existing communication at 0.60 (s.e.=0.15). The additional controls shift somewhat the impact of the treatment effect size (0.23, s.e.=0.13).

For another sense of the magnitudes, Figure D.6 computes the area under the curve (AUC) that measures the accuracy of prediction. Using just the evidence-based determinants (Column 1) yields an AUC of 0.58, and using all the determinants in Column 4 except the indicator for pre-existence yields an AUC of 0.72. In comparison, using just one variable, whether the communication was pre-existing, yields a higher AUC of 0.78.

In Column 7 we estimate the same specifications using a logit model, leading to parallel results. Pre-existing communication is estimated to have an impact on adoption of 291 log points (s.e.=67), that is an increase of over 1,000 percent over the baseline.

Model estimate In Column 8, we present estimates for the model in Section 5.3, including the controls from Column 4. The prior μ_0 is slightly positive at 0.40 (s.e.=1.09), with a fairly narrow standard deviation $\sigma_0 = 0.23$ (s.e.=0.08); as an implication, the model implies only a modest weight on the signal, that is the treatment effect, estimated at 0.13 for the median and 0.03 for the average RCT. This reproduces the flat responsiveness in adoption to the effectiveness, as shown in Figure 5.5b. The model also reproduces the finding that pre-existing communication is the largest predictor.[65]

Robustness We consider a series of robustness checks in Table D.6: (i) using robust standard errors (as opposed to clustering by city); (ii) dropping four observations in which the current communication suggests adoption but is not as straightforward as in the other cases (detailed in Appendix D.4); and (iii) considering only the cases of adoption in which we received and verified the current template of the communication and dropping cases in which the city just stated their adoption (though we did confirm with follow-up questions). Across these specifications, we replicate the results.

5.4.7 Other forms of adoption

So far, we considered the adoption of the nudge treatment by the city department. However, there are other dimensions of adoption, such as an RCT inspiring the city to use treatment wording for different purposes or to collect more experimental evidence. We recorded such mentions of broader adoption in our communications with the city department, as detailed in Section 5.2.3, but we should caution that this analysis is exploratory, since we rely necessarily on self-reports of this form of adoption.

[65] This is the interior solution. Since the effect size has little predictive power, the corner solution with $\hat{\sigma}_0 = 0, \hat{\mu}_0 = -2.6$ (moving toward the logit estimates in Column 7) has a superior log likelihood.

The broad adoption of evidence (Column 2 of Table 5.4) is more correlated with effect size and is not positively predicted by pre-existing communication, compared to the adoption of the specific nudge in the trial by a city department (Column 1).

5.5 Interpretation and implications

5.5.1 Interpretations

The most important determinant of adoption is whether the communication is pre-existing, while other determinants play more limited roles. We discuss four potential explanations of this finding and related evidence in support or against each explanation.

Cost of materials While pre-existing communications already have a line in the budget, the new communications may not have such secured funding in the following years. In Figure 5.8a we compare the impact of pre-existing communication for online communications, which have near zero marginal cost, and for paper communications, which require financing the mailer. We find a nearly identical effect size, suggesting that the cost of the materials is not the primary reason for the key finding.

State capacity City departments with pre-existing communications may have better *state capacity*, which could explain why they were already sending communications and why they implement more nudge innovations. City-level variation in state capacity cannot explain the results, given that the estimates are unaffected by controlling for city fixed-effects (Column 5 of Table 5.3). Still, there may be within-city variation in staff and decision-making capacity across departments.

As a proxy for each department's capacity for communication, we measure the availability of online forms and communications on city department websites, such as business license forms and code enforcement brochures. Conditional on availability, we also measure the rate at which the forms and communications are updated over time, a proxy of willingness to "experiment". In Figure D.2, we compare such variables in the BIT-NA city departments to the same departments in the non-BIT cities closest in population size within the same census region. We find no difference, economically or statistically, in either variable. Thus, BIT-NA departments with pre-existing communication do not appear to have special state capacity, compared to similar departments. We do find some evidence that the departments with pre-existing communication are more likely to post forms (though no more likely to change them), in both the BIT-NA cities and the matching ones. This suggests that some types of departments tend to have more frequent communications, the explanation for which we leave to future research.

Other unobservables Other unobservable variables, such as prior beliefs of the policymakers, may be correlated with pre-existing communication in a way that explains the results. While prior beliefs likely explain the adoption of some nudge treatments with negative effect sizes—e.g., the wording is clearer than the control wording—it seems implausible that they would explain the impact of pre-existing communications. For the new communications, city staff priors likely were *more* positive to enable an experiment, given the higher complexity relative to experiments set up on pre-existing communication. Further, controlling for additional features in Columns 5 and 6 of Table 5.3 slightly *increases* the estimated impact of pre-existing communication. Under the assumptions of Altonji et al. (2005)—that the

unobservables are positively related to the observables—this makes it less likely that unobservables are driving the key finding.

Organizational inertia Consider a two-step decision process with organizational costs to adoption, $C = C_1 + C_2$ in the model, where the first step is whether to send any communication, and the second step is designing the content of the communication. Setting up a new communication can have substantial costs C_1, while content changes conditional on communication have a low cost C_2.[66]

In cases with pre-existing communication, there is infrastructure and staffing to send the communication each year, so the first step is not a hurdle ($C_1 \approx 0$), and incorporating the most effective wording in the communication is relatively straightforward, leading to high adoption. When the communication was instead set up for the experiment, there is no routine pathway to send it again (C_1 is high), leading to low adoption.[67]

A first prediction of this model is that the low adoption in the *new communication* trials should be due to inertia in the first step, not in the choice of content. Indeed, Figure 5.8b shows that for the RCTs with new communication, the non-adoption is *entirely* due to nothing being sent after the RCT.

A second prediction is that, when communications are sent post-RCT, the content should be more responsive to evidence. Indeed, in cases of nudge adoption for RCTs with multiple nudge arms, in 5 out of 6 cases the nudge with the largest effect size was adopted (Figure 5.5d). Further, for the trials with pre-existing communication, for which the first-step hurdle is minimal, the adoption of evidence rises from 45% for non-statistically significant results to 90% for statistically significant results (Figure 5.9a), though the evidence is more muted for the response to effect size (Figure 5.9b). Overall, the organizational hurdles in the first step appear much more significant than in the second step of content formation. This suggests that legal, communications, or political preferences over content are not the main barrier to evidence adoption in this context.[68]

5.5.2 Survey evidence on adoption

Organizational inertia is an umbrella term nesting distinct explanations for non-adoption, each implying different potential interventions. For example, would it be enough to remind cities to adopt the results for new communications, or would additional staff be necessary? Is low prioritization of the communication an issue?

To provide additional evidence, we ran a short survey of city officials in two samples. First, we contacted cities that conducted the 31 trials that did not result in adoption of the nudge communication despite a positive effect size (≥ 1 pp. or $t > 1.96$) and obtained

[66] A third of forecasters mention factors related to inertia in the open-ended responses (Figure 5.4). Even these forecasters do not anticipate the channel through which inertia operates, as on average they expect the same impact of pre-existing communication as those who do not mention inertia.

[67] Inertia also explains the different findings for broad adoption, since whether the specific communication in the trial was pre-existing has no bearing on the inertial barriers for adoption in other contexts.

[68] Figure D.5b partitions trials into thirds by effect size; the findings are similar. Figure D.7a-f provides interaction effects for staff retention and by a median split in the control take-up, which may proxy for the difficulty of affecting an outcome variable. Pre-existing communication remains the only reliable predictor of adoption, statistically and economically, across these splits.

responses for 25 trials, for an 81% response rate.[69] Second, we contacted city policymakers in other government innovation networks, yielding 45 additional responses. The survey asks on a 1 (not at all) to 5 (extremely) Likert-scale how helpful each of seven channels (presented in random order) would be for adopting a successful nudge in ongoing city communication: (1) prioritization from key decision-makers, (2) timely reminders, (3) logistics and technical support, (4) more staff full-time equivalent (FTE) hours, (5) city staff receive training from external consultants, (6) funding for the costs of communication, and (7) stronger evidence of effectiveness. These channels are similar to those in other surveys of policymakers (e.g., Figure D.1 in Toma and Bell, 2022). We also asked for open-ended feedback on supporting evidence adoption in cities.

Figure 5.10 shows the average ratings across the two samples. Prioritization from decision-makers is indicated as the key factor, followed by human capital solutions such as outsourcing via logistical support, staff training, and additional staffing, as well as funding for communication costs. Demand for stronger evidence and the provision of reminders are rated as less important. Figure D.8a reports the average response in each sample, further splitting the second sample into city workers who self-report that their city has made policy adoption choices based on evidence, versus not. The patterns are similar across the three samples, with prioritization rated as the top factor throughout.

While these responses should be taken with the necessary caveats, we identify some takeaways: (i) stronger evidence is not seen as a priority, indicating that the bottlenecks are likely downstream of evidence collection; (ii) a light-touch intervention, a reminder, is not seen as sufficient; and (iii) to overcome the *organizational inertia* of defaulting to the status quo, respondents claim that decision-makers should prioritize the adoption of evidence by assigning personnel and training resources to this purpose. The open-text responses often touch on this last point, as indicated also by the word cloud in Figure D.8b. One respondent explains: "*Our evaluation work has been an "extra" on top of employees doing their regular jobs, so even if the employee sees value in it, if it's not part of what their manager expects them to do, it falls off their priority list. The only place I've seen evaluation done routinely, and findings applied, is in a team where the manager sees value in evaluation and prioritizes it for their team. They've encouraged their staff to take evaluation trainings and included evaluation in project plans.*"

5.5.3 Implications and counterfactuals

How much did the evidence collected from the RCT improve the targeted policy outcome, and how much could it have improved it under other counterfactuals?

We assume that the treatment effects of the RCTs would replicate in subsequent years if the same treatments were adopted, and when no nudge treatment is adopted, we assume an improvement of 0 pp. That is, for each trial i, we take the highest effect size $\hat{\beta}_i$ across treatment arms and compute the average actual "improvement" as $\frac{1}{73}\sum_{i=1}^{73}\hat{\beta}_i\mathbf{1}\{i \text{ is adopted}\}$. The first bar of Figure 5.11 shows that across all 73 trials, the evidence from the RCTs is

[69] As Table D.7 documents, the 6 trials for which we could not obtain a response do not differ on key dimensions. Given that the large majority of trials with non-adoption are for the *new communication* case (22) versus *pre-existing communication* (3), we are not powered to study the difference between the two types of trials, and report the results for the pooled sample.

predicted to have improved policy outcomes by 0.89 pp. based on actual adoptions, a statistically significant improvement.

The second bar presents a counterfactual of how much the RCTs would have improved outcomes, had all the treatments with positive effect size been adopted: 2.70 pp. This comparison highlights the importance of bottlenecks to policy adoption: the achieved gains from the RCTs of 0.89 pp. are only one third of the achievable gains of 2.70 pp.

For the 52 trials with new communication, in comparison to the achievable 2.48 pp. under optimal adoption, the actual adoption creates an improvement of only 0.32 pp., less than one tenth of the possible surplus. Conversely, for the 21 trials with pre-existing communication, the estimated policy improvements from actual adoptions is 2.31 pp., closer to the optimal counterfactual of 3.24 pp. Thus, for the cases in which organizational inertia is more conducive to adoption, the evidence collected in the RCTs largely translated into actual significant policy improvements.

A third benchmark is the effect size implied by the forecasts. Forecasters predict the average adoption rate to be 13% for trials with effect sizes in the lowest third, and 49% for trials in the highest third. An average with these weights implies a predicted improvement of 1.23 pp. Thus, the forecasters are slightly optimistic.

5.6 Generalizability of results

How applicable are the lessons from this study? The adoption rate of 27% in our study is in the range of the (few) estimates in the literature (Table 5.1).

A separate question is whether organizational inertia also impacts the adoption of evidence in other settings through the pre-existing channel. In line with our results, Kremer et al. (2019) find that USAID-funded interventions that were distributed through pre-existing platforms were three times more likely to be adopted widely than those establishing new distribution networks (see Table 5.20 in their paper). They note, however, that the pre-existing channel in their context may be confounded with lower costs.

The experiment in Hjort et al. (2021) provides further evidence. Brazilian mayors attending a conference who were randomized into a treatment group were invited to a session on taxpayer reminder letters. The session presented evidence on the cost-effectiveness of a nudge intervention and provided a template (Figure D.9) with three mechanisms: (1) a deadline, (2) the risk of fines and audits, and (3) social norms.

Between 15 and 24 months after the conference, the researchers contacted the municipalities to ask whether the city sends any communication for taxpayer reminders. If so, they asked whether the communication is a letter (as opposed to an email, for example) and whether it includes each of the three behavioral mechanisms. While the researchers did not ask cities whether the communication was pre-existing prior to the conference, they did contact municipalities in both the treatment group and the control group.

Re-analyzing the data from Hjort et al. (2021), in Figure 5.12 we compare the treatment and control share of observations in a 2×2 matrix for (i) whether the city is sending a reminder *letter* (L) and (ii) whether the communication has all three *nudge* (N) mechanisms. A first benchmark model, aiming to mirror the specification in Hjort et al. (2021), posits that the intervention effect is monotonic—that is, the info session moves cities only toward, not away from, adopting either the letter or the nudge as indicated by the arrows, with a

uniform persuasion rate f. This yields a system of three equations (given that the fourth cell is a linear combination of the others):

$$P^T_{L=0,N=0} = P^C_{L=0,N=0}(1-3f)$$
$$P^T_{L=1,N=0} = P^C_{L=1,N=0}(1-f) + P^C_{L=0,N=0}f$$
$$P^T_{L=0,N=1} = P^C_{L=0,N=1}(1-f) + P^C_{L=0,N=0}f$$

where $P^g_{L,N}$ is the rate in group $g \in \{T, C\}$ for treatment and control.

Column 1 of Table 5.5 shows the results from a minimum-distance estimation of this baseline model, accounting for the first-stage session attendance of 37%. The baseline persuasion rate is positive and statistically significant at 0.035 (s.e.=0.017).

We then enrich this baseline model to allow for a different persuasion rate f_{pe} for pre-existing communication: the persuasive impact may be larger for cities that were already sending a letter (see Figure 5.12). Column 2 shows that the estimated persuasion rate for the pre-existing cases is indeed higher at 0.42 (s.e.=0.21) by an order of magnitude, if fairly imprecise. In Panels B and C, we re-estimate the results for alternative definitions of the nudge adoption, yielding similar qualitative patterns.[70]

An important caveat is that alternative models are possible, for example allowing a separate persuasion rate along the diagonal, f_{diag}, which also fits well (Column 3). In a horse-race between the two models (Column 4), which persuasion rate plays a larger role depends on the definition of nudge (Panel A versus B and C). Ultimately, while we cannot conclusively prove a larger adoption impact for pre-existing communication in Hjort et al. (2021), this strikes us as a reasonable interpretation of the data.

The Hjort et al. (2021) data set also allows us to further investigate whether the pre-existing effect is confounded with the selection of cities. The data include a rich set of characteristics of the mayor (e.g., education, vote margin, term effects, and ideology) and the city (e.g., population, college educated, poverty, inequality, income, and tax revenue). In the control group, cities that are, or are not, sending a letter are not significantly different in these observables (Table D.8b), which alleviates selection concerns.

5.7 Discussion and conclusion

Organizations from the World Bank to U.S. federal agencies run experiments to gather evidence on how to best achieve outcomes of public policy interest. In our context, U.S. cities experimented by testing behavioral science interventions in their communications with citizens to achieve policy goals such as the timely payment of municipal taxes. But does the gathering of evidence guarantee the improvement of the outcomes, or are there bottlenecks to the adoption of evidence, even under such favorable conditions?

At least in our context, the bottlenecks are substantial: the innovations from the RCTs yield only about one third of their potential direct benefits.[71] This is because the rate

[70] See Table D.8a for the treatment and control group moments under these alternate definitions for nudge adoption. Hjort et al. (2021) define policy adoption as sending any taxpayer reminder communication (not just letters) with or without any of the three nudge mechanisms.

[71] We acknowledge that there are further benefits to policy RCTs not captured in our estimates. For example, policy leaders note that they often look to RCTs in peer cities for innovations.

of adoption is fairly low, 27%, and is only modestly sensitive to the effectiveness of the intervention. As a consequence, several high-return nudge innovations are not adopted by the city in years subsequent to the experiment. Even organizations that value and produce rigorous evidence are not immune to challenges in evidence adoption.

To an extent this is bad news for evidence-based policy-making. But there is good news too: the barriers to adoption, in our context, do not appear to be due to intractable problems such as political divisions or funding challenges for the roll-out, but more "simply" due to organizational inertia. When the RCTs take place in the context of ongoing communication to residents—such as altering a yearly mailer about registering business taxes—the adoption rate is high at 67% and, to an extent, more sensitive to evidence. For such ongoing communications there is a routine process, and organizations incorporate the successful changes. For the new communications which were not pre-existing, instead, the adoption rate is very low, at 12%. Following the experiment, inertia tilts the organization back to the previous status quo of non-communication.

A first implication of these findings is that targeting such bottlenecks should achieve a higher adoption rate post-RCT. Nudge units already frame experimentation as an opportunity to test "what works" for the purposes of scaling. Given that adoption still does not arise organically and that leadership prioritization after the RCT is not guaranteed (as our survey suggests), heavier investments could be made to support the adoption after a trial, in the same way that heavy investments have been made in the past decade to increase the implementation of RCTs in government. Moreover, government agencies, in their initial choice of interventions to test, could consider whether the infrastructure and sustained agency support exists to scale up a particular treatment.

A second implication is that we should collect systematic evidence on such bottlenecks and overall adoption, and keep track of relevant variables, such as the pre-existence of communication. A natural consequence of having sparse evidence on adoption is that experts and practitioners alike understand that barriers exist but are less able to predict the relative importance of the barriers. Figure D.10a plots the average predictions of the bottlenecks against the actual impact on adoption. The forecasters are mostly directionally correct, but they are unable to discern the most important factor, to the point that the predictions are negatively correlated to the actual determinants. Interestingly, this pattern is near identical for both researchers and practitioners.

An important caveat is that the findings are, to an extent, specific to our context. To have some sense on perceived bottlenecks in other contexts, we asked respondents of the forecasting survey to compare our context to A-B experiments in firms and to RCTs in low-income countries. The respondents thought on average that evidence-based adoption would be higher in firms, but that the development RCTs would be similar in terms of adoption (Figure D.10b). Indeed, the impact of pre-existing communication appears to play a role in adoption also in Kremer et al. (2019) and Hjort et al. (2021).

Regarding A-B experimentation in firms, we know of no comprehensive data set on adoption, beyond specific instances (e.g., Cho and Rust, 2010; List, 2022). Profit motives make it less likely that researchers will be able to access comprehensive records for a set of A-B experiments, compared to the transparency with which BIT-NA shared their records. Lacking such evidence, we conjecture that bottlenecks are likely to be an issue even in firms with online platforms for experimentation, given that the adoption post A-B testing requires

an active decision. Only platforms that automatically adopt the most successful arm, used in some companies, would remove the inertial barriers.

Finally, in other settings, the political barriers to adoption may be higher, or the costs of rolling out an innovation at scale often will be larger than the cost of sending a mailer. Given that those bottlenecks may be harder to address, it is even more important to put systems in place to address the organizational inertia. Good architecture design should apply to experimentation as well.

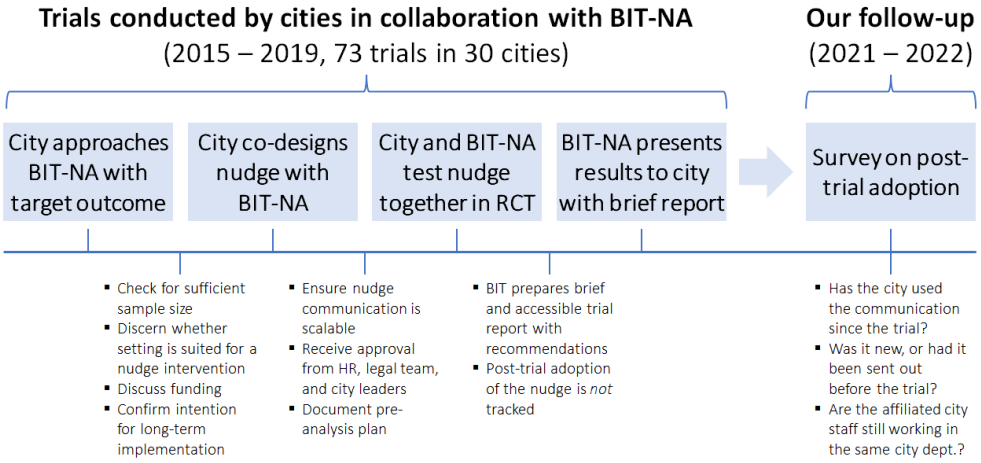
Data Availability

The replication data and code are available at DellaVigna, Kim, and Linos (2023) (`https://doi.org/10.7910/DVN/XOCJOF`).

5.8 Figures and tables

Figure 5.1: Study design and sample restrictions

(a) Study design



(b) Sample restrictions

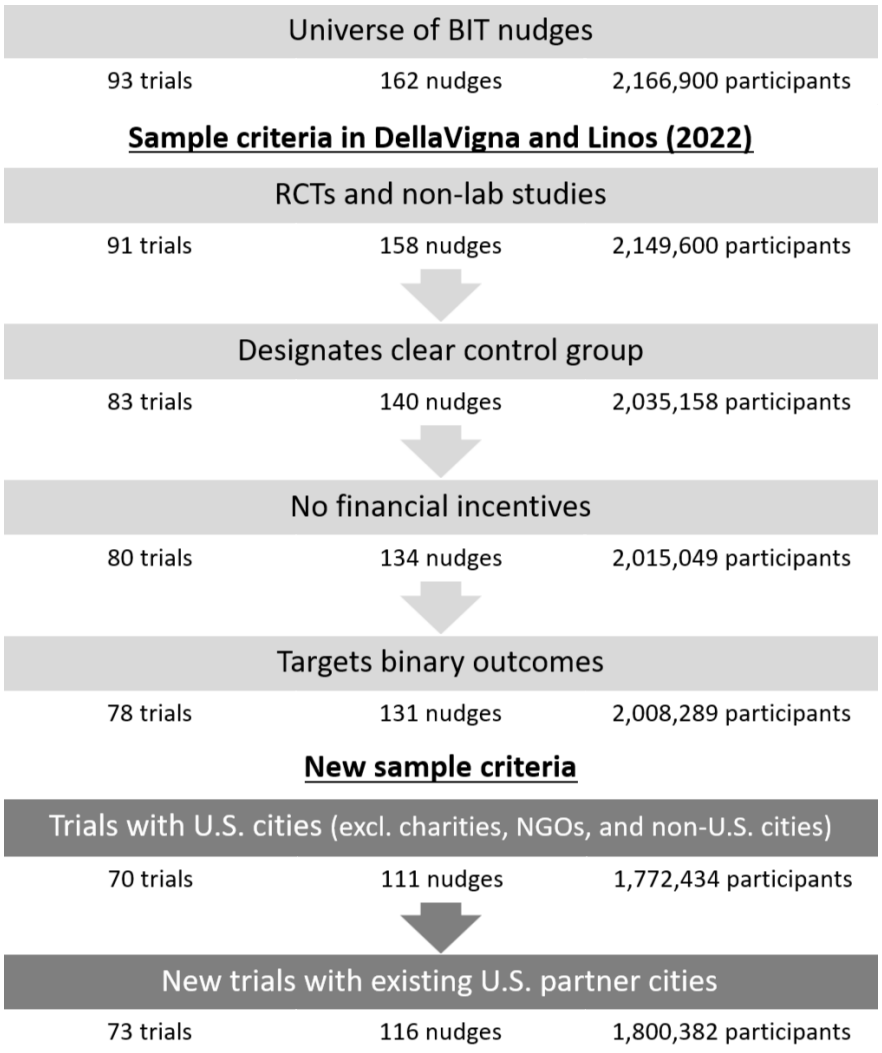
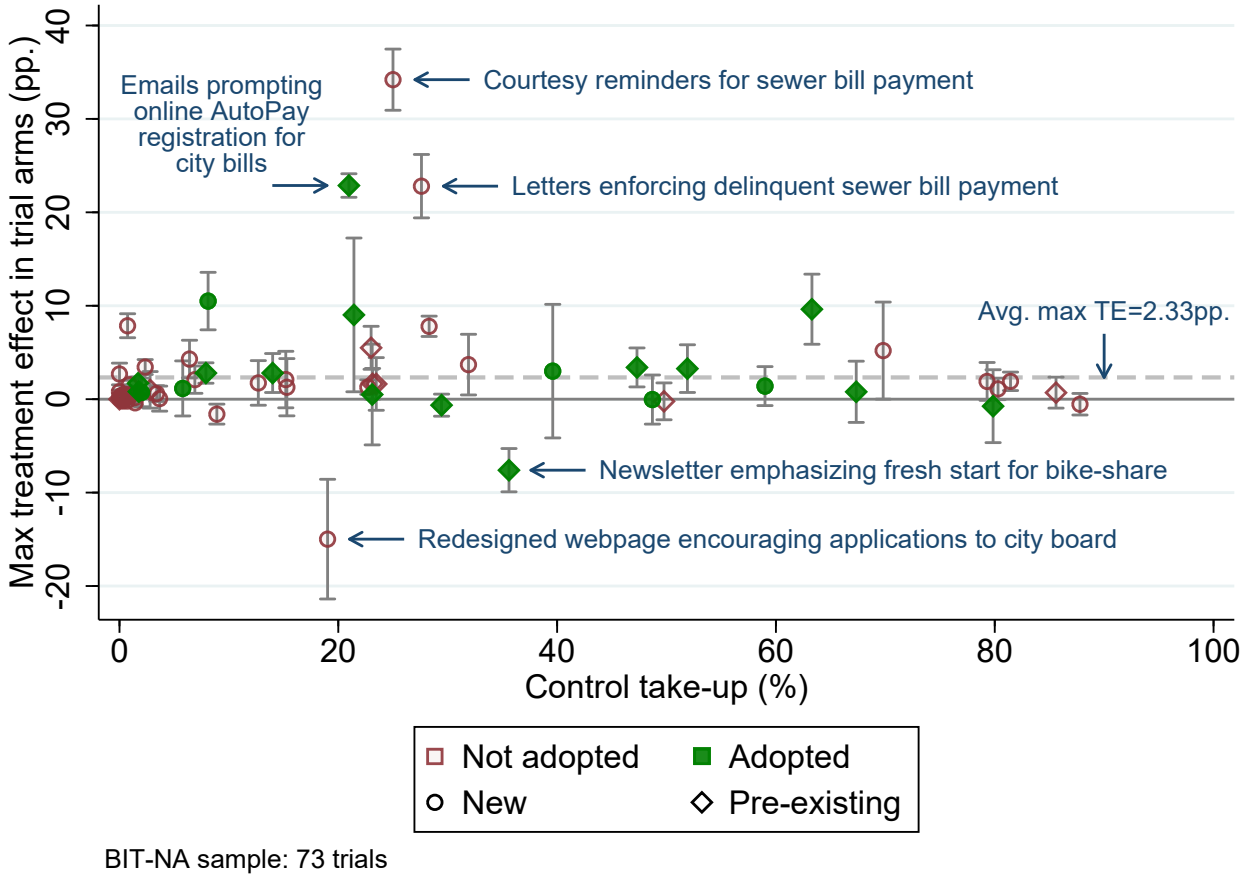
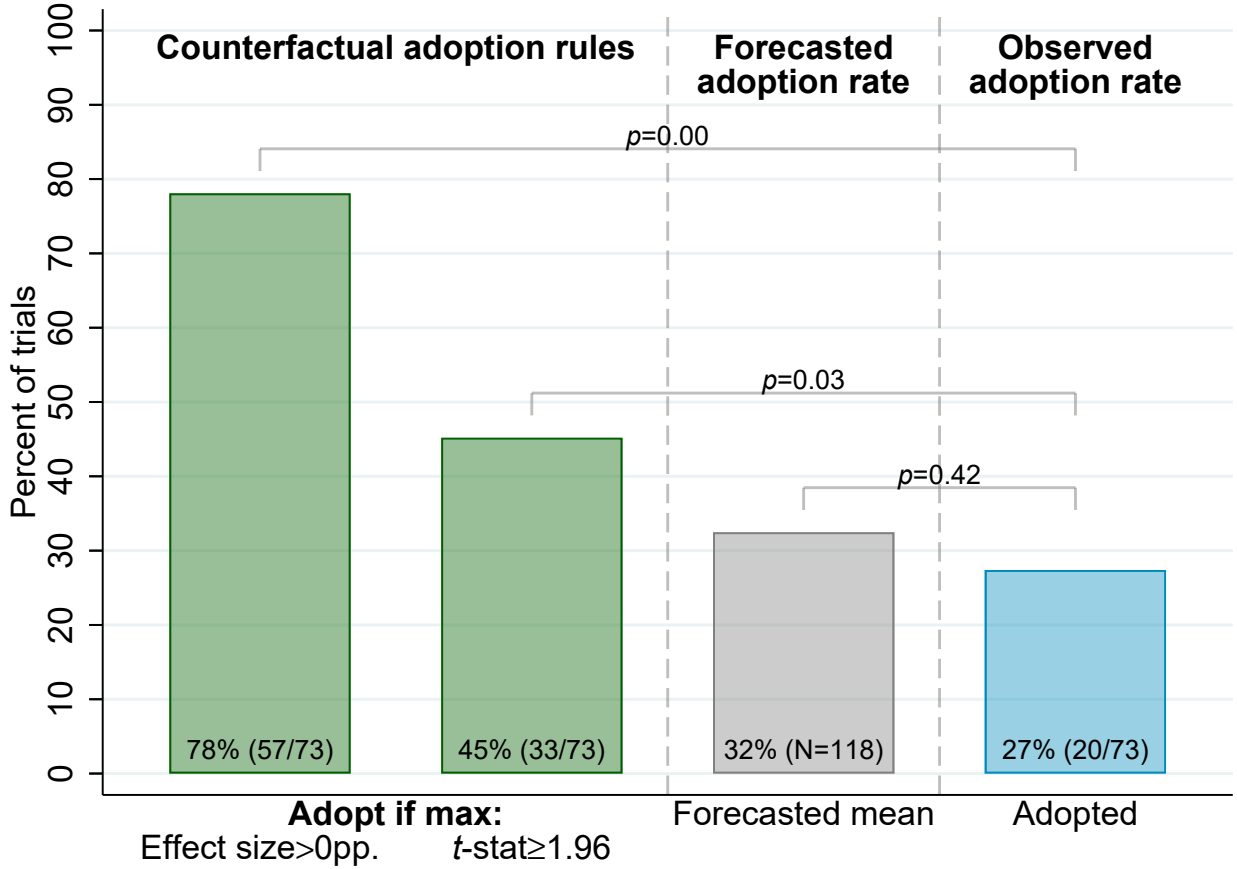


Figure 5.2: Trial-by-trial adoption and effect sizes



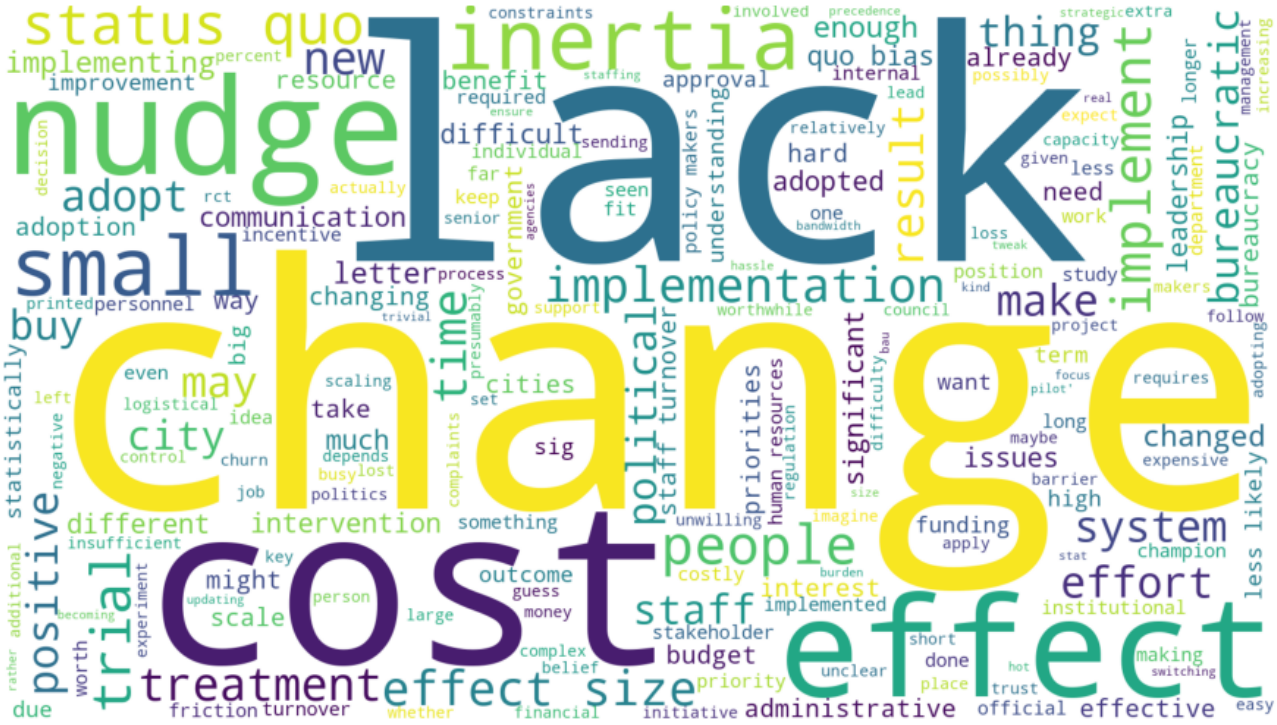
This figure plots the trial-by-trial treatment effect and control take-up. For trials with multiple treatment arms, the figure shows the effect of the arm with the highest effect size.

Figure 5.3: Adoption of nudges: Observed compared to benchmarks



This figure compares the observed adoption rate in the sample with two counterfactual adoption rules and with the overall adoption rate forecasted by experts. The first counterfactual rule is to adopt all trials that found a positive effect size, and the second is to adopt all trials that found a positive *and* statistically significant effect size.

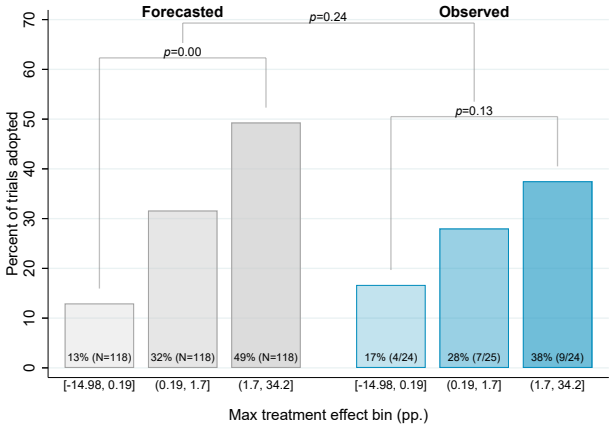
Figure 5.4: Word cloud from open-ended forecasts of adoption determinants



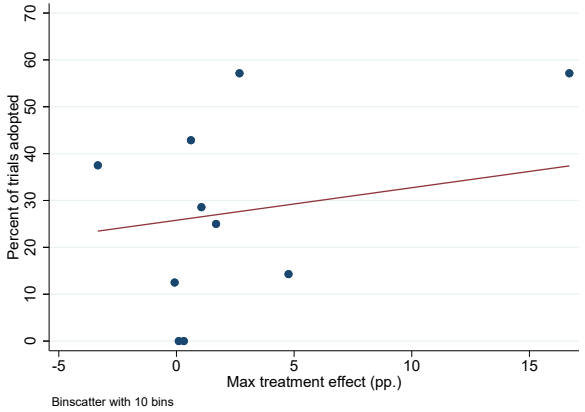
This word cloud is based on the responses in the forecasting survey to the open-ended question "When cities do not adopt the nudges from the trials, what do you think are the main reasons?" The size of the words is proportional to their frequency in the responses.

Figure 5.5: Adoption of nudges by effectiveness

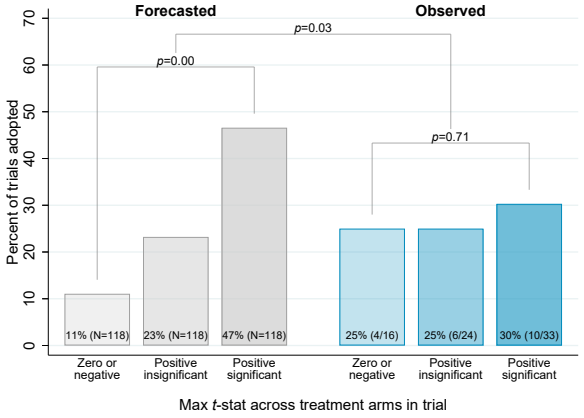
(a) By effect size



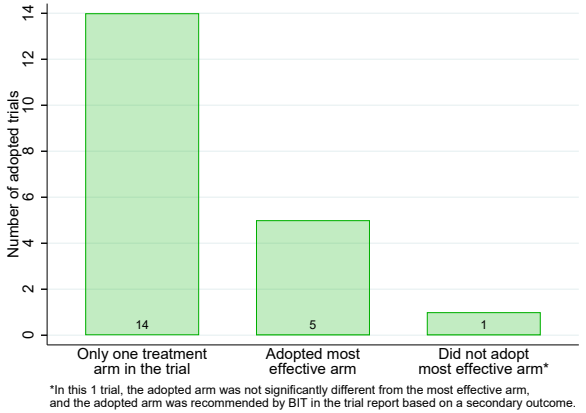
(b) By effect size (bin scatter)



(c) By statistical significance



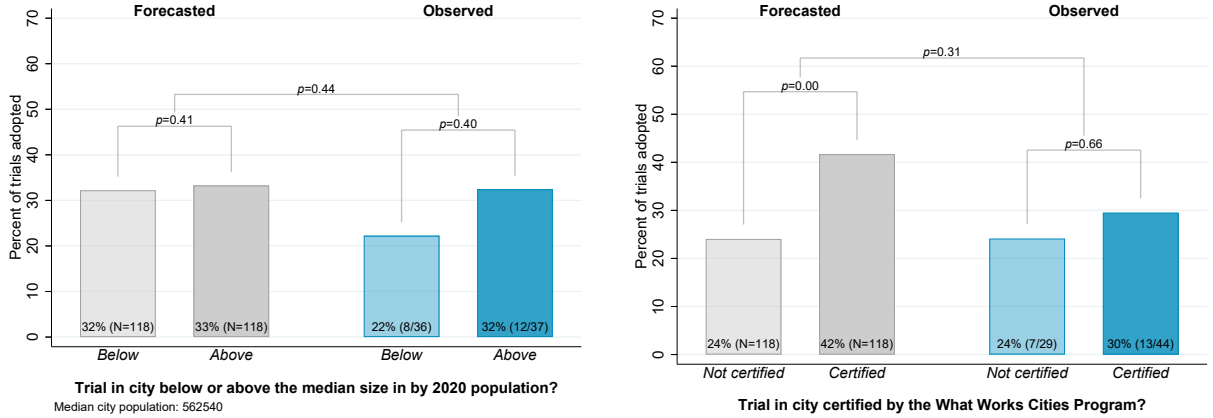
(d) By treatment arm among adoptions



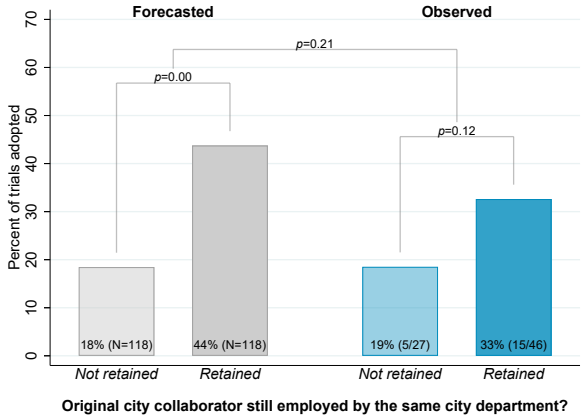
Figures 5.5a and 5.5c show the forecasted (left bars) and actual (right bars) adoption rates of trials conditional on two measures of effectiveness: (a) effect size in percentage points and (b) statistical significance at the 95% level. In Figure 5.5a, trials are partitioned into thirds by their effect sizes. In Figure 5.5c, trials are categorized based on whether they found a zero or negative effect, a positive but insignificant effect, or a positive and significant effect. Figure 5.5b is a bin scatter of the actual adoption rate of trials across 10 bins for the treatment effect size. Figure 5.5d categorizes the actual adoption of trials into cases when the city adopted: the only treatment arm in the trial, the most effective arm if there were multiple, or did not adopt the most effective arm.

Figure 5.6: Adoption based on city context

(a) By "state capacity" (proxied by city size) **(b)** By certification of evidence-based cities



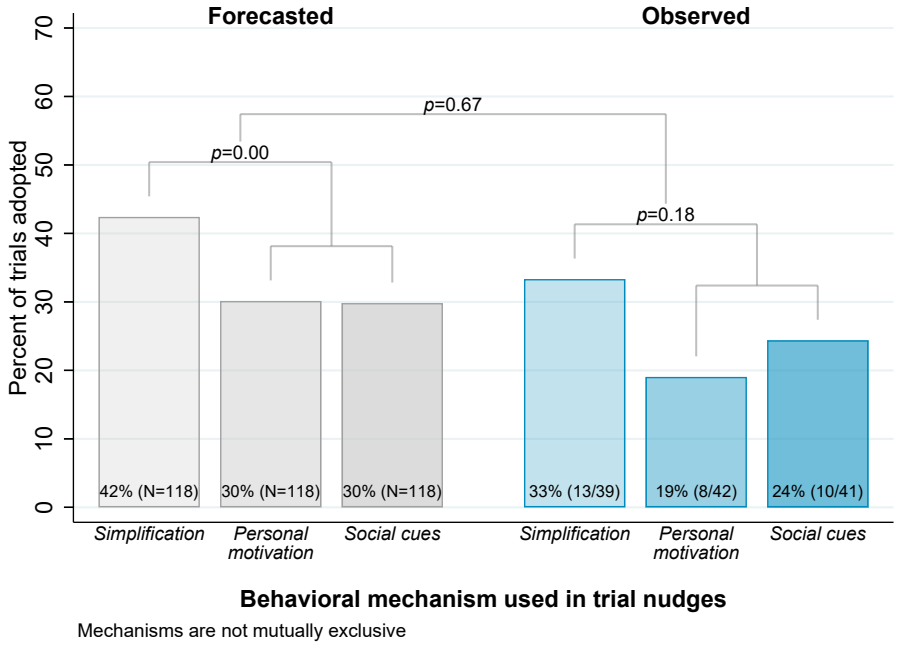
(c) By staff retention



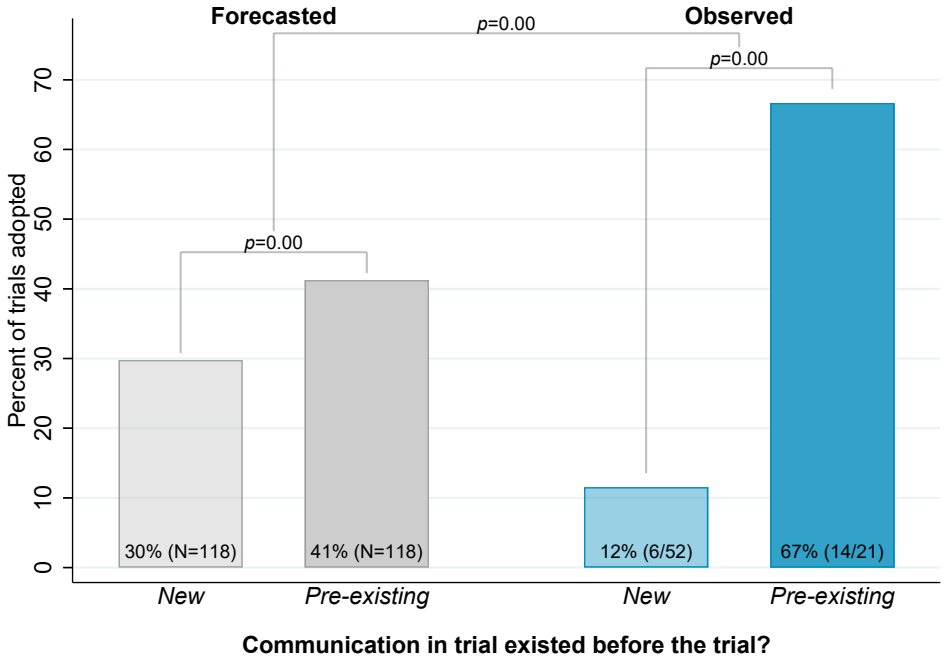
Figures 5.6a-5.6c show the forecasted (left bars) and actual (right bars) adoption rates of trials conditional on whether the collaborating city: (a) is below or above the median 2020 city population in the sample, (b) has been certified by What Works Cities as a "data-driven, well-managed local government", and (c) has retained the original city collaborator on the trial in the same city department.

Figure 5.7: Adoption based on experimental design

(a) By behavioral mechanism



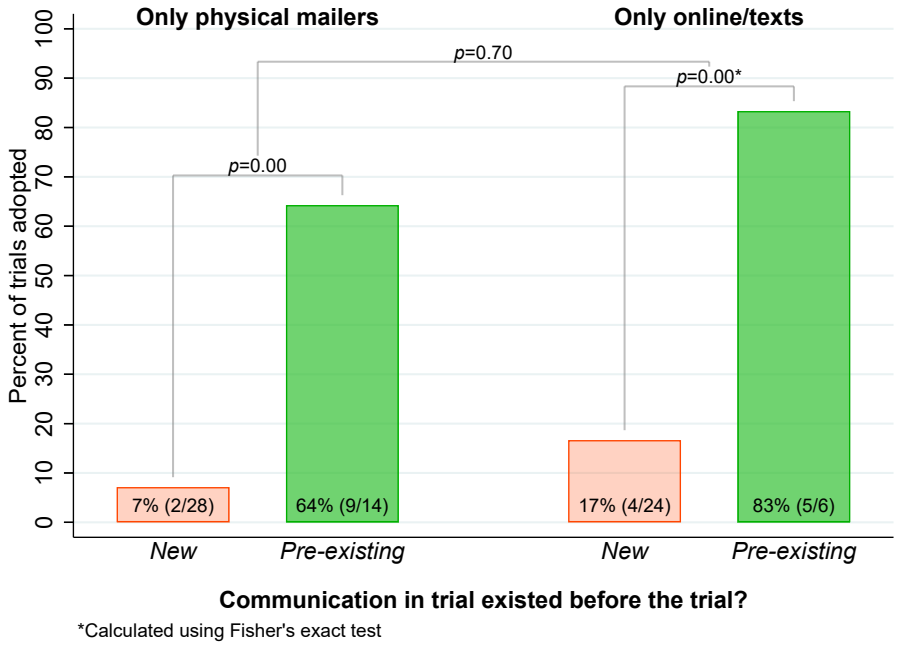
(b) By pre-existence



Figures 5.7a and 5.7b show the forecasted (left bars) and actual (right bars) adoption rates of trials conditional on whether the trial: (a) uses simplification, personal motivation, or social cues in the nudge intervention, and (b) tests a nudge in a new communication that the city had not sent prior to the trial or in a pre-existing communication that the city had already been sending.

Figure 5.8: Mechanisms behind the effect of pre-existence

(a) Marginal cost of communication



(b) Any communication adopted post-trial

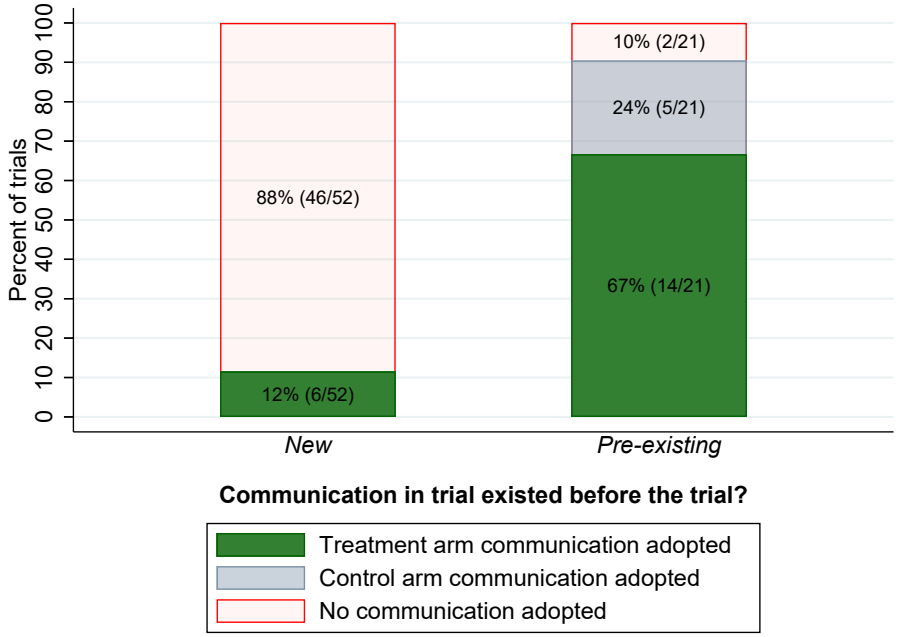
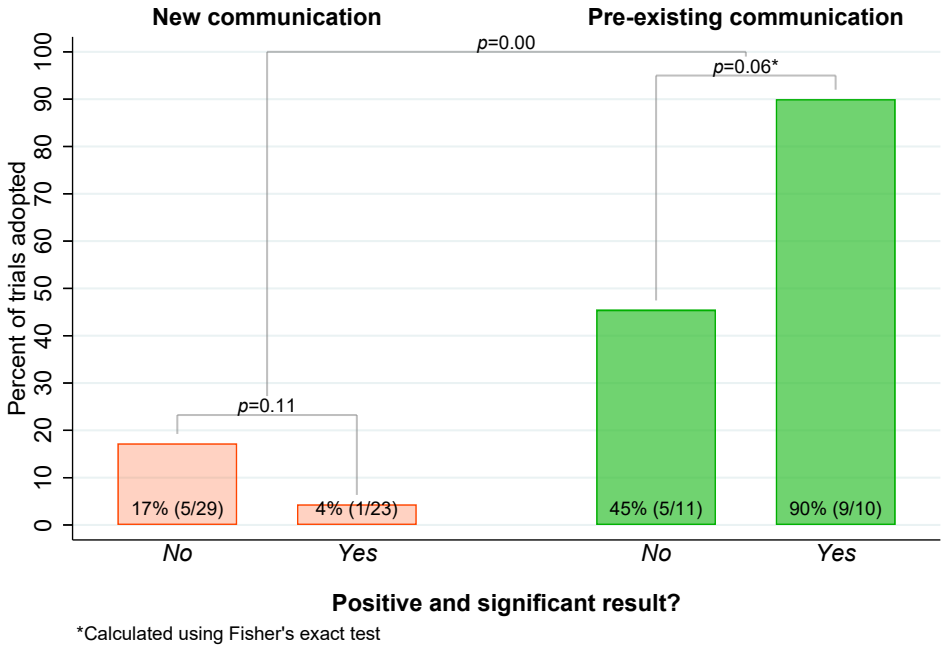


Figure 5.8a compares the adoption rate of interventions in new versus pre-existing communications separately for those delivered by a physical medium (e.g., letter or postcard) and those by a digital or online medium (e.g., email or text).
Figure 5.8b shows the rates of adoptions of the treatment arm communication as well as the control arm communication for new and pre-existing trials separately. For pre-existing trials, the control arm is typically the status-quo communication that the city was sending prior to the trial.

Figure 5.9: Pre-existence and evidence based adoption

(a) Pre-existence and statistical significance



(b) Pre-existence and effect size (bin scatter)

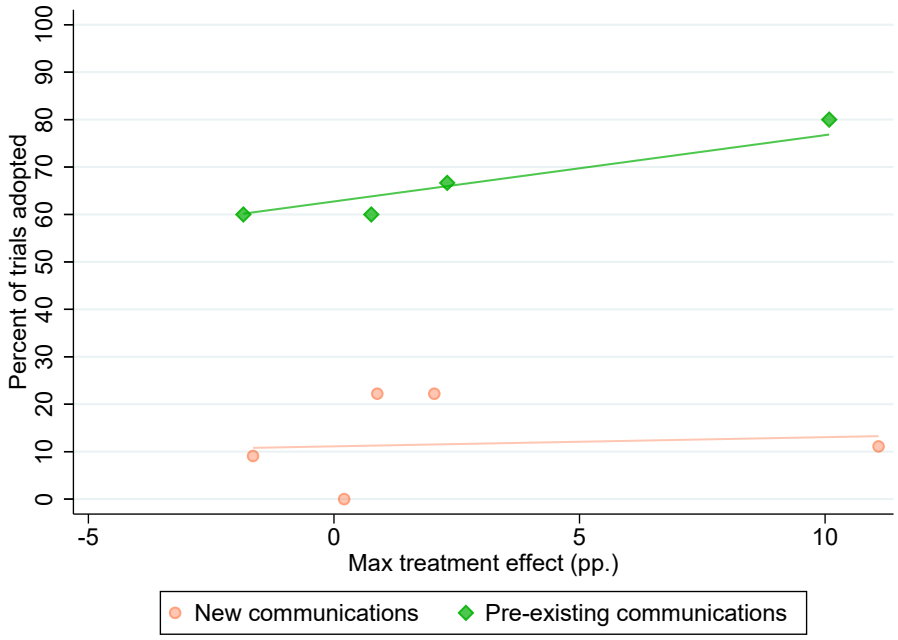
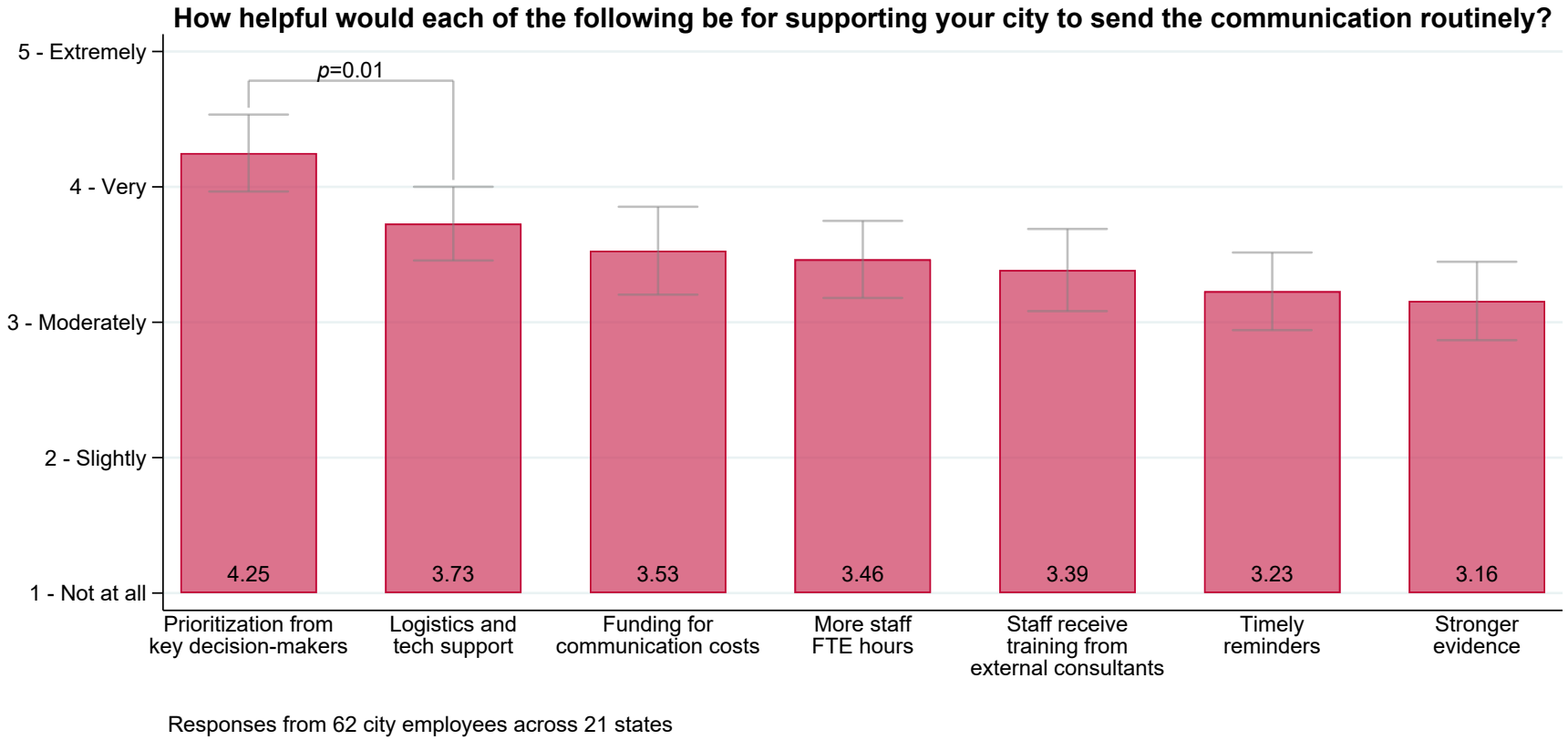


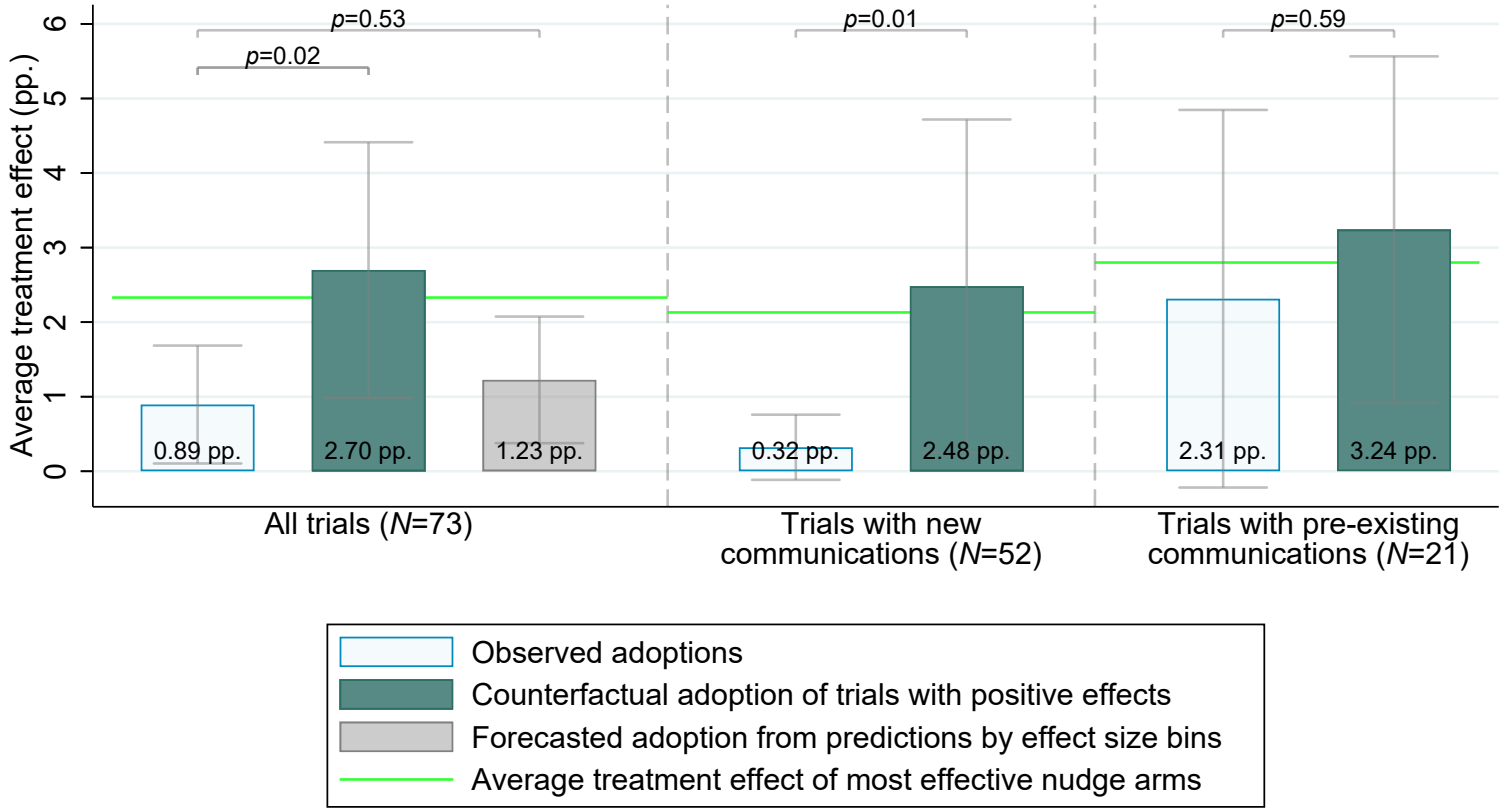
Figure 5.9a shows the adoption rates conditional on finding an effect that is positive and significant for new and pre-existing trials separately. Figure 5.9b shows the bin scatter of adoption rates on bins of effect sizes for new and pre-existing trials separately.

Figure 5.10: Survey evidence on organizational inertia



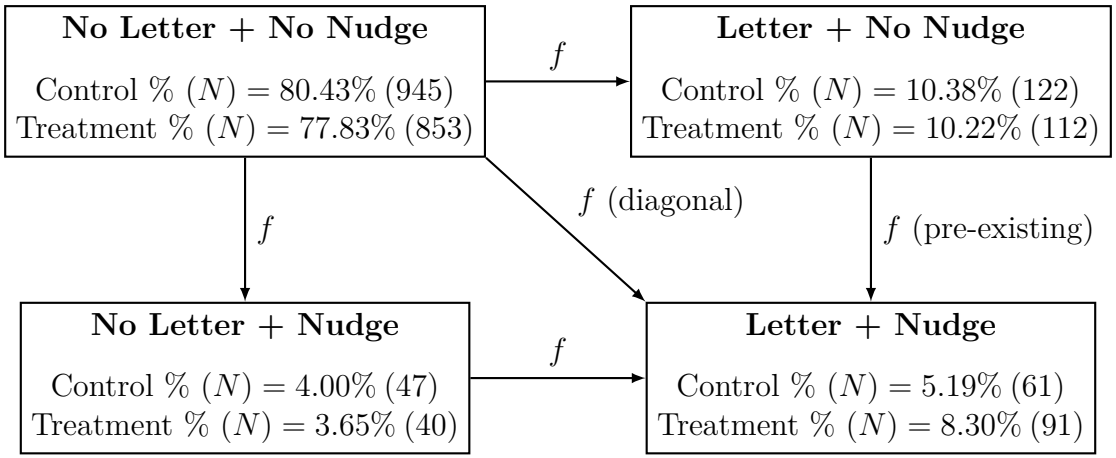
The respondents in this survey are (i) staff from BIT-trial cities where the nudge was not adopted though the effect size was either positively significant ($t > 1.96$) or greater than 1 pp., with responses from 17 employees in 14 cities answering for 25 of the 31 trials that meet this criteria (81% response rate), and (ii) 45 staff-members from a broader sample of U.S. cities with exposure to evidence-based communications (e.g., Chief Innovation Officers). 95% confidence intervals are shown with standard errors clustered by respondent.

Figure 5.11: Counterfactual adoption rules



This figure shows the average *adopted* treatment effect under: (1) actual adoptions, (2) a counterfactual rule of adopting all trials that found a positive effect, and (3) the forecasted adoption rates predicted by experts within the three effect size bins from Figure 5.5a. Specifically, we assign all non-adopted trials an adopted treatment effect of 0 pp. and assign all adopted trials the same effect size as their most effective treatment arm. Then we take the average of the adopted treatment effects across all trials. The average adopted treatment effects under actual adoptions and the counterfactual rule are shown separately for trials on new and pre-existing communications. See Section 5.5.3 for further details. 95% confidence intervals are shown.

Figure 5.12: Hjort et al. (2021) policy adoption experiment: Letter and nudge adoption in treatment and control groups



In the policy adoption experiment of Hjort et al. (2021), the researchers invite Brazilian mayors in the treatment group during a conference to a session providing evidence from research on tax payment reminder letters. The mayors attending the session were provided with a template for the letter highlighting three mechanisms: (1) the deadline, (2) the threat of audits or fines, and (3) social norm language. Mayors in the control group were not invited to this session.
15 to 24 months after the session, the researchers contacted the municipalities of the Brazilian mayors and asked whether the city sends a reminder communication for tax payments, and if so, (i) whether the communication is a physical letter and (ii) whether the language mentions the deadline, the threat of audits or fines, and social cues.
Using the data from this policy adoption experiment of Hjort et al. (2021), this figure shows the frequency in each cell, separately for the treatment and control groups. The adoption of the nudge is defined as including all 3 mechanisms (the deadline, the threat of audits or fines, and social cues) in the communication.

Table 5.1: Summary of papers on adoption of evidence

Paper	(1) No. of Decision-making Units	(2) No. of Interventions	(3) Intervention(s)	(4) Adoption Measure	(5) Average Adoption	(6) Moderators
Papers on Hypothetical Adoption						
Nakajima (2021)	2079 employees in U.S. state and local educational agencies	1	Charter schools	Choice between evidence from two studies	N/A	Sample size, sample population, research design, effect size, beliefs
Toma and Bell (2022)	192 employees across 22 U.S. federal agencies	5	Hypothetical government programs in health, education, and international development	Assessment of program value	N/A	Effect size, scale, policy outcome, policymaker numeracy, experience, cognitive noise
Vivalt and Coville (2023)	400 participants at World Bank or IDB workshops and headquarters	2	Cash transfer, school meals programs	Allocation of external funds to programs	N/A	Prior beliefs, effect size, variance, professions
Mehmood et al. (2022)	301 Pakistani deputy ministers	1	AI education training	Support for AI in policy	N/A	-
Papers on Adoption of One Best Practice						
Cho and Rust (2010)	10 sites of a U.S. car rental firm	1	Allow car rental price to vary by car age	Adoption of varied prices	0	-
Atkin et al. (2017)	132 soccer ball firms in Pakistan	1	Provide evidence of a more efficient ball-producing technology	Producing more than 1000 balls using the new method	0.14	Firm size, production quality, manager and employee skill, employee incentives
Bloom et al. (2020)	28 plants across 17 textile firms in India	1	Consultants introduce 38 standard management practices (e.g., quality control, inventory, HR, sales management)	Proportion of management practices adopted 9 years after consulting	0.46	Managerial turnover, director time, spillovers
Hjort et al. (2021)	1465 municipalities in Brazil	1	Encourage use of letter for timely tax payment	Use of tax reminder letter 1 year later	0.36	Mayor characteristics (e.g., gender, age, education, term), municipal characteristics (e.g., population, poverty rate), beliefs
Papers on Adoption of Multiple Interventions						
Kremer et al. (2019)	41 organizations awarded grants from USAID DIV	41	Various development RCTs (e.g., home solar systems, cook stoves)	Scaled to over 1 mil. beneficiaries	0.24	For-profit vs. non-profit, local partner, country population, academic affiliation, prior experimental evidence, pre-existing distribution network, cost of innovation
Wang and Yang (2022)	98 central ministries and commissions in China	633	Various policies before scaling nationally in China (e.g., carbon emission trading policy, agriculture catastrophe insurance)	National roll-out after regional experimentation	0.54	Local socioeconomic conditions, background of involved politicians, politician assignment process, complexity, ex ante uncertainty, effectiveness (growth rate of GDP per capita), policy domain, administrative level, fiscal shocks
DellaVigna et al. (2022)	67 departments across 30 U.S. cities	73	RCT within the city department to evaluate use of nudges in city communication	Use of nudge communication 2-6 years later	0.27	Effect size, staff retention, resources, behavioral mechanisms, pre-existence of communication

Table 5.2: Sample characteristics

	Overall	Effect size>median		City staff retained		Comm. pre-existed	
Frequency in category (%)	(1)	(2) No	(3) Yes	(4) No	(5) Yes	(6) No	(7) Yes
Nudge effectiveness							
Max $t \geq 1.96$	45.21	21.62	69.44*	44.44	45.65	44.23	47.62
Max treatment effect $\geq$ 1 pp.	46.58	0.00	94.44*	40.74	50.00	42.31	57.14
Organizational features							
City certified by What Works Cities	60.27	64.86	55.56	62.96	58.70	63.46	52.38
City staff member from trial retained	63.01	59.46	66.67	0.00	100.00*	59.62	71.43
Partner city dept. in charge of implementing	79.45	75.68	83.33	85.19	76.09	75.00	90.48
Senior city staff on trial (Director/Chief)	53.42	56.76	50.00	48.15	56.52	61.54	33.33*
Experimental design							
Communication pre-existed before trial	28.77	21.62	36.11	22.22	32.61	0.00	100.00*
Nudge communication uses Simplification	53.42	48.65	58.33	59.26	50.00	44.23	76.19*
Nudge communication uses Personal Motivation	57.53	56.76	58.33	70.37	50.00	61.54	47.62
Nudge communication uses Social Cues	56.16	59.46	52.78	51.85	58.70	55.77	57.14
Policy area							
Revenue collection & debt repayment	24.66	16.22	33.33	29.63	21.74	17.31	42.86
Registration & regulation compliance	20.55	13.51	27.78	14.81	23.91	19.23	23.81
Workforce & education	20.55	29.73	11.11	25.93	17.39	23.08	14.29
Take-up of benefits and programs	13.70	16.22	11.11	11.11	15.22	15.38	9.52
Community engagement	13.70	18.92	8.33	11.11	15.22	17.31	4.76
Health	5.48	5.41	5.56	7.41	4.35	5.77	4.76
Environment	1.37	0.00	2.78	0.00	2.17	1.92	0.00
Medium							
Physical letter	38.36	29.73	47.22	51.85	30.43	25.00	71.43*
Email	30.14	27.03	33.33	22.22	34.78	32.69	23.81
Postcard	21.92	27.03	16.67	22.22	21.74	30.77	0.00*
Text message	10.96	10.81	11.11	3.70	15.22	11.54	9.52
Website	4.11	5.41	2.78	0.00	6.52	3.85	4.76
Number of trials	73	37	36	27	46	52	21

This table shows the frequencies of trials for each category listed in the leftmost column. Column 1 shows the frequencies for all trials. Columns 2 and 3 partition the sample along the median of the maximum effect size in each trial. Columns 4 and 5 consider trials for which all the city collaborators from the trial have departed versus trials that have at least one original staff member still working in the same city department. Columns 6 and 7 distinguish between trials that tested nudges in a new communication and those that added nudges to a pre-existing communication that the city had been sending before the trial.

*Asterisk indicates that the p-value of the difference<0.05. Standard errors are clustered by city. Except when there are fewer than 5 trials in one of the 2×2 cells, p-values are calculated using the two-sided Fisher's exact test instead.

Table 5.3: Determinants of nudge adoptions

	OLS						Logit	ML
Dep. var.: Nudge adopted (0/1)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Max $t \geq 1.96$	0.02			-0.03	-0.16	-0.24	-0.20	-0.69
	(0.13)			(0.08)	(0.10)	(0.11)	(0.59)	(0.51)
Max treatment effect (10pp.)	0.06			0.10	0.14	0.23	0.78	
	(0.12)			(0.07)	(0.09)	(0.13)	(0.52)	
City staff retained		0.13		0.07	0.00	-0.06	0.61	-0.63
		(0.08)		(0.08)	(0.11)	(0.13)	(0.53)	(0.52)
Above-median city population		0.07		0.06			0.26	0.27
		(0.12)		(0.08)			(0.68)	(0.62)
What Works Cities certified		0.06		0.14			1.10	-0.07
		(0.12)		(0.11)			(0.86)	(0.61)
Communication pre-existed			0.53	0.52	0.59	0.60	2.91	2.58
			(0.13)	(0.13)	(0.14)	(0.15)	(0.67)	(0.79)
Mechanism								
Simplification & information			0.01	0.03	0.06	0.21	0.21	-0.46
			(0.10)	(0.10)	(0.13)	(0.14)	(0.80)	(0.81)
Personal motivation			-0.13	-0.12	-0.00	0.02	-0.93	-1.66
			(0.11)	(0.12)	(0.14)	(0.10)	(0.89)	(0.77)
Social cues			-0.06	-0.08	0.06	0.08	-0.66	-0.91
			(0.08)	(0.08)	(0.06)	(0.08)	(0.58)	(0.37)
Control take-up (10%)						0.02		
						(0.03)		
Uses online mediums						0.32		
						(0.12)		
Years since trial						-0.00		
						(0.06)		
City dept. in charge of implementing						0.29		
						(0.19)		
Senior city staff on trial (Director/Chief)						0.07		
						(0.14)		
Prior parameters								
μ_0								0.40
								(1.09)
σ_0								0.23
								(0.08)
Constant	0.25	0.12	0.22	0.05	0.07	-0.38	-2.71	
	(0.07)	(0.12)	(0.10)	(0.15)	(0.11)	(0.46)	(1.19)	
Average adoption rate	0.27	0.27	0.27	0.27	0.27	0.27	0.27	0.27
City fixed effects					✓	✓		
Policy area fixed effects						✓		
Number of trials	73	73	73	73	73	73	73	73
Number of cities	30	30	30	30	30	30	30	30
(Pseudo-)R^2	0.01	0.03	0.34	0.38	0.69	0.79	0.33	0.25

Standard errors clustered by city are shown in parentheses. Policy area fixed effects includes a dummy for each of the policy areas (Community engagement; Environment; Health; Registration & regulation compliance; Revenue collection & debt repayment; Take-up of benefits and programs; and Workforce & education). 3 trials are missing the data on the seniority of the city staff member working on the trial (Column 6); these trials are included with an indicator for missing. Column 8 estimates the model from Section 3 via maximum likelihood. The model specifies the distribution of the policymaker's prior on the percentage point effectiveness of the nudge as $N(\mu_0, \sigma_0^2)$. The policymaker updates after observing the treatment effect of the nudge from the trial. The weight placed on the signal is $\sigma_0^2/(\sigma_s^2 + \sigma_0^2)$, where σ_s^2 is the sampling variance or the square of the standard error, and the weight on the prior is $\sigma_s^2/(\sigma_s^2 + \sigma_0^2)$. The average sampling variance is 1.51, which gives a weight on the signal of 0.03, and the median is 0.35, which provides a signal weight of 0.13.

Table 5.4: Comparison of specific nudge adoption and broad adoption

Dep. var.: Adoption (0/1, OLS)	Nudge adoption (1)	Broad adoption (2)	Difference (3)
Max $t \geq 1.96$	-0.03	0.26	-0.28
	(0.08)	(0.11)	(0.15)
Max treatment effect (10pp.)	0.10	-0.13	0.23
	(0.07)	(0.08)	(0.12)
City staff retained	0.07	0.09	-0.02
	(0.08)	(0.08)	(0.11)
Above-median city population	0.06	-0.23	0.28
	(0.08)	(0.13)	(0.16)
What Works Cities certified	0.14	0.12	0.02
	(0.11)	(0.10)	(0.17)
Communication pre-existed	0.52	-0.08	0.61
	(0.13)	(0.09)	(0.18)
Mechanism			
Simplification & information	0.03	-0.00	0.03
	(0.10)	(0.08)	(0.14)
Personal motivation	-0.12	0.00	-0.12
	(0.12)	(0.10)	(0.16)
Social cues	-0.08	0.14	-0.22
	(0.08)	(0.10)	(0.15)
Constant	0.05	0.07	-0.02
	(0.15)	(0.12)	(0.20)
Average adoption rate	0.27	0.22	
Number of trials	73	73	
Number of cities	30	30	
R^2	0.38	0.23	

Standard errors clustered by city are shown in parentheses. In Column 1, the dependent variable is the same binary indicator from Table 3 for whether the city adopted the specific nudge in the trial. Column 1 replicates the baseline specification of Column 4 in Table 3. In Column 2, the dependent variable is a binary indicator for whether the city broadly adopted a similar nudge or the method of experimentation in other contexts.

Table 5.5: Hjort et al. (2021) policy adoption experiment: Persuasion rates

Persuasion rates (treatment-on-treated)	(1)	(2)	(3)	(4)
Panel A. Nudge adoption definition: All 3 mechanisms				
f	0.035	0.030	-0.010	-0.012
	(0.017)	(0.018)	(0.025)	(0.029)
f_{pe} (pre-existing)		0.417		-0.053
		(0.207)		(0.417)
f_{diag} (diagonal)			0.106	0.111
			(0.037)	(0.064)
MSE	1.696	0.517	0.003	0.000
Panel B. Nudge adoption definition: ≥ 2 of 3 mechanisms				
f	0.050	0.045	-0.002	0.026
	(0.023)	(0.022)	(0.028)	(0.059)
f_{pe} (pre-existing)		2.122		1.431
		(0.808)		(1.890)
f_{diag} (diagonal)			0.131	0.077
			(0.053)	(0.095)
MSE	2.062	0.059	0.196	0.000
Panel C. Nudge adoption definition: Social cues				
f	0.044	0.036	-0.016	0.006
	(0.018)	(0.019)	(0.027)	(0.033)
f_{pe} (pre-existing)		0.724		0.363
		(0.233)		(0.453)
f_{diag} (diagonal)			0.138	0.093
			(0.041)	(0.068)
MSE	3.178	0.239	0.202	0.000

This table shows the treatment-on-treated persuasion rates estimated from the model in Figure 12. The 3 mechanisms mentioned in the template for the tax reminder letter are the due date, the threat of audits or fines, and social norm language. MSE is the mean squared error in the 4 moments for the treatment group. The MSE for (4) is 0 since the model is exactly identified. Standard errors from 1000 bootstrap samples (resampled at the municipal level) are shown in parentheses.

6 Bibliography

References for Television and American consumerism

References

Advertising Age. (2003). "History: 1950s."

Advertising Age. (2005). "1950s TV turns on America."

Allen, S. and Thompson, R. (2017). "Television in the United States." *Encyclopedia Britannica.*

Bagwell, K. (2007). The economic analysis of advertising. *Handbook of Industrial Organization* 3: 1701-1844.

Bagwell, L. and Bernheim, B. (1996). "Veblen effects in a theory of conspicuous consumption." *American Economic Review* 86 (3): 349-373.

Baker, M. and George, L. (2010). "The role of television in household debt: Evidence from the 1950's." *The B E Journal of Economic Analysis & Policy* 10 (1).

Barnouw, E. (1990). *Tube of Plenty: The Evolution of American Television* (2nd ed.). New York, NY: Oxford University Press.

Baum-Snow, N. (2007). "Did highways cause suburbanization?" *Quarterly Journal of Economics* 122 (2): 775-805.

Beckinger, E. (1950). "Will television absorb radio?" *Advertising Agency* January issue.

Berkin, C., Miller, C., Cherny, R., Gormly, J., Egerton, D., and Woestman, K. (2011). *Making America: A History of the United States* (5th ed.). MA: Wadsworth.

Black, S., Devereux, P., and Salvanes, K. (2008). "Staying in the classroom and out of the maternity ward? The effect of compulsory schooling laws on teenage births." *The Economic Journal* 118: 1025-1054.

Bogart, L. (1958). *The Age of Television: A Study of Viewing Habits and the Impact of Television on American Life.* NY: Frederick Ungar Publishing Co.

Boustan, L. (2010). "Was Postwar suburbanization 'White Flight'? Evidence from the Black Migration." *Quarterly Journal of Economics* 125 (1): 417-443.

Brooks, T. and Marsh, E. (2003). *The Complete Directory to Prime Time Network and Cable TV 1946-Present* (8th ed.). NY: Ballantine Books.

Bursztyn, L. and Cantoni, D. (2016). "A tear in the Iron Curtain: The impact of Western television on consumption behavior." *The Review of Economics and Statistics* 98 (1): 25-41.

Campante, F. and Hojman, D. (2013). "Media and polarization: Evidence from the introduction of broadcast TV in the United States." *Journal of Public Economics* 100, 79-92.

Callaway, B. and Sant'Anna, P. (2020). "Difference-in-differences with multiple time periods." *Journal of Econometrics.*

Chaisemartin, C. and D'Haultfoeuille, X. (2020). "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9): 2964-2996.

Conley, T. (1999). "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92 (1): 1-45.

Coffin, T. (1955). "Television's impact on society." *American Psychologist* 10: 630-641.

Cohen, L. (2003). *A Consumer's Republic: The Politics of Mass Consumption in Postwar America*. NY: Vintage Books.

Cullen, J. (2003). *The American Dream: A Short History of an Idea That Shaped a Nation*. New York, NY: Oxford University Press.

DellaVigna S. and Kaplan, E. (2007). "The Fox News effect: Media bias and voting." *The Quarterly Journal of Economics* 122 (3): 1187-1234.

DellaVigna, S. and La Ferrara, E. (2015). "Economic and social impacts of the media." *Handbook of Media Economics* 1.

Durante, R., Pinotti, P., and Tesei, A. (2019). "The political legacy of entertainment TV." *American Economic Review* 109 (7): 2497-2530.

Edgerton, G. (2007). *The Columbia History of American Television*. NY: Columbia University Press.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). "Media and Political Persuasion: Evidence from Russia." *American Economic Review* 111: 3253–3285.

Fenton, G. and Koenig, F. (2021). "Labor Supply and Entertainment Innovations: Evidence From the U.S. TV Rollout." Working paper.

Fidler, F. (1949). "Television may lead to revolution in selling." *Advertising Agency* June issue.

Flory, Jr., J. (2008). *American Cars, 1946-1959: Every Model, Year by Year.* Jefferson, NC: McFarland & Company, Inc., Publishers.

Freidan, B. (1963). *The Feminine Mystique*. New York, NY: W. W. Norton & Company, Inc.

Galbraith, J. (1958). *The Affluent Society*. NY: Houghton Mifflin Harcourt.

Gentzkow, M. (2006). "Television and voter turnout." *Quarterly Journal of Economics* 121 (3): 931-972.

Gentzkow, M. and Shapiro, J. (2004). "Media, education and anti-Americanism in the Muslim world." *Journal of Economic Perspectives* 18: 117–133.

Gentzkow, M. and Shapiro, J. (2008). "Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study." *Quarterly Journal of Economics* 123 (1): 279-323.

George, L. (2009). "National television and the market for local products: The case of beer." *The Journal of Industrial Economics* 57 (1): 85-111.

Goodman-Bacon, A. (2021). "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*.

Haines, M. and Inter-university Consortium for Political and Social Research. (2010). *Historical, Demographic, Economic, and Social Data: The United States, 1790-2002*. https://doi.org/10.3886/ICPSR02896.v3

Halberstam, D. (1994). *The Fifties*. NY: Fawcett Columbine.

Heffetz, O. (2011). "A test of conspicuous consumption: Visibility and income elasticities." *The Review of Economics and Statistics* 93 (4): 1101-1117.

Hennigan, K., Del Rosario, M., Heath, L., Cook, T., Wharton, J., and Calder, B. (1982). "Impact of the introduction of television on crime in the United States: Empirical findings and theoretical implications." *Journal of Personality and Social Psychology* 42 (3): 461-477.

Inter-university Consortium for Political and Social Research (ICPSR). (2012). *United States Bureau of the Census: County and City Data Book Consolidated File: County Data, 1947-1977*. https://doi.org/10.3886/ICPSR07736.v2

Jensen, R. and Oster, E. (2009). "The power of TV: Cable television and women's status in India." *Quarterly Journal of Economics* 143: 1057–1094.

Katona, G. (1964). *The Mass Consumption Society*, MI: McGraw-Hill Inc.

Kearney, M. and Levine, P. (2015). "Media influences on social outcomes: The impact of MTV's 16 and Pregnant on Teen Childbearing." *American Economic Review* 105 (12): 3597-3632.

Kennedy, D., Cohen, L., and Bailey, T. (2010). *The American Pageant: A History of the American People Volume II since 1865* (Vol. 15th). MA: Wadsworth.

Krähmer, D. (2006). "Advertising and conspicuous consumption." *Journal of Institutional and Theoretical Economics* 162 (4): 661-682.

Koenig, F. (2021). "Technical change and superstar effects: Evidence from the rollout of television." Working paper.

Kopecky, K. and Suen, R. (2010). "A quantitative analysis of suburbanization and the diffusion of the automobile." *International Economic Review*, 51 (4): 1003-1037.

La Ferrara, E., Chong, A., and Duryea, S. (2012). "Soap operas and fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4: 1–31.

Leibman, N. (1995). *Living Room Lectures: The Fifties Family in Film and Television.* TX: University of Texas Press.

Lipsitz, G. (1990). *Time Passages: Collective Memory and American Popular Culture.* MN: University of Minnesota Press.

Mintz, S. and Kellogg, S. (1988). *Domestic Revolutions: A Social History of American Family Life.* NY: The Free Press.

Nacey, S. (2015). "Blurring fiction with reality: American television and consumerism in the 1950s." In A. Hulme (Ed.), *Consumerism on TV: Popular Media from the 1950s to the Present*, NY: Routledge.

National Broadcasting Company (NBC). (1952a). "Summer television advertising."

National Broadcasting Company (NBC). (1952b). "TV today: Report 2. Additional findings and research methods of the second NBC-Hofstra study."

O'Guinn, T. and Shrum, L. (1997). "The role of television in the construction of consumer reality." *Journal of Consumer Research* 2 (4): 278-294.

Olken, B. (2009). "Do TV and radio destroy social capital? Evidence from Indonesian villages." *American Economic Journal: Applied Economics* 1: 1–33.

Rayo, L. and Becker, G. (2006). "Peer comparisons and consumer debt." *The University of Chicago Law Review* 73: 231-248.

Roosevelt, E. (1961). "What has happened to the American Dream?" *The Atlantic.*

Samuel, L. (2001). *Brought to You By: Postwar Television Advertising and the American Dream.* TX: University of Texas Press.

Samuel, L. (2012). *The American Dream: A Cultural History.* NY: Syracuse University Press.

Sant'Anna, P. and Zhao, J. (2020). "Doubly robust difference-in-differences estimators." *Journal of Econometrics* 219 (1): 101–122.

Schor, J. (1999). *The Overspent American: Why We Want What We Don't Need.* NY: Harper-Collins.

Smith, S. (1990). *In All His Glory: The Life of William S. Paley.* NY: Random House Inc.

Shapiro, B., Hitsch, G., and Tuchman, A. (2021). "TV advertising effectiveness and profitability: Generalizable results from 288 brands." *Econometrica* 89 (4): 1855-1879.

Spigel, L. (2013). *Make Room for TV: Television and the Family Ideal in Postwar America.* IL: The University of Chicago Press.

Sterling, C. and Kittross, J. (2002). *Stay Tuned: A History of American Broadcasting* (3rd ed.). NJ: Lawrence Erlbaum Associates Publishers.

Sun, L. and Abraham, S. (2020). "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics.*

TIME. (1953). "Radio: Birth of a memo."

U.S. Bureau of Labor Statistics (2021). "Consumer price index for all urban consumers: All items in U.S. city average [CPIAUCNS]." Retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/CPIAUCNS.

Veblen, T. (1899). *The Theory of the Leisure Class: An Economic Study of Institutions.* U.K.: George Allen & Unwin Ltd.

Weiss, J. (2000). *To Have and to Hold: Marriage, the Baby Boom, and Social Change*. IL: The University of Chicago Press.

References for "Don't you plan on voting?" The motives and effects of peer pressure in voter mobilization

References

Ali, S. and C. Lin. 2013. "Why people vote: Ethical motives and social incentives." *American Economic Journal: Microeconomics*, 5 (2): 73-98

Balafoutas, L. and N. Nikiforakis. 2012. "Norm enforcement in the city: A natural field experiment." *European Economic Review*, 56: 1773-1785.

Balafoutas, L., N. Nikiforakis, and B. Rockenbach. 2014. "Direct and indirect punishment among strangers in the field." *Proceedings of the National Academy of Sciences*, 111 (45): 15924-15927.

Bandiera, O., I. Barankay, and I. Rasul. 2005. "Social preferences and response to incentives: Evidence from personnel data." *The Quarterly Journal of Economics,* 120 (3): 917-962.

Bennion, E. and D. Nickerson. 2011. "The cost of convenience: An experiment showing e-mail outreach decreases voter registration." *Political Research Quarterly*, 64 (4): 858-869.

Bond, R., C. Fariss, J. Jones, et al. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature*, 489, 295–298.

Bowles, S. and H. Gintis. 2002. "Social capital and community governance." *Economic Journal*, 112: 419-436.

Brennan, G. and A. Hamlin. 1998. "Expressive voting and electoral equilibrium." *Public Choice*, 95: 149-175.

Bronchetti, E., D. Huffman, and E. Magenheim. 2015. "Attention, intentions, and follow-through in preventive health behavior: Field experimental evidence on flu vaccination." *Journal of Economic Behavior and Organization*, 116: 270-291.

Burszytn, L. and R. Jensen. 2015. "How does peer pressure affect educational investments?" *The Quarterly Journal of Economics*, 130 (3): 1329-1367.

Burszytn, L. and R. Jensen. 2017. "Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure." *Annual Review of Economics*, 9:131-153.

Burszytn, L., G. Egorov, I. Haaland, A. Rao, and C. Roth. 2023. "Justifying dissent." *The Quarterly Journal of Economics*, 138 (3): 1403-1451.

Butera, L., R. Metcalfe, W. Morrison, and D. Taubinsky. 2022. "Measuring the welfare effects of shame and pride." *American Economic Review*, 112 (1): 122-168.

Danz, D., L. Vesterlund, and A. Wilson. 2022. "Belief Elicitation and Behavioral Incentive Compatibility." *American Economic Review*, 112 (9): 2851-83.

DellaVigna, S., J. List, and U. Malmendier. 2012. "Testing for altruism and social pressure in charitable giving." *The Quarterly Journal of Economics*, 127 (1): 1-56.

DellaVigna, S., J. List, U. Malmendier, and G. Rao. 2017. "Voting to tell others." *Review of Economic Studies*, 84: 143-181.

Feddersen, T. 2004. "Rational choice theory and the paradox of not voting." *Journal of Economic Perspectives*, 18 (1):99-112.

Fehr, E. and S. Gachter. 2000. "Cooperation and punishment in public goods experiments." *American Economic Review*, 90 (4): 982-993.

Fehr, E. and S. Gachter. 2002. "Altruistic punishment in humans." *Nature*, 415: 137-140.

Fudenberg, D. and P. Pathak. 2010. "Unobserved punishment supports cooperation." *Journal of Public Economics*, 94: 78-86.

Funk, P. 2010. "Social incentives and voter turnout: Evidence from the Swiss mail ballot system." *Journal of the European Economic Association*, 8 (5): 1077-1103.

Gerber, A., D. Green, and C. Larimer. 2008. "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American Political Science Review*, 102 (1): 33-48.

Gerber, A., D. Green, and C. Larimer. 2017. "Field experiments on voter mobilization: An overview of a burgeoning literature." In Banerjee, A. V. and E. Duflo (Eds), *Handbook of Economic Field Experiments*, Vol. 1 (pp. 395-438). North-Holland: Amsterdam, Netherlands.

Green, D. and A. Gerber. 2019. *Get Out the Vote: How to Increase Voter Turnout*. Brookings Institution Press.

Haenschen, K. 2016. "Social pressure on social media: Using Facebook status updates to increase voter turnout." *Journal of Communication*, 66: 542-563.

Jones, Z. 2020. "2020 Presidential Election odds, predictions & state-by-state electoral college projections as Biden betting odds slip." *Forbes*. Internet Archive (link). Retrieved 3 November 2020.

Kandel, E. and E. Lazear. 1992. "Peer pressure and partnerships." *Journal of Political Economy*, 100 (4): 801-817.

Karing, A. 2023. "Social signaling and childhood immunization: A field experiment in Sierra Leone." Working paper.

Levine, D. and A. Mattozzi. 2020. "Voter turnout and peer punishment." *American Economics Review*, 110 (10): 3298-3314.

Mas, A. and E. Moretti. 2009. "Peers at work." *American Economic Review*, 99 (1): 112-145.

Matland, R. and G. Murray. 2012. "An experimental test for 'backlash' against social pressure techniques used to mobilize voters." *American Politics Research*, 41 (3): 359-386.

Ostrom, E., J. Walker, and R. Gardner. 1992. "Covenants with and without a sword: Self-governance is possible." *American Political Science Review*, 86 (2): 404-417.

Perez-Truglia, R. and U. Troiano. 2017. "Shaming tax delinquents: Theory and evidence from a field experiment in the United States." *Journal of Public Economics*, 167: 120-137.

Schein, A., K. Vafa, D. Sridhar, et al. 2021. "Assessing the effects of friend-to-friend texting on turnout in the 2018 US Midterm Elections." In *Proceedings of the Web Conference 2021*: 2025-2036.

University of California Undergraduate Experience Survey (UCUES) Data Tables. 2020. Internet Archive (link). Retrieved 16 September 2023.

References

Angelucci, Charles, Elliott Ash, and Nicolas Longuet Marx. 2022. "The Nationalization of American Lawmaking? Evidence from State Statutes." Working paper.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, 107 (3), 797–817.

Barrios, T., Rebecca Diamond, Guido W. Imbens, and Michal Kolesar. 2012. "Clustering, Spatial Correlations, and Randomization Inference." *Journal of the American Statistical Association*, 107 (498), 578-591.

Bernecker, Andreas, Pierre C. Boyer, and Christina Gathmann. 2021. "The Role of Electoral Incentives for Policy Innovation: Evidence from the U.S. Welfare Reform." *AEJ: Economic Policy*, 13 (2), 26-57.

Berry, William D., Evan J. Ringquist, Richard C. Fording, and Russell L. Hanson. 1998. "Measuring Citizen and Government Ideology in the American States, 1960-93." *American Journal of Political Science*, 42 (1), 327-348.

Besley, Timothy and Anne C. Case. 1995. "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition." *American Economic Review*, 85 (1), 25-45.

Besley, Timothy and Torsten Persson, "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review*, September 2009, 99 (4), 1218–44.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100 (5), 992-1026.

Boehmke, Frederick and Paul Skinner. 2012. "State Policy Innovativeness Revisited." *State Politics and Policy Quarterly*, 12 (3), 303-329.

Boehmke, Frederick, Mark Brockway, Bruce Desmarais, Jeffrey Harden, Scott LaCombe, Fridolin Linder, and Hanna Wallach. 2020. "SPID: A New Database for Inferring Public Policy Innovativeness and Diffusion Networks." *Policy Studies Journal*, 48 (2), 517-545.

Bowen, Daniel C. and Zachary Greene. 2014. "Should We Measure Professionalism with an Index? A Note on Theory and Practice in State Legislative Professionalism Research." *State Politics & Policy Quarterly*, 14 (3), 277-296.

Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. Forthcoming. "Cross-country Trends in Affective Polarization." *Review of Economics and Statistics*.

Buera, Francisco J., Alexander Monge-Naranjo, and Giorgio E. Primiceri. 2011. "Learning the Wealth of Nations." *Econometrica*, 79 (1), 1–45.

Callander, Steven, and Bård Harstad. 2015. "Experimentation in Federal Systems." *Quarterly Journal of Economics*, 130 (2), 951–1002.

Canen, Nathan, Chad Kendall, and Francesco Trebbi. 2020. "Unbundling polarization." *Econometrica*, 88 (3), 1197-1233.

Case, Anne C., Harvey S. Rosen, and James R. Hines Jr. 1993. "Budget spillovers and fiscal policy interdependence: Evidence from the states." *Journal of Public Economics*, 52 (3), 285-307.

Caughey, Devin, and Christopher Warshaw. 2016. "The Dynamics of State Policy Liberalism, 1936–2014." *American Journal of Political Science,* 60 (4), 899–913.

Caughey, Devin, and Christopher Warshaw. 2018. "Policy Preferences and Policy Change: Dynamic Responsiveness in the American States, 1936–2014." *American Political Science Review,* 112 (2), 249–266.

Caughey, Devin, Christopher Warshaw, and Yiqing Xu. 2017. "Incremental democracy: The policy effects of partisan control of state government." *The Journal of Politics*, 79 (4), 1342-1358.

Cantoni, Enrico and Vincent Pons. 2021. "Strict ID Laws Don't Stop Voters: Evidence from a U.S. Nationwide Panel, 2008–2018." *The Quarterly Journal of Economics*, 136 (4), 2615-2660.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics* 134 (3), 1405-1454.

Collins, William J. 2003. "The Political Economy of State-level Fair Employment Laws, 1940-64." *Explorations in Economic History*, 40, 24-51.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *American Economic Review*, 100 (1), 35-69.

Cui, Zhihan, Geoffrey Heal, Howard Kunreuther and Lu Li. 2021. "The Political Economy of Responses to COVID-19 in the U.S.A." NBER Working Paper w28578.

de Chaisemartin, Clément and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110 (9): 2964-96.

de Paula, Aureo, Imran Rasul, and Pedro CL Souza. 2020. "Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition." Working paper.

Desmarais, Bruce A., Jeffrey J. Harden, and Frederick J. Boehmke. 2015. "Persistent Policy Pathways: Inferring Diffusion Networks in the American States." *American Political Science Review*, 109 (2), 392-406.

Erikson, Robert S. Gerald C. Wright, Jr., and John P. McIver. 1989. "Political Parties, Public Opinion, and State Policy in the United States." *American Political Science Review*, 83: 729-750.

Freyaldenhoven, Simon, Christian Hansen, Jorge Perez Perez, and Jesse M. Shapiro. Forthcoming. "Visualization, Identification, and Estimation in the Linear Panel Event-Study Design." *Advances in Economics and Econometrics: Twelfth World Congress.*

Fiorina, Morris P. and Samuel J. Abrams. 2008. "Political polarization in the American public" *Annual Review of Political Science*, 11, 563-588.

Geary, R. C. 1954. "The Contiguity Ratio and Statistical Mapping." *The Incorporated Statistician*, 5 (3), 115-146.

Graham, E. R., Charles R. Shipan, and Craig Volden. 2012. "Review Article: The Diffusion of Policy Diffusion Research in Political Science." *British Journal of Political Science*, 43, 673-701.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica,* 25 (4), 501-522.

Gruber, Jonathan and Benjamin Sommers. 2020. "Fiscal Federalism and the Budget Impacts of the Affordable Care Act's Medicaid Expansion." Working paper.

Grumbach, J. M. 2018. "From Backwaters to Major Policymakers: Policy Polarization in the States, 1970–2014." *Perspectives on Politics*, 16 (2), 416-435.

Goodman-Bacon, Andrew. 2021. "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes." *American Economic Review,* 111 (8), 2550-2593.

Heckman, James and Sidharth Moktan. 2020. "Publishing and Promotion in Economics: The Tyranny of the Top Five." *Journal of Economic Literature*, 58 (2), 419-470.

Hjort, Jonas, Diana Moreira, Gautam Rao and Juan Francisco Santini. 2021. "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities." *American Economic Review*, 111 (5), 1442-80.

Hoynes, Hilary W. and Diane Whitmore Schanzenbach. 2009. "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program." *AEJ: Applied Economics,* 1 (4), 109-139.

Knight, Brian. 2013. "State Gun Policy and Cross-State Externalities: Evidence from Crime Gun Tracing." *American Economic Journal: Economic Policy*, 5 (4): 200–229.

Lagnodny, Julius, Rebekah Jones, Julianna Koch, and Peter K. Enns. 2022. "A Validation and Extension of State-Level Public Policy Mood: 1956 to 2020." *State Politics and Policy Quarterly*, 1-14.

Mallinson, Daniel J. 2020. “The spread of policy diffusion studies: A systematic review and meta-analysis, 1990-2018.” Working paper.

Mallinson, Daniel J. 2021. “Who are your neighbors? The role of ideology and decline of geographic proximity in the diffusion of policy innovations.” *Policy Studies Journal*, 49 (1), 67-88.

Manski, C. 1993. “Identification of endogenous social effects: The reflection problem.” *The Review of Economic Studies*, 60 (3): 531-542.

McCarty, Nolan. 2019. *Polarization: What Everyone Needs to Know*. Oxford University Press.

Mulligan, Casey B. and Andrei Shleifer. 2005. “The Extent of the Market and the Supply of Regulation.” *The Quarterly Journal of Economics*, 120 (4), 1445-1473.

Poole, Keith T. and Howard Rosenthal. 1985. “A spatial model for legislative roll call analysis.” *American Journal of Political Science*. 29 (2), 357-384.

Shor, Boris and Nolan McCarty. 2011. “The ideological mapping of American legislatures.” *American Political Science Review*, 105 (3), 530-551.

Strumpf, Koleman S., and Felix Oberholzer-Gee. 2002. "Endogenous policy decentralization: Testing the central tenet of economic federalism." *Journal of Political Economy* 110.1: 1-36.

Volden, Craig, Michael Ting, and Daniel Carpenter. 2008. “A formal model of learning and policy diffusion.” *American Political Science Review*, 102 (3), 319-332.

Walker, Jack L. 1969. “The Diffusion of Innovations among the American States.” *The American Political Science Review*, 63 (3), 880-899.

Wang, Shaoda and David Y. Yang. 2021. “Policy Experimentations in China: the Political Economy of Policy Learning.” Working paper.

References for Bottlenecks for evidence adoption

References

Allcott, Hunt. 2015. “Site Selection Bias in Program Evaluation.” *The Quarterly Journal of Economics* 130 (3): 1117-1165.

Altonji, Joseph G., Todd E. Elder and Christopher R. Taber. 2005. “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools.” *Journal of Political Economy* 113(1): 151-184.

Argote, Linda and Ella Miron-Spektor. 2011. "Organizational Learning: From Experience to Knowledge." *Organization Science* 22 (5): 1123-1137.

Atkin, David, Azam Chaudhry, Shamyla Chaudry, Amit K. Khandelwal, and Eric Verhoogen. 2017. "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan." *The Quarterly Journal of Economics* 132 (3): 1101-1164.

Athey, Susan and Michael Luca. 2019. "Economists (and Economics) in Tech Companies." *Journal of Economic Perspectives* 33 (1): 209-230.

Banerjee, Abhijit V. and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1: 151-178.

Baron, J. 2018. "A Brief History of Evidence-based Policy." *The Annals of the American Academy of Political and Social Science*, 678 (1): 40-50.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041-1055.

Besley, Tim and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218-1244.

Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts. 2020. "Do Management Interventions Last? Evidence from India." *American Economic Journal: Applied Economics* 12(2): 198–219.

Cho, Sungjin and John Rust. 2010. "The Flat Rental Puzzle." *The Review of Economic Studies* 77 (2): 560–594.

DellaVigna, Stefano, Woojin Kim, and Elizabeth Linos. 2023. "Replication Data for: 'Bottlenecks for evidence adoption'." Harvard Dataverse `https://doi.org/10.7910/DVN/XOCJOF`.

DellaVigna, Stefano and Elizabeth Linos. 2022. "RCTs to scale: Comprehensive evidence from two nudge units" *Econometrica* 90, 81-116.

DellaVigna, Stefano, Devin Pope, and Eva Vivalt. 2019. "Predict science to improve science." *Science* 366 (6464): 428-429.

de Vries, Hanna, Victor Bekkers, and Lars Tummers. 2015. "Innovation in the Public Sector: A Systematic Review and Future Research Agenda." *Public Administration.*

Development Impact Evaluation (DIME). 2019. "Science for Impact: Better Evidence for Better Decisions." *World Bank Group.*

Fernandez, Sergio and Lois Wise. 2010. "An Exploration of Why Public Organizations 'Ingest' Innovations." *Public Administration* 88 (4): 979-998.

Foundations for Evidence-Based Policymaking Act, H.R. 4174, 115th Cong. 2018. https://www.congress.gov/bill/115th-congress/house-bill/4174.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2021. "How research affects policy: Experimental evidence from 2,150 Brazilian municipalities" *American Economic Review* 111 (5), 1442-80.

Kasy, Maximilian and Anja Sautmann. 2021. "Adaptive treatment assignment in experiments for policy choice." *Econometrica* 89 (1): 113-132.

Kremer, Michael, Sasha Gallant, Olga Rostapshova, and Milan Thomas. 2019. "Is Development Innovation a Good Investment? Which Innovations Scale? Evidence on social investing from USAID's Development Innovation Ventures." Working paper.

Levitt, Barbara and James G. March. 1988. "Organizational Learning." *Annual Review of Sociology*, 14, 319-338.

List, John. 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale.* New York, NY: Random House

Meager, Rachael. 2019. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11 (1): 57-91.

Mehmood, Sultan, Shaheen Naseer, and Daniel Chen. 2022. "AI Education as State Capacity: Experimental Evidence from Pakistan." Working paper.

Milkman, Katherine L., Dena Gromet, Hung Ho, et al. 2021. "Megastudies Improve the Impact of Applied Behavioural Science." *Nature* 600, 478-483.

Muralidharan, Karthik and Paul Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives* 31 (4): 103-24.

Nakajima, Nozomi. 2021. "Evidence-Based Decisions and Education Policymakers." Working paper.

Naranjo-Gil, D. 2009. "The Influence of Environmental and Organizational Factors on Innovation Adoptions: Consequences for Performance in Public Sector Organizations." *Technovation* 29 (12): 810-818.

Simon, Herbert A. 1997. *Administrative Behavior*. New York, NY: The Free Press.

Thaler, Richard and Cass Sunstein. 2008. *Nudge*. New Haven, CT: Yale University Press.

Toma, Mattie and Elizabeth Bell. 2022. "Understanding and Improving Policymakers' Sensitivity to Program Impact." Working paper.

Vivalt, Eva. 2020. "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economic Association* 18 (6), 3045–3089.

Vivalt, Eva and Aidan Coville. 2023. "How Do Policymakers Update Their Beliefs?" *Journal of Development Economics* 165, 1-14.

Wang, Shaoda and David Yang. 2022. "Policy Experimentation in China: The Political Economy of Policy Learning." *NBER Working Paper No. 29402.*

Appendix A

Appendix for Television and American consumerism

A.1 Business categories in the *Census of Business*

Food group. These establishments sell food mainly for home consumption. They include grocery stores; meat markets; fish markets; fruit stores; vegetable markets; candy, nut, and confectionery stores; dairy product stores; milk dealers; bakeries; egg and poultry dealers; delicatessen stores; and specialized food stores.

Eating and drinking places. Distinct from the food group, eating and drinking places offer food for consumption *on premises*. These include restaurants; cafeterias; caterers; lunch counters; refreshment stands; and drink places.

General merchandise. These stores sell an array of products such as dry goods, clothes, accessories, furniture, and food. They include country general stores (typically in rural areas); department stores; dry goods (e.g., clothes and hardware) stores; general merchandise stores; and variety stores, which sell items such as stationery and gifts usually at a low price (e.g, "5 and 10 cents" stores, the equivalent to a contemporary dollar store).

Apparel group. Apparel stores specialize in clothing and accessories. These also include footwear stores; custom tailors; hosiery and lingerie stores; fur shops; and horse-back riding apparel stores.

Furniture, furnishings, and appliance group. This group offers home furniture and appliances, as well as furnishings such as drapes and flooring. They include furniture stores; floor covering stores; drapery, curtain, and upholstery stores; china, glassware, and metalware stores; antique stores; household appliance stores (including televisions); and radio stores.

Automotive group. These stores mainly sell cars and automobile accessories, but can also be shops for other vehicles such as boats. They include motor vehicles (new and used) dealers; tire, battery, and accessory dealers; motorcycle dealers; aircraft dealers; and boat dealers.

Gasoline service stations. These stations sell gasoline and lubricating oils. They may also offer repair and maintenance services.

Lumber, building, and hardware group. This group offers materials and hardware for construction to both contractors and individuals. They include lumber yards; building

material dealers; paint, glass, and wallpaper stores; heating and plumbing equipment dealers; electrical supply stores; hardware stores; and farm equipment stores.

Drug and proprietary stores. These stores can sell pharmaceutical drugs, medicines, and miscellaneous items such as cosmetics, toiletries, candy, and liquor.

Other retail stores. The remaining category encompasses a wide variety of remaining establishments. They include fuel dealers (e.g., coal, wood, and bottled gas); ice dealers; hay, grain, and feed stores; farm and garden supply stores; jewelry stores; bookstores; stationery stores; sporting goods stores; bicycle stores; florists; cigar stores; news dealers and newsstands; gift, novelty, and souvenir stores; music stores; luggage and leather goods stores; optical goods stores; camera and photography supply stores; and office, store machine, and equipment dealers.

A.2 Construction Permit-holders vs. Pending Applicants

The "Pending Applicants" (PA) and "Construction Permit-holders" (CP) are a subset of DMAs that had applications filed with the FCC for the first TV stations in their markets when the Freeze was announced. These broadcasters who had applied for markets in the Pending Applicants group—including some who had even cleared the expensive hearing process—had their plans put on an indefinite hold (Sterling and Kittross, 2002; Figure A.13a). Instead of the typical six-month process, it would eventually take them over four years to get a license.

The same could have happened to the broadcasters targeting the Construction Permit-holders group. These broadcasters had cleared the application process when the Freeze was announced and had just been issued construction permits, but had yet to build their stations. Instead of suspending those broadcasters as well, the FCC somewhat arbitrarily decided to let them "build without fear" and go on air during the Freeze (Figure A.13b).

The Gentzkow and Shapiro (2008) dataset records when the first local TV stations went on air, but does not list when the stations applied for broadcasting licenses. To find the Pending Applicants, I search historical issues of *Television Digest* for DMAs that had an applicant on file *before* the Freeze was announced but received TV service only *after* the Freeze ended in 1952. The Construction Permit-holders are identified as the markets that went on air during the Freeze.

Figure A.14 shows the distribution of TV broadcasting start dates for the Construction Permit-holders and Pending Applicants, and Table A.1 summarizes the pre-TV differences in the retail outcomes and demographic variables. Relative to the full Pre- and Post-Freeze groups in Table 5.2, these subgroups are more comparable in virtually every dimension. In fact, even the initial difference in log retail sales is not statistically significant.

Table A.9 tests the selection of Construction Permit-holders versus Pending Applicants using all pre-TV years in the panel. The difference is large in magnitude yet statistically insignificant without any demographic controls (Column 1) and shrinks close to zero when the controls are included (Column 3).

As mentioned in Section 2.5.2, Column 5 of Table C.6 replicates the main finding on retail sales comparing these subgroups.

A.3 Appendix figures and tables

Figure A.1: *Census of Business* record example

Table 103.—COUNTIES AND CITIES OF 2,500 INHABITANTS OR MORE[1]

ESTABLISHMENTS, SALES, PAY ROLL, AND PERSONNEL—TOTAL RETAIL STORES AND SALES, BY KIND-OF-BUSINESS GROUP: 1948

County and city	Total						Kind-of-business group																	
	Establishments	Sales, entire year	Pay roll, entire year	Active proprietors, November	Paid employees, workweek ended nearest Nov. 15		Food group		Eating and drinking places		General merchandise group, general stores		Apparel group		Furniture furnishings, appliance group		Automotive group		Gasoline service stations		Lumber, building, hardware group		Drug and proprietary stores		All other retail stores	
					Total	Full workweek	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales	Stores	Sales
	(Number)	(Thousand dollars)	(Thousand dollars)	(Number)	(Number)	(Number)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)	(Number)	(Thous. dol.)
State, total	121,290	11,019,804	1,369,411	127,925	546,816	460,406	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Warehouses, etc.	179	...	53,279	...	12,601	11,856	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Stores	12,111	11,019,804	1,316,132	127925	534,215	448,550	23698	2,637,116	26367	1,202,595	3921	1,312,696	9,372	769,829	8,771	704,983	6,055	1,722,993	14,197	593,421	5,959	902,443	3,354	364,618	39,021	1,009,246
Alameda County, total	8,554	808,445	106,637	8,995	40,197	33,509	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Warehouses, etc.	17	...	6,011	...	1,166	1,151	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Stores	[2] 537	808,445	100,626	8,995	39,031	32,358	2,035	203,697	1,910	66,848	231	127,703	659	63,542	496	56,569	391	119,290	853	35,595	291	41,540	269	26,412	1,408	67,329
Alameda	420	39,214	4,164	445	1,469	1,234	109	15,434	79	2,932	12	1,276	43	1,835	26	2,188	20	8,112	40	2,103	14	812	16	2,563	61	1,959
Albany	158	14,005	1,328	180	427	347	34	3,286	38	1,684	4	189	12	380	11	527	12	4,833	13	566	5	1,373	4	328	23	857
Berkeley	1,128	103,099	12,065	1,145	4,674	3,879	290	32,522	160	7,509	40	8,859	95	8,475	72	5,497	53	15,234	119	6,423	40	4,513	47	4,770	212	9,365
Emeryville	77	5,619	865	80	284	230	7	2,823	40	1,004	...	...	1	(x)	...	...	4	681	8	290	2	(x)	1	(x)	14	367
Hayward	253	27,727	2,691	277	1,024	871	48	6,411	30	1,327	10	1,990	30	2,236	17	2,016	19	5,603	20	751	12	1,310	7	829	40	5,282
Livermore	93	6,319	644	101	269	231	14	(x)	22	(x)	5	300	5	(x)	9	487	9	671	7	204	6	(x)	2	(x)	14	978
Oakland	5,317	548,380	73,049	5,547	28,509	23,627	1,274	115,769	1,297	45,834	127	112,418	415	47,430	309	43,851	236	76,439	482	19,820	149	26,369	163	15,826	905	44,424
Piedmont	7	651	66	6	29	16	1	(x)	1	(x)	...	...	...	...	...	...	...	...	4	270	...	...	1	(x)	...	...
San Leandro	338	26,752	2,529	346	987	757	78	10,977	58	1,452	12	1,485	32	1,655	18	874	13	4,505	36	1,667	20	1,686	15	1,165	56	1,286
Remainder of County	746	36,649	3,225	868	1,349	1,146	100	14,404	205	4,404	21	986	26	1,112	28	1,129	25	3,172	122	3,521	43	4,434	13	676	83	2,811
Alpine County	6	115	9	5	4	2	1	(x)	1	(x)	...	...	...	...	...	...	...	...	3	50	...	...	...	...	1	(x)
Amador County	147	8,229	645	168	300	238	30	2,511	45	1,178	8	626	6	307	5	173	10	1,320	16	485	6	870	7	234	14	525
Butte County	667	72,267	7,060	913	2,991	2,485	170	19,446	174	5,282	31	6,187	67	3,209	42	2,568	67	12,464	100	3,286	70	11,428	22	1,834	124	6,963
Chico	285	30,445	3,356	290	1,362	1,147	32	6,039	47	1,707	10	3,094	31	1,648	19	1,154	31	6,640	29	1,273	21	5,022	9	883	56	2,963
Oroville	158	18,181	1,908	166	816	693	24	5,298	30	1,360	6	2,559	19	971	10	850	17	3,580	14	553	10	1,255	6	498	22	1,257
Remainder of County	424	23,641	1,796	457	813	645	114	8,109	97	2,215	15	534	17	590	13	564	19	2,244	57	1,460	39	5,129	7	453	46	2,343
Calaveras County	140	6,262	440	174	230	194	34	2,083	40	(x)	11	960	8	160	3	(x)	7	930	19	430	5	281	4	193	9	166
Colusa County	187	14,744	1,319	208	585	499	31	2,984	47	1,440	10	944	5	178	6	262	13	1,511	26	1,056	23	5,272	6	433	20	664
Contra Costa County	2,632	192,722	19,078	2,994	7,538	6,214	625	67,627	683	18,989	86	11,211	191	9,860	141	11,729	130	27,476	265	12,484	124	13,477	71	6,067	316	13,324
Antioch	152	9,864	856	181	358	272	37	3,362	26	659	5	446	13	376	12	653	13	1,291	12	608	9	1,184	5	341	18	744
El Cerrito	153	11,031	1,148	167	444	348	26	4,919	46	1,518	4	114	10	208	9	809	6	219	22	983	9	1,442	6	398	15	421
Martinez	166	12,132	1,304	170	477	366	29	2,770	37	959	5	684	21	1,170	13	1,040	16	2,956	13	438	9	855	5	428	18	772
Pittsburg	276	17,660	1,775	321	875	699	62	4,818	90	2,332	9	2,533	33	1,870	16	1,241	8	1,536	12	669	7	658	6	664	33	1,339
Richmond	753	71,194	7,571	858	2,751	2,274	203	25,406	169	6,676	26	5,786	58	4,283	40	5,022	40	12,356	67	3,677	27	2,546	17	1,970	106	5,472
Remainder of County	1,132	70,851	6,430	1,297	2,633	2,155	268	26,352	315	8,845	37	1,648	56	1,733	51	2,964	47	9,118	137	6,099	63	6,792	32	2,266	126	6,074
Del Norte County	152	8,084	727	152	362	304	24	2,197	47	1,446	9	517	7	380	7	281	8	912	19	821	12	799	3	254	16	477
El Dorado County	257	13,811	1,127	275	492	426	57	5,131	72	1,780	8	397	17	790	12	562	15	2,034	37	1,005	11	688	3	254	25	1,152
Placerville	95	8,743	864	89	353	315	14	2,347	21	844	2	(x)	11	712	6	425	12	1,995	8	433	5	475	3	254	13	(x)
Remainder of County	162	5,068	263	186	139	121	43	2,784	51	936	6	(x)	6	78	6	137	3	39	29	572	6	213	...	...	12	(x)
Fresno County, total	3,222	285,519	29,141	3,450	12,443	10,529	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Warehouses, etc.	4	...	739	...	148	148	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Stores	3,218	285,519	28,402	3,450	12,295	10,381	781	66,176	683	20,898	117	27,350	188	20,025	197	16,295	182	45,991	393	19,434	192	37,831	77	8,442	408	27,077
Coalinga	83	7,297	627	91	333	278	11	2,253	12	570	4	439	8	290	7	268	11	1,688	12	495	5	640	3	(x)	12	(x)
Fresno	1,562	185,945	21,135	1,615	8,906	7,569	307	32,066	311	12,566	50	22,657	127	17,674	118	12,328	87	21,385	187	8,009	79	24,628	47	6,609	249	18,023
Reedley	105	10,666	927	123	387	335	18	2,333	17	903	4	575	8	329	10	609	11	2,187	10	298	11	2,647	4	261	12	942
Sanger	110	8,451	762	131	329	269	28	2,504	22	498	5	694	8	346	6	595	9	1,523	12	531	7	893	2	(x)	11	(x)
Selma	113	9,458	674	133	323	259	23	2,457	21	574	5	532	10	504	8	746	13	1,558	9	482	5	706	3	254	16	1,645
Remainder of County	1,243	63,702	4,247	1,357	2,017	1,671	394	24,563	300	6,225	49	2,403	27	882	48	1,749	51	7,650	163	5,619	85	8,317	18	941	108	5,353
Glenn County	211	17,689	1,648	239	709	602	30	3,491	46	1,529	12	1,959	10	358	15	593	16	2,499	19	986	24	3,002	5	308	34	3,014
Humboldt County	843	75,204	7,051	994	3,012	2,481	169	18,568	212	7,698	39	8,667	48	2,877	42	4,369	38	13,180	99	4,296	48	6,762	17	1,466	111	7,321
Arcata	69	9,014	895	73	381	322	10	2,447	14	668	2	(x)	5	292	4	384	5	1,063	7	489	9	1,598	2	(x)	11	897
Eureka	376	43,714	4,424	430	1,871	1,520	59	8,859	92	3,877	11	5,818	28	1,987	26	2,909	36	10,338	36	1,574	16	2,883	8	805	64	4,664
Remainder of County	398	22,476	1,732	491	760	639	100	7,262	106	3,153	26	(x)	15	598	12	1,076	17	1,779	56	2,233	23	2,281	7	(x)	36	1,800
Imperial County	733	72,951	7,809	724	3,203	2,754	152	16,904	196	5,342	22	5,688	40	4,705	25	1,527	40	11,577	120	5,771	42	12,352	17	2,163	73	6,922
Brawley	184	15,680	1,782	175	770	676	38	4,306	53	1,503	6	1,021	8	463	5	406	9	2,158	24	1,086	10	3,034	4	415	27	1,568
Calexico	83	10,684	902	85	458	390	20	3,334	12	960	4	1,730	11	1,310	4	187	3	1,031	9	385	6	1,481	4	402	10	404
El Centro	182	26,170	3,473	162	1,242	1,117	24	3,761	50	1,743	3	2,064	17	2,726	8	517	17	6,273	25	1,542	9	4,531	4	925	25	2,086
Holtville	51	3,839	363	56	163	138	7	1,010	13	554	1	(x)	1	(x)	2	(x)	1	(x)	14	451	4	939	1	(x)	3	291
Imperial	36	4,909	490	39	237	177	7	1,469	7	87	2	(x)	...	...	3	(x)	2	(x)	7	241	2	(x)	1	(x)	3	387
Westmorland	38	1,310	88	38	55	46	5	469	21	295	...	...	2	(x)	...	...	...	...	7	432	2	(x)	1	(x)	...	...
Remainder of County	159	10,419	711	169	278	250	51	2,855	38	836	6	356	1	(x)	3	(x)	8	654	34	1,634	9	1,645	2	(x)	5	2,186

See footnotes at end of table.

Figure A.2: Merchandise Line Sales data from 1948 *Census of Business*

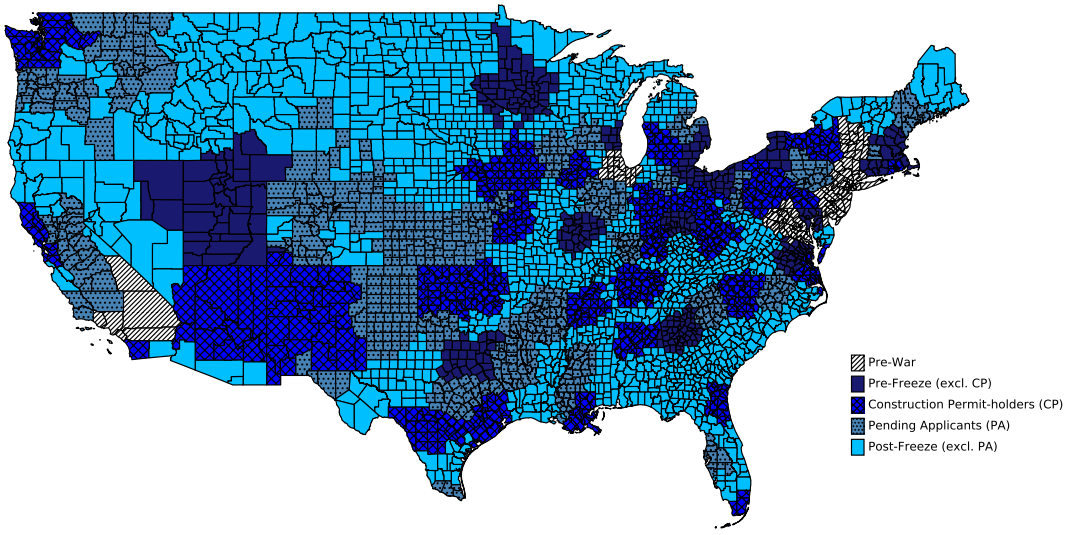
16.02 **RETAIL TRADE—MERCHANDISE LINE SALES**

Table 16A.—MERCHANDISE LINE SALES—FOOD GROUP—UNITED STATES, BY KIND OF BUSINESS: 1948

[Data based on sample. See text for description of sample and other explanatory material.]

Kind of business and merchandise line	Merchandise line sales: Amount (Thousand dollars) (1)	Merchandise line sales: Percent distribution (2)	Stores handling merchandise line: Number (3)	Stores handling merchandise line: Total sales (Thousand dollars) (4)	Col. 1 as percent of col. 4 (5)
GROCERY STORES, WITHOUT FRESH MEAT					
Stores analyzing merchandise line sales, total	4,007,742	100.0	152,185	...	...
Canned goods, other groceries	3,183,911	79.5	152,185	4,007,742	79.5
Meats (exc. fresh), sea food, poultry	134,879	3.4	33,818	865,450	15.6
Fresh fruits, vegetables	345,640	8.6	59,485	2,141,575	16.1
Gasoline, lubricants, fuel oil	133,066	3.3	38,697	628,831	21.2
Other nongrocery items	210,246	5.2	...	...	...
Stores not analyzing merchandise line sales	68,744	...	2,100	...	...
GROCERY STORES, WITH FRESH MEAT					
Stores analyzing merchandise line sales, total	20,641,824	100.0	218,840	...	...
Canned goods, other groceries	11,597,473	56.2	218,840	20,641,824	56.2
Meats (all kinds), sea food, poultry	6,026,729	29.2	218,840	20,641,824	29.2
Fresh fruits, vegetables	2,381,844	11.5	185,003	19,037,932	12.5
Gasoline, lubricants, fuel oil	169,910	0.8	31,034	1,249,974	13.6
Other nongrocery items	465,868	2.3	...	...	...
Stores not analyzing merchandise line sales	308,174	...	4,815	...	...
MEAT MARKETS					
Stores analyzing merchandise line sales, total	1,629,901	100.0	23,832	...	...
Meats (all kinds), sea food, poultry	1,574,465	96.6	23,832	1,629,901	96.6
Fresh fruits, vegetables	14,059	0.9	1,579	127,526	11.0
Canned goods, other groceries	33,916	2.1	4,359	330,480	10.3
Other sales	7,461	0.4	...	...	...
Stores not analyzing merchandise line sales	6,117	...	53	...	...
FISH (SEA FOOD) MARKETS					
Stores analyzing merchandise line sales, total	148,204	100.0	5,509	...	...
Sea food, meats (all kinds), poultry	142,736	96.3	5,509	148,204	96.3
Fresh fruits, vegetables	1,657	1.1	217	7,259	22.8
Canned goods, other groceries	2,669	1.8	554	14,327	18.6
Other sales	1,142	0.8	...	...	...
Stores not analyzing merchandise line sales	2,257	...	69	...	...
FRUIT STORES, VEGETABLE MARKETS					
Stores analyzing merchandise line sales, total	396,523	100.0	15,669	...	...
Fresh fruits, vegetables	367,332	92.7	15,669	396,523	92.7
Meats (all kinds), sea food, poultry	6,817	1.7	699	27,782	24.5
Canned goods, other groceries	17,529	4.4	2,436	78,380	22.4
Other sales	4,845	1.2	...	...	...
Stores not analyzing merchandise line sales	3,765	...	84	...	...
CANDY, NUT, CONFECTIONERY STORES					
Stores analyzing merchandise line sales, total	649,094	100.0	32,063	...	...
Candy, nuts, confectionery	421,060	64.8	32,063	649,094	64.8
Cigars, cigarettes, tobacco	67,859	10.4	18,787	293,460	23.1
Bakery products	11,445	1.8	3,879	74,975	15.3
Fountain and meals	79,714	12.3	12,003	251,529	31.7
Canned goods, beverages, other foods	24,420	3.8	8,569	125,895	19.4
Other sales	44,596	6.9	...	...	...
Stores not analyzing merchandise line sales	16,321	...	810	...	...
DAIRY PRODUCTS STORES					
Stores analyzing merchandise line sales, total	312,341	100.0	5,596	...	...
Milk and cream	81,953	26.2	3,337	211,824	38.7
Ice cream, frozen custard, etc	94,493	30.3	3,986	205,218	46.0
Other dairy products	61,525	19.7	2,980	214,413	28.7
Eggs	22,190	7.1	2,409	171,714	12.9
Groceries, other foods	34,695	11.1	2,454	141,082	24.6
Other sales	17,485	5.6	...	...	...
Stores not analyzing merchandise line sales	16,607	...	202	...	...
MILK DEALERS					
Stores analyzing merchandise line sales, total	1,552,222	100.0	5,800	...	...
Milk and cream	1,351,191	87.1	5,800	1,552,222	87.1
Ice cream, frozen custard, etc	49,155	3.2	1,247	475,084	10.3
Other dairy products	99,336	6.4	2,052	1,106,156	9.0
Eggs	27,023	1.7	1,065	672,760	4.0
Other sales	25,517	1.6	...	...	...
Stores not analyzing merchandise line sales	29,378	...	134	...	...
BAKERY PRODUCTS STORES					
Stores analyzing merchandise line sales, total	743,164	100.0	19,985	...	...
Bakery products	712,426	95.9	19,985	743,164	95.9
Candy, nuts, confectionery	4,936	0.7	1,580	62,792	7.9
Fountain and meals	9,753	1.3	891	36,404	26.8
Canned goods, beverages, other foods	6,894	0.9	1,207	40,309	17.1
Other sales	9,155	1.2	...	...	...
Stores not analyzing merchandise line sales	6,368	...	152	...	...
EGG, POULTRY DEALERS					
Stores analyzing merchandise line sales, total	261,078	100.0	5,428	...	...
Eggs	45,434	17.4	2,987	152,012	29.9
Poultry	196,569	75.3	5,110	245,359	80.1
Groceries, other foods	3,769	1.4	466	21,852	17.2
Other sales	15,306	5.9	...	...	...
Stores not analyzing merchandise line sales	5,578	...	154	...	...
DELICATESSEN STORES					
Stores analyzing merchandise line sales, total	310,033	100.0	8,633	...	...
Delicatessen meats, salads, etc	150,358	48.5	8,633	310,033	48.5
Candy, nuts, confectionery	8,100	2.6	4,030	116,408	7.0
Cigars, cigarettes, tobacco	12,162	3.9	4,212	131,988	9.2
Bakery products	16,734	5.4	4,338	143,453	11.7
Fountain and meals	13,654	4.4	1,450	57,667	23.7
Canned goods, beverages, other foods	94,744	30.6	6,846	241,080	39.3
Other sales	14,281	4.6	...	...	...
Stores not analyzing merchandise line sales	7,533	...	212	...	...

Figure A.3: Geographic distribution of counties by TV group



*The Pre-war group, which had access to television programming before the end of World War II and consists of only six market areas, is excluded from the analysis.

Figure A.4: Counties from Baum-Snow (2007)'s highway dataset

(a) 1947 highway plan reproduced from Baum-Snow (2007)

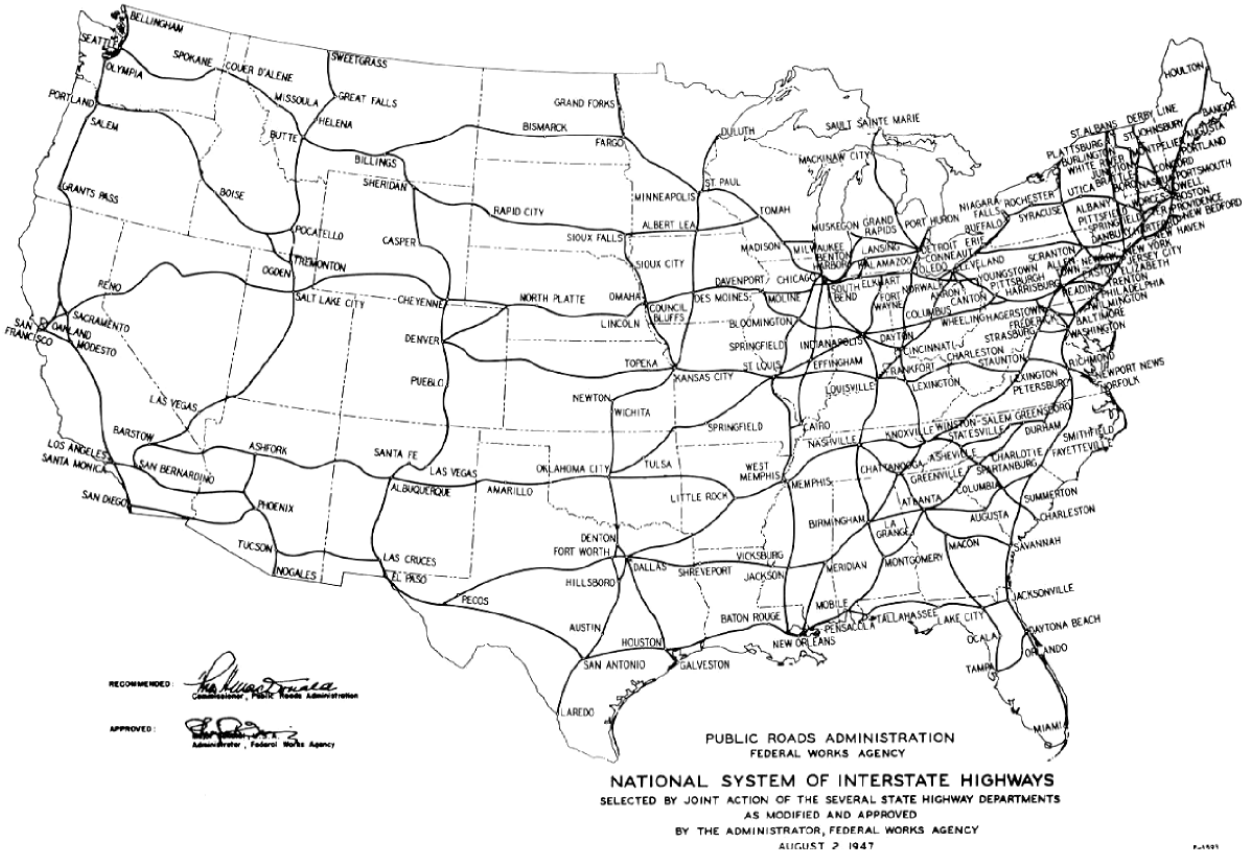


FIGURE I
The Projected System of Interstate Highways in 1947

(b) Counties in dataset

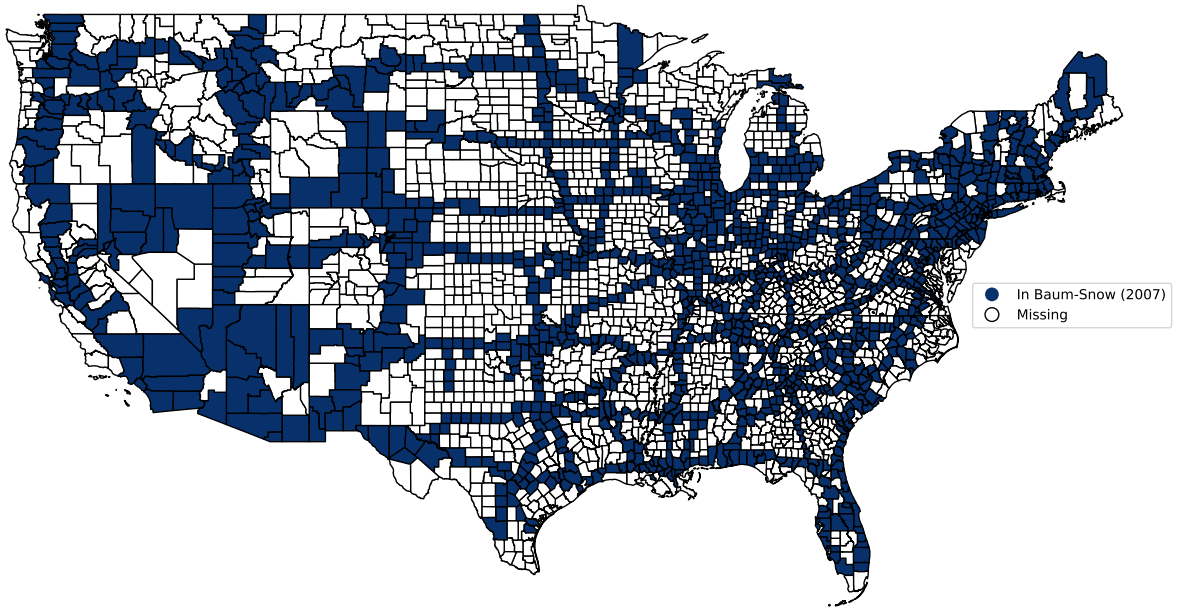
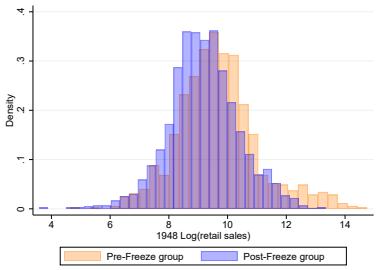
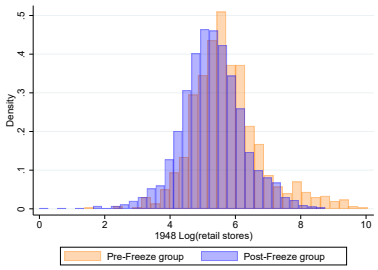


Figure A.5: Distributions of demographics for Pre- and Post-Freeze groups

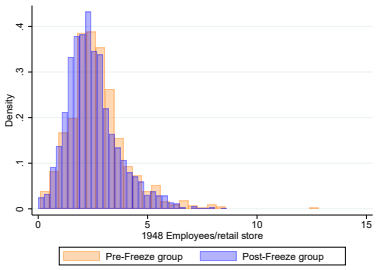
(a) Log 1948 retail sales



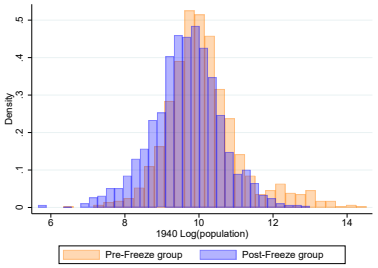
(b) Log 1948 retail stores



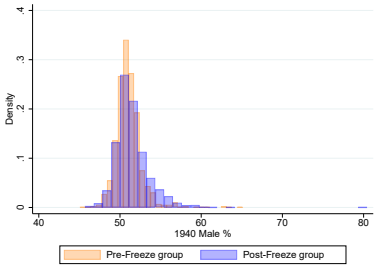
(c) Log 1948 employees/retail store



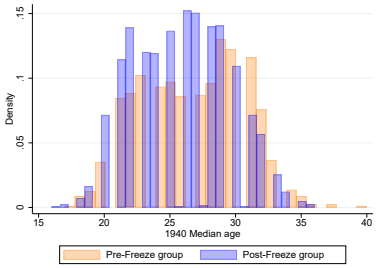
(d) Log 1940 population



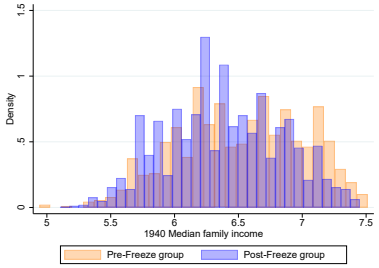
(e) 1940 Male %



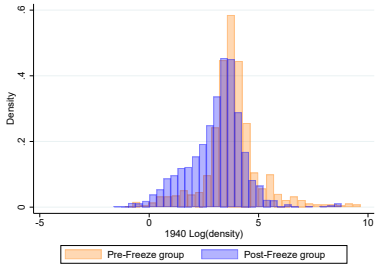
(f) 1940 Median age



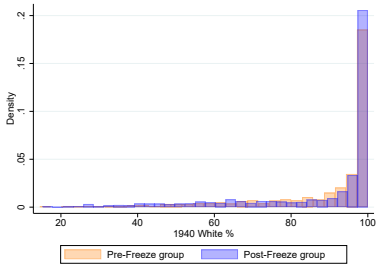
(g) Log 1940 median family income



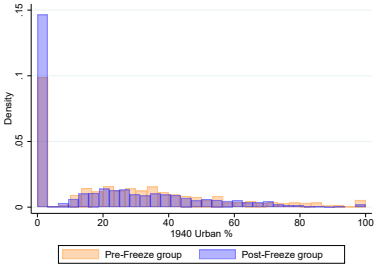
(h) Log 1940 density



(i) 1940 White %



(j) 1940 Urban %



(k) 1940 High school educated %

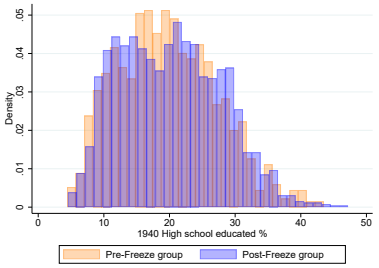
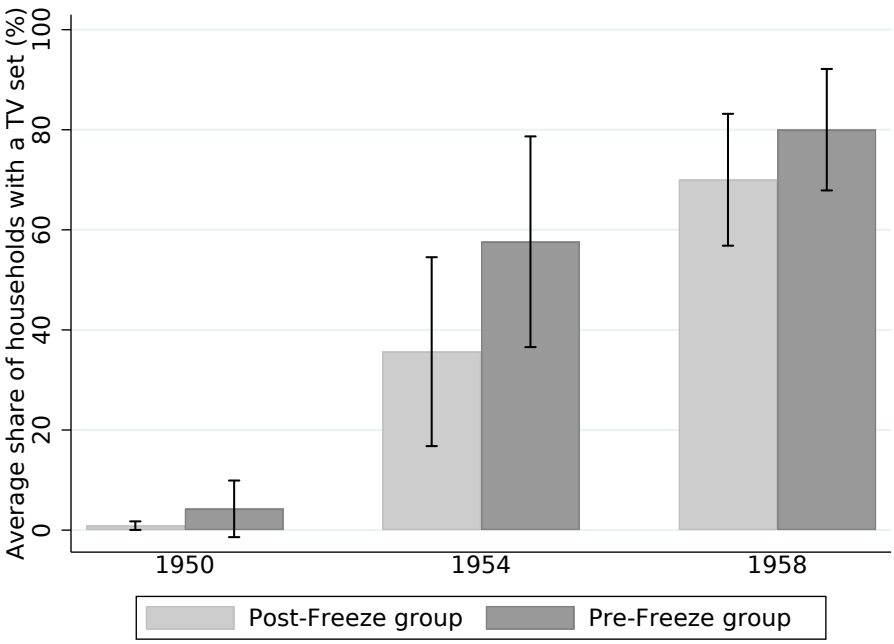


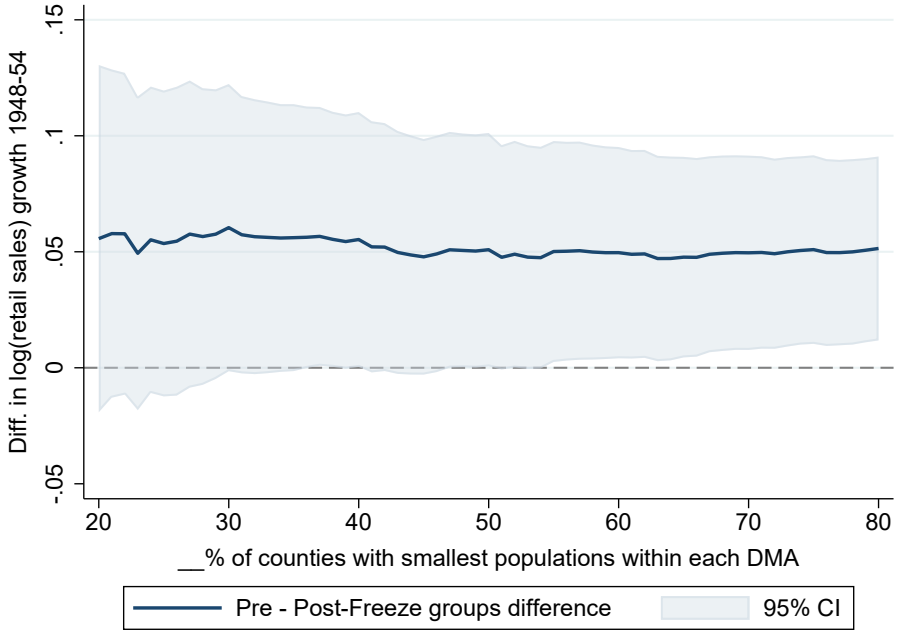
Figure A.6: TV ownership by year



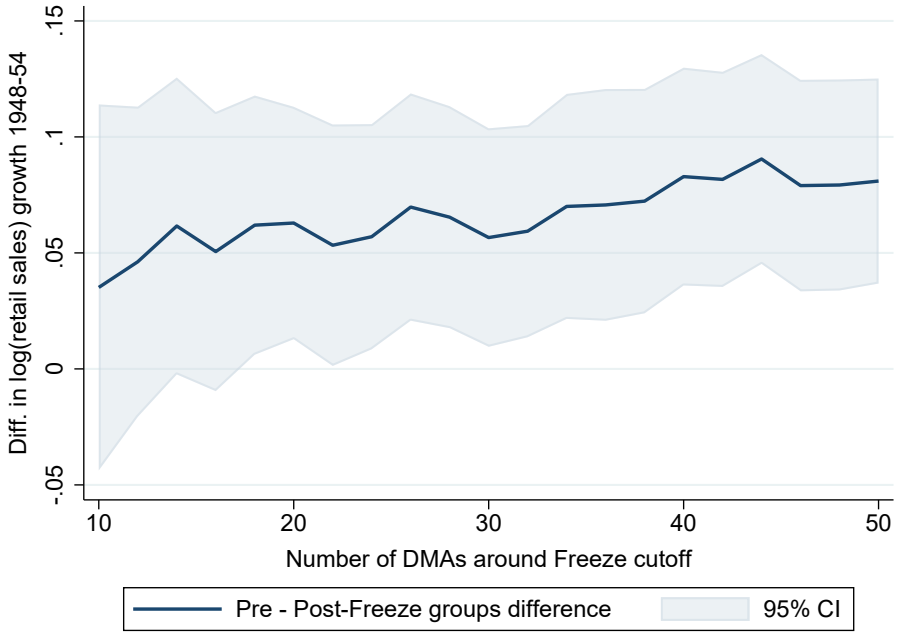
Error bars indicate the range of ±1 standard deviation. 1950 is the earliest year for which TV set ownership data are available.

Figure A.7: Robustness checks

(a) Smallest counties within each DMA



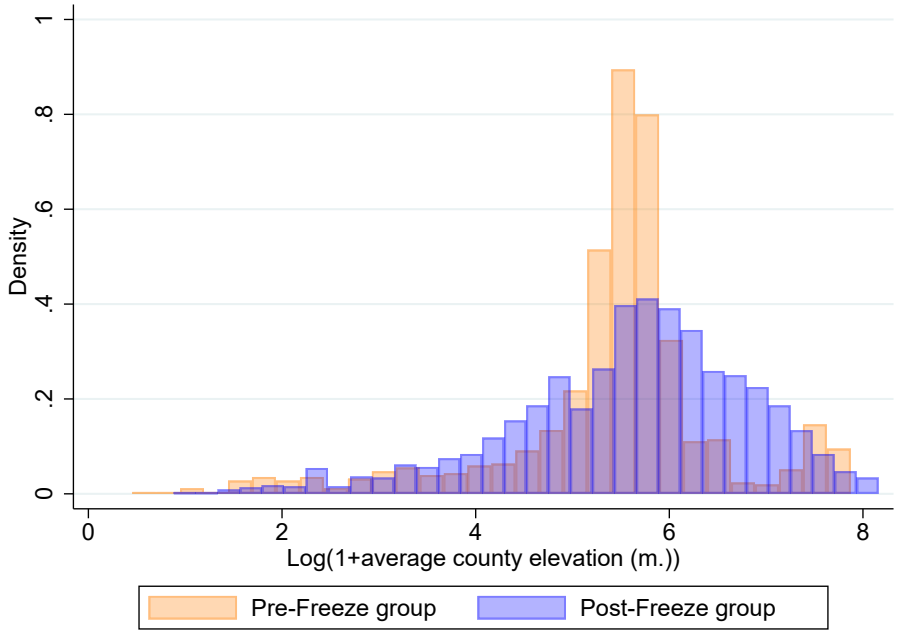
(b) Robustness around Freeze discontinuity



Figures A.7a and A.7b show the difference in average log retail sales growth over 1948 to 1954 between subsamples of the Pre- and Post-Freeze groups. Figure A.7a compares the smallest p% of counties by population within each DMA in both the Pre- and Post-Freeze groups, where p is varied along the horizontal axis. Figure A.7b compares the counties in the last N DMAs in the Pre-Freeze group and counties in the first N DMAs in the Post-Freeze group to gain TV access, where N is varied along the horizontal axis. The shaded regions indicates the 95% confidence intervals with standard errors clustered by DMA.

Figure A.8: Distribution of topographical features

(a) Elevation



(b) Area

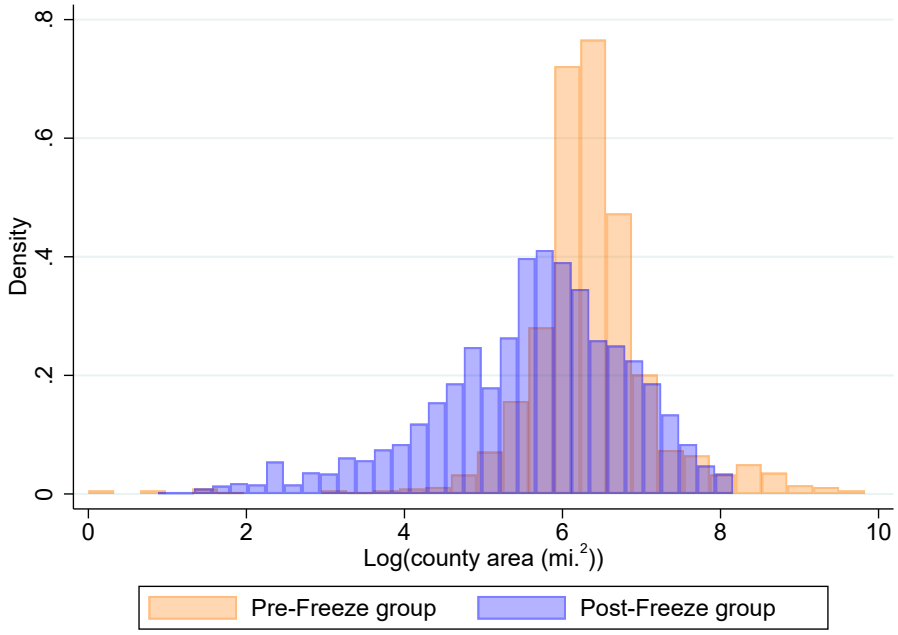
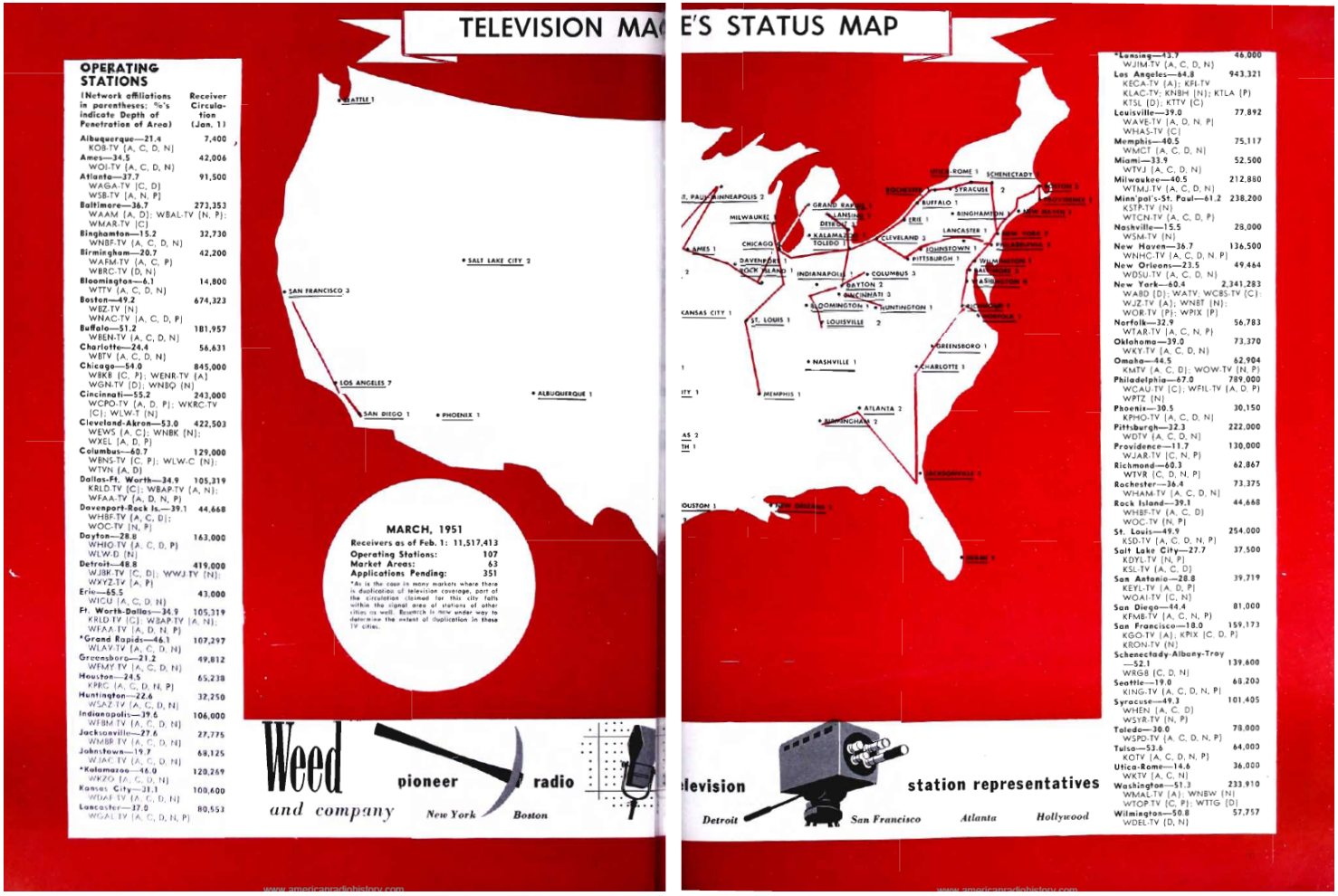


Figure A.9: Status map of main broadcasting companies and affiliate stations



Source: *TV Magazine* 1951 March issue. Abbrevations: A - American Broadcasting Company (ABC); C - Columbia Broadcsting System (CBS); N - National Broadcasting Company (NBC); D - DuMont Television Network (dissolved 1956); P - Paramount Television Network (dissolved 1956).

Figure A.10: Leading TV advertisers 1951-54

THE TOP TEN NETWORK ADVERTISERS

Estimated gross time and production expenditures

	1954		1953		1952		1951	
Company	**Rank**	**Expenditure**	**Rank**	**Expenditure**	**Rank**	**Expenditure**	**Rank**	**Expenditure**
Procter & Gamble	1	$31,025,000	1	$21,300,000	1	$20,574,000	1	$12,205,000
Colgate-Palmolive-Peet	2	19,485,000	2	16,260,000	2	12,464,000	4	6,831,000
General Motors	3	18,480,000	4	12,880,000	7	9,044,000		
R. J. Reynolds	4	17,810,000	3	13,275,500	4	11,481,000	3	7,586,000
Gillette	5	17,190,000	7	11,000,000	9	7,286,000		
General Foods	6	16,450,000	6	11,432,000	3	11,870,500	2	12,111,000
American Tobacco	7	15,700,000	5	12,246,500	6	10,693,000	6	6,371,000
General Electric	8	13,800,000	10	8,958,000				
Chrysler	9	12,730,000						
General Mills	10	12,545,000					10	4,087,000
Lever Brothers			9	10,763,000	5	11,130,500	8	4,974,000
P. Lorillard							9	4,784,000
Ford Motor					10	7,091,000	5	6,752,000
Liggett & Myers			8	10,777,000	8	7,402,500	7	5,942,000

Source: *TV Magazine* 1955 March issue.

Figure A.11: TV advertising revenue data from *Advertising Age*

Advertising Expenditures by Industries (January-September 1952 and 1951)

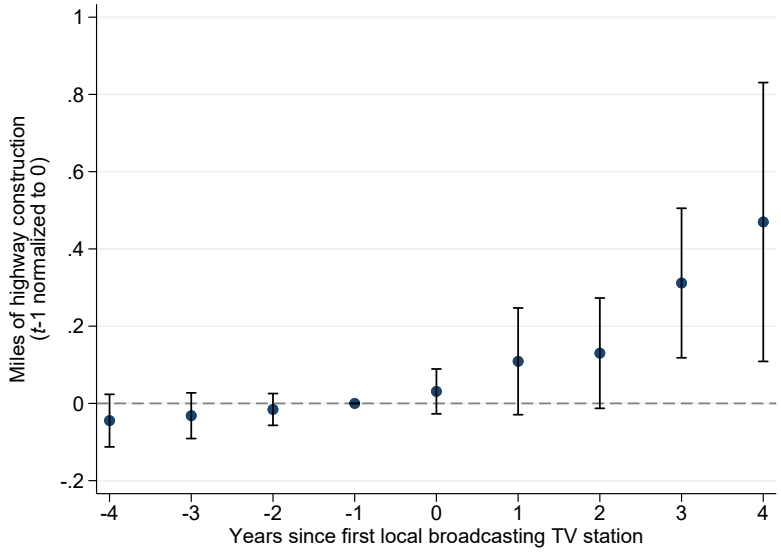
Industry Class	Total		General And Farm Magazines		Newspaper Sections		Network Radio		Network Television	
	1952	1951	1952	1951	1952	1951	1952	1951	1952	1951
Food and Food Products	$105,230,640	$102,261,252	$ 43,505,813	$ 40,291,011	$ 9,253,678	$12,178,319	$26,933,657	$31,876,460	$25,537,492	$17,915,462
Toiletries and Toilet Goods	72,677,317	66,814,840	29,541,098	28,378,558	9,324,938	8,464,949	16,646,080	19,700,414	17,165,201	10,270,919
Automotive	50,847,094	43,089,746	35,536,426	30,533,818	1,811,538	1,841,881	2,819,104	3,113,596	10,680,026	7,600,451
Smoking Materials	44,337,420	39,165,580	10,335,998	10,126,606	1,571,217	1,874,071	12,069,433	15,120,444	20,360,772	12,044,459
Household Soaps and Cleansers	37,287,631	31,396,252	7,325,817	7,855,011	2,304,582	3,251,479	13,532,568	13,340,141	14,124,664	6,939,621
Household Equipment	37,196,464	33,128,177	24,547,030	23,952,964	1,872,692	1,807,845	3,572,759	2,111,100	7,203,983	5,256,268
Apparel and Accessories	36,431,660	36,781,676	32,000,765	32,047,148	2,003,588	2,112,173	235,681	519,588	2,191,626	2,102,767
Drugs and Remedies	35,590,656	32,499,601	12,190,475	10,493,298	4,268,685	3,735,297	15,286,674	16,777,832	3,844,822	1,493,174
Industrial Materials	34,016,830	27,123,941	29,437,380	23,364,120	328,079	113,925	1,422,867	1,512,697	2,828,504	2,133,199
Beer, Wine and Liquor	26,092,728	27,528,649	19,927,631	20,615,632	267,036	262,201	1,701,275	2,636,355	4,196,786	4,014,461
Building Materials	23,209,283	20,476,517	21,871,758	19,097,345	194,470	313,772	786,786	1,053,510	356,269	11,890
Household Furnishings	21,573,501	22,789,381	18,155,006	18,801,902	856,070	769,571	974,662	720,659	1,587,763	2,497,249
Transportation, Hotels, Industrial & Agricultural Development	16,589,346	14,593,404	15,606,007	13,407,491	301,655	272,189	681,684	913,724	.....	.....
Fuels and Lubricants	16,522,072	14,921,137	9,372,355	8,307,408	525,867	460,953	3,988,787	4,305,544	2,635,063	1,847,232
Miscellaneous	15,000,828	13,515,971	8,960,150	8,051,151	915,190	796,221	3,808,615	3,776,816	1,316,873	891,783
Confectionery and Soft Drinks	14,475,540	12,626,990	5,918,230	5,365,926	616,269	456,285	4,300,755	4,618,585	3,640,286	2,186,194
Consumer Services	13,064,384	13,168,893	10,320,915	9,906,776	849,110	852,817	1,637,673	2,072,698	256,686	336,562
Office Equipment and Writing Supplies	11,630,573	9,235,535	9,704,169	8,000,632	362,836	367,873	427,113	453,900	1,136,455	413,130
Insurance	11,184,702	9,340,542	7,672,511	6,572,240	567,077	91,274	2,602,376	2,387,678	342,738	289,350
Radios, TV Sets, Phonographs, Musical Instruments	10,564,887	13,369,083	5,806,860	6,669,069	513,498	959,539	1,185,756	2,143,839	3,058,773	3,596,536
Retail and Direct by Mail	10,545,823	9,895,595	8,477,780	7,479,326	1,130,763	853,349	21,438	28,286	915,842	1,524,634
Jewelry, Optical Goods, Cameras	8,743,996	7,734,983	6,271,328	5,207,456	282,489	284,166	587,565	873,341	1,602,614	1,370,020
Publishing and Media	6,758,630	6,731,184	3,868,969	3,494,956	1,761,516	2,323,263	492,220	268,707	635,925	644,258
Agriculture and Farming	6,337,326	4,509,797	5,755,966	4,019,537	43,406	5,125	537,954	485,135	.....	.....
Sporting Goods and Toys	4,634,388	4,209,017	4,575,994	4,186,891	58,394	22,126	.....	.....	.....	.....
Entertainment	3,912,474	3,907,639	3,286,995	3,470,666	617,889	431,250	6,090	5,723	1,500	.....
Horticulture	3,640,140	3,149,630	3,244,718	2,823,889	273,129	237,847	109,923	87,894	12,370	.....
Aviation	2,547,455	1,238,741	2,547,455	1,224,241	.....	14,500	.....	.....	.....	.....
Political	971,590	.....	.....	.....	.....	.....	506,289	.....	465,301	.....
Total	$681,615,378	$625,203,713	$395,765,599	$363,745,068	$42,875,661	$45,174,360	$116,875,784	$130,904,666	$126,098,334	$85,379,619

*Source: Publishers Information Bureau Records, Compiled by Leading National Advertisers, Inc.

*National political convention programs not included

Figure A.12: TV access and highway construction: TWFE and problematic weights

(a) "Problematic" two-way fixed-effects (TWFE) specification



(b) Goodman-Bacon (2021) decomposition of weights in TWFE difference-in-differences

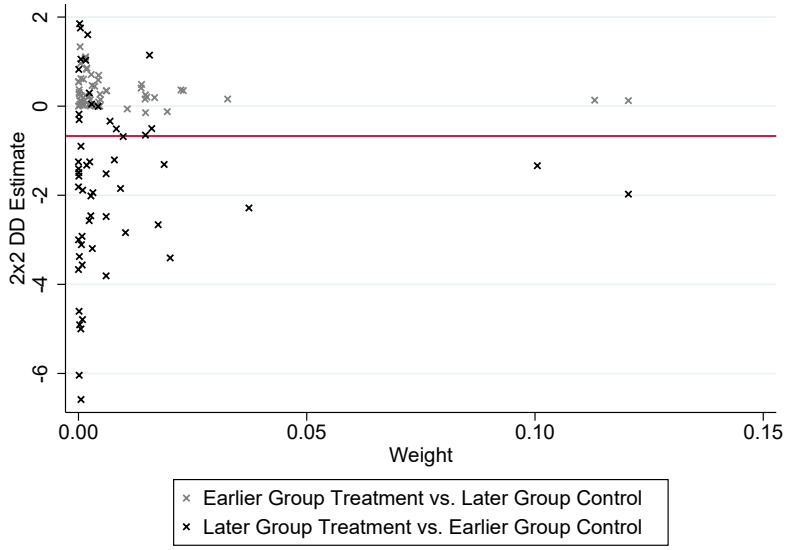
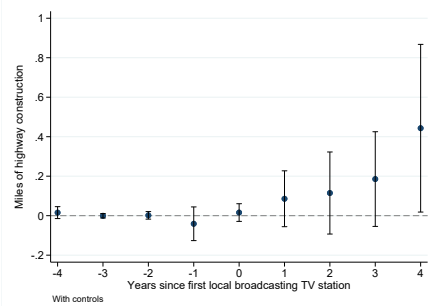


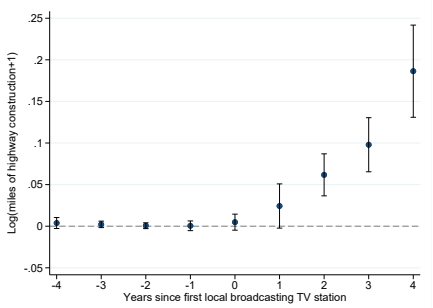
Figure A.12a plots the event study of highway construction around the event of gaining TV acess relative to the year before the event using the two-way fixed-effects (TWFE) estimator with year and county fixed effects. Years up to 1956 are included and counties that gained TV after 1956 are used as the control group.
Figure A.12b shows the Goodman-Bacon (2021) decomposition of the weights in the simple TWFE difference-in-differences estimate of the effect of TV access on highway construction. The figure is produced by the `bacondeomp` Stata package written by Goodman-Bacon, Goldring, and Nichols.

Figure A.12: Event study of TV access and highway construction (additional specifications)

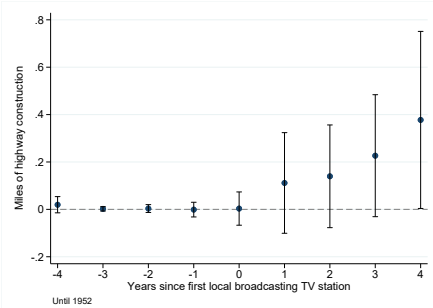
(c) Demographic controls



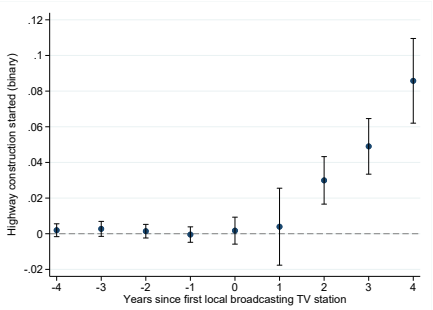
(d) Outcome: log highways



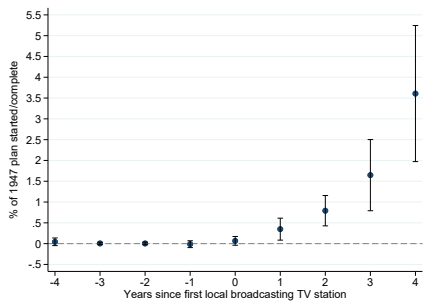
(e) Until 1952



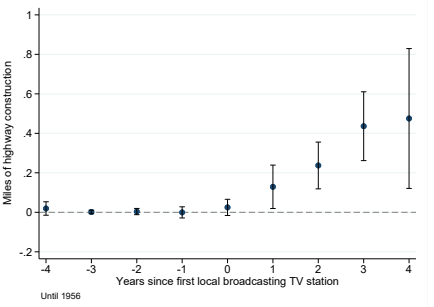
(f) Outcome: started construction



(g) Outcome: fraction of plan



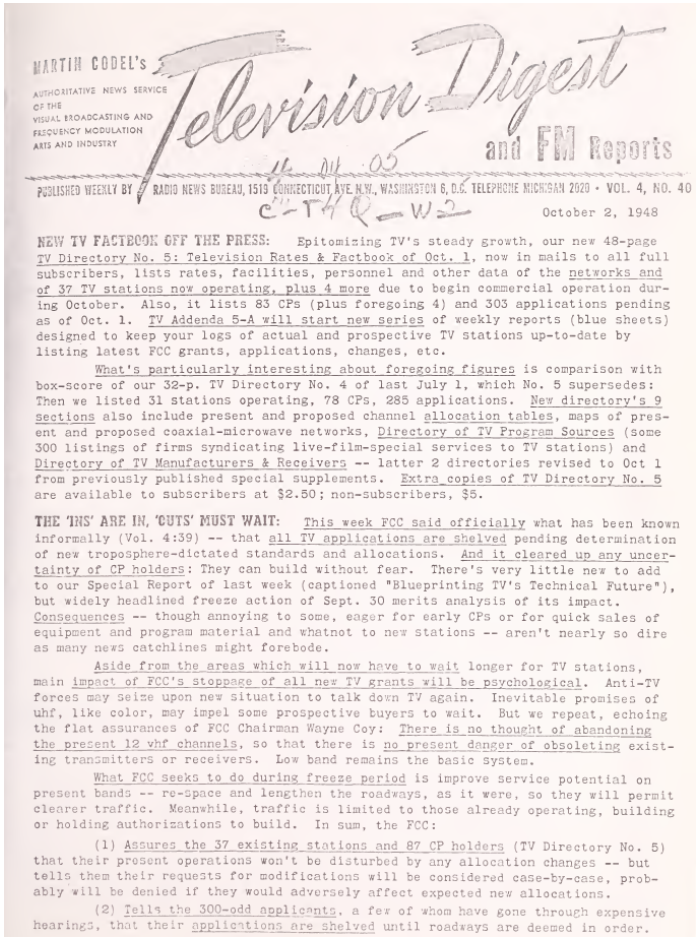
(h) Until 1956



Figures A.12c-A.12h present alternate specifications of the event study of highway construction around the event of gaining TV access. All estimates are from the Callaway and Sant'Anna (2020) estimator. Coefficients for the years after the event are estimated relative to the year before the event, and the coefficients for the years before the event are estimated relative to the preceding year. Standard errors for the 95% CIs are clustered by DMA. Year and county fixed effects are included in every specification. Figure A.12c incorporates the effect of base-year demographics (log population, median age, male %, log density, log median income, white %, urban %, and high-school graduate %, all from 1940) in both the outcome regression model and the propensity score estimator (see Callaway and Sant'Anna, 2020, and Sant'Anna and Zhou, 2020, for details). Figure A.12d takes the natural log of highway construction miles as the outcome, with a constant of 1 added inside the logarithm to avoid dropping observations with zero miles of construction. Figure A.12e only includes years until 1952 when the FCC Freeze ended to isolate the variation in TV access induced by the Freeze. The outcome in Figure A.12f is a binary variable for whether the county has started highway construction; the event study estimates the change in the probability of starting highway construction. The outcome in Figure A.12g is the percent of highway miles in the 1947 plan that have started or finished construction. Figure A.12h only includes years until 1956 when the Highway Revenue Act was passed and federal funding for highway construction became widely available.

Figure A.13: *Television Digest* example excerpts

(a) Front page (1948, Q4)



MARTIN CODEL'S

AUTHORITATIVE NEWS SERVICE OF THE VISUAL BROADCASTING AND FREQUENCY MODULATION ARTS AND INDUSTRY

Television Digest and FM Reports

PUBLISHED WEEKLY BY RADIO NEWS BUREAU, 1519 CONNECTICUT AVE. N.W., WASHINGTON 6, D.C. TELEPHONE MICHIGAN 2020 • VOL. 4, NO. 40

October 2, 1948

NEW TV FACTBOOK OFF THE PRESS: Epitomizing TV's steady growth, our new 48-page TV Directory No. 5: Television Rates & Factbook of Oct. 1, now in mails to all full subscribers, lists rates, facilities, personnel and other data of the networks and of 37 TV stations now operating, plus 4 more due to begin commercial operation during October. Also, it lists 83 CPs (plus foregoing 4) and 303 applications pending as of Oct. 1. TV Addenda 5-A will start new series of weekly reports (blue sheets) designed to keep your logs of actual and prospective TV stations up-to-date by listing latest FCC grants, applications, changes, etc.

What's particularly interesting about foregoing figures is comparison with box-score of our 32-p. TV Directory No. 4 of last July 1, which No. 5 supersedes: Then we listed 31 stations operating, 78 CPs, 285 applications. New directory's 9 sections also include present and proposed channel allocation tables, maps of present and proposed coaxial-microwave networks, Directory of TV Program Sources (some 300 listings of firms syndicating live-film-special services to TV stations) and Directory of TV Manufacturers & Receivers -- latter 2 directories revised to Oct 1 from previously published special supplements. Extra copies of TV Directory No. 5 are available to subscribers at $2.50; non-subscribers, $5.

THE 'INS' ARE IN, 'OUTS' MUST WAIT: This week FCC said officially what has been known informally (Vol. 4:39) -- that all TV applications are shelved pending determination of new troposphere-dictated standards and allocations. And it cleared up any uncertainty of CP holders: They can build without fear. There's very little new to add to our Special Report of last week (captioned "Blueprinting TV's Technical Future"), but widely headlined freeze action of Sept. 30 merits analysis of its impact. Consequences -- though annoying to some, eager for early CPs or for quick sales of equipment and program material and whatnot to new stations -- aren't nearly so dire as many news catchlines might forebode.

Aside from the areas which will now have to wait longer for TV stations, main impact of FCC's stoppage of all new TV grants will be psychological. Anti-TV forces may seize upon new situation to talk down TV again. Inevitable promises of uhf, like color, may impel some prospective buyers to wait. But we repeat, echoing the flat assurances of FCC Chairman Wayne Coy: There is no thought of abandoning the present 12 vhf channels, so that there is no present danger of obsoleting existing transmitters or receivers. Low band remains the basic system.

What FCC seeks to do during freeze period is improve service potential on present bands -- re-space and lengthen the roadways, as it were, so they will permit clearer traffic. Meanwhile, traffic is limited to those already operating, building or holding authorizations to build. In sum, the FCC:

(1) Assures the 37 existing stations and 87 CP holders (TV Directory No. 5) that their present operations won't be disturbed by any allocation changes -- but tells them their requests for modifications will be considered case-by-case, probably will be denied if they would adversely affect expected new allocations.

(2) Tells the 300-odd applicants, a few of whom have gone through expensive hearings, that their applications are shelved until roadways are deemed in order.

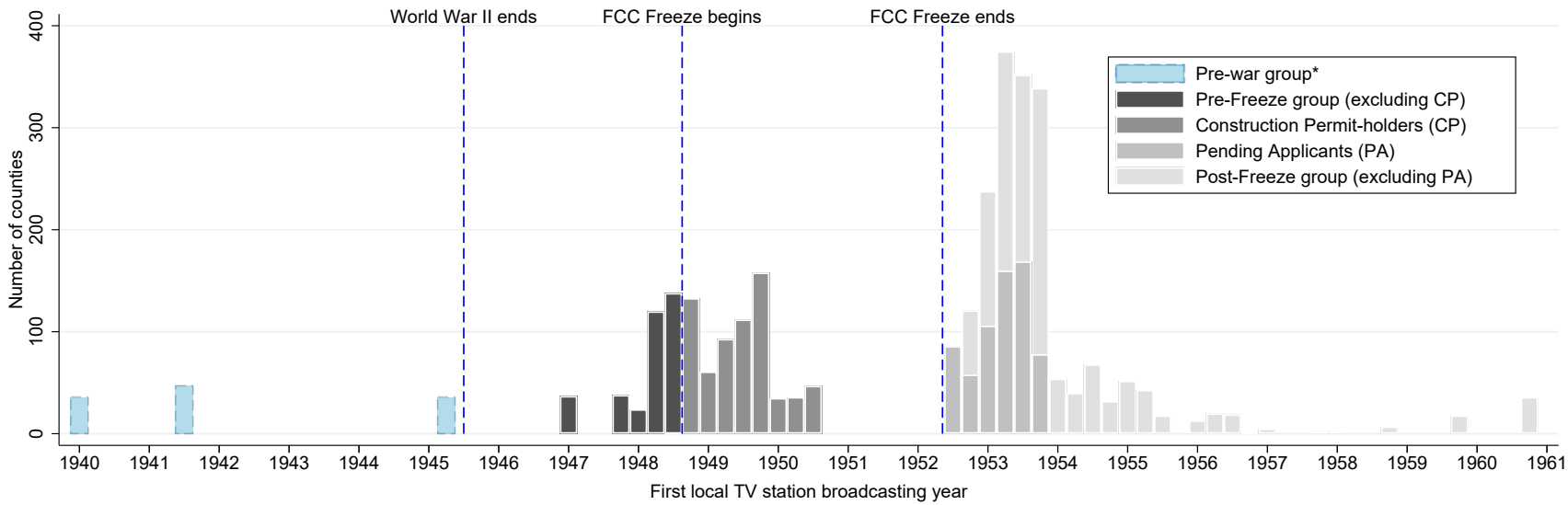
(b) New applicants record

BING WANTS A HUNK OF THAT TV, TOO: Eight new applications for TV were filed this week, probably prepared before FCC freeze order was made official. Bing Crosby filed for Channel No. 7 in Spokane, his home town, No. 9 in Tacoma, No. 6 in Yakima. Applications stated Bing proposed to spend $592,011 on combined construction, $340,000 per year on operation, listed his net worth as $1,649,117, showed 1947 income after taxes as $269,000, disclosed he owns 3,885 out of 200,000 shares in Dick Richards' KMPC, Hollywood.

Los Angeles used car dealer Leland Holzer, advertising as "Honest John," plopped in 4 applications: For Tucson, Ariz., No. 6; Boise, Ida., No. 6; Albuquerque, N.M., No. 5; Wichita Falls, Tex., No. 10. Previously an applicant for San Francisco and San Diego, Holzer withdrew on eve of those hearings.

Eighth applicant of week (310th now pending) was Clyde Weatherby, for No. 7 in El Paso, Tex.; he's owner of daytime local KCLW, Hamilton, Tex. Two applications were withdrawn: From San Jose, Cal., by Video Broadcasting Co., Hollywood group holding CP for Portland, Ore., still applicant for San Diego. From Harlingen, Tex., by KGBS interests. [For details about foregoing applicants and withdrawees, see TV Directory No. 5 and Addenda 5-A.]

Figure A.14: Distribution of first local TV station broadcasting date (with CP and PA)



This histogram plots the distribution of the first quarter that a local TV station began broadcasting for each U.S. county. The *Pre-Freeze* group had a local station that began broadcasting after WWII but before/during the FCC Freeze, which lasted from late-1948 to mid-1952. Within the Pre-Freeze group, the subset of *Construction Permit-holders* constructed their first local TV station *during* the Freeze; at the start of the Freeze, those stations only had permits to do so. The first local TV stations for the *Post-Freeze* group were built *after* the Freeze. Within the Post-Freeze group, the subset of *Pending Applicants*, like the Construction Permit-holders, had filed an application before or near the onset of the Freeze. However, their applications were "frozen" until the Freeze ended.
*The *Pre-war* group, which had access to television programming before the end of World War II and consists of only six market areas, is excluded from the analysis.

Table A.1: Summary statistics for other groups

	Pre-war group	Construction Permit-holders	Pending Applicants	
	(1)	(2)	(3)	(2)-(3)
1948 Retail sales ($1000s)	270667.17	46473.42	28074.18	18399.24
	[747521.69]	[118581.68]	[53458.17]	(7059.11)
1948 Log(retail sales)	10.99	9.69	9.41	0.27
	[1.74]	[1.30]	[1.24]	(0.17)
1948 Retail stores	3428.58	618.39	406.36	212.03
	[8177.95]	[1241.26]	[614.11]	(86.23)
1948 Employees per retail store	3.08	2.73	2.67	0.07
	[1.37]	[1.30]	[1.12]	(0.13)
Demographic controls				
1940 Population	232351.40	47286.07	30248.27	17037.80
	[577314.65]	[93371.12]	[43550.74]	(6257.82)
1940 Male %	50.80	50.97	51.47	-0.50
	[1.74]	[1.64]	[1.94]	(0.31)
1940 Median age	29.59	26.61	26.70	-0.09
	[3.20]	[4.06]	[3.39]	(0.81)
1940 Median family income ($)	1206.25	729.34	721.24	8.10
	[441.99]	[337.81]	[310.00]	(55.78)
1940 Persons per sq. mile	2137.53	167.72	54.60	113.12
	[9276.13]	[852.36]	[228.78]	(67.42)
1940 White %	92.76	89.10	89.58	-0.48
	[9.64]	[16.31]	[17.53]	(3.94)
1940 Urban %	44.40	27.12	22.19	4.94
	[34.03]	[25.78]	[23.46]	(2.37)
1940 High school educated %	22.54	19.67	22.55	-2.89
	[7.74]	[7.64]	[7.48]	(1.70)
	121	678	657	
	6	38	38	

This table shows the average 1948 retail outcomes and 1940 demographic characteristics of counties in the Pre-war, Construction Permit-holders, and Pending Applicants groups (see notes in Figure A.14 for definitions of these groups). Retail sales and median family income are unadjusted for inflation. Standard deviations are shown in brackets. The last column shows the average difference between the Construction Permit-holders and Pending Applicants with standard errors clustered by DMA (Designated Market Area) in parentheses.

Table A.2: Access to TV and retail sales (additional specifications)

	TV years	Dependent variable		Subsamples			Conley SEs	
	(1)	(2) Sales/household	(3) Sales/income	(4) Dist. hypo.	(5) CP vs. PA	(6) Highway	(7) 350mi	(8) 500mi
Years since TV access	0.0097							
	(0.0048)							
Differences in Pre-Freeze group growth relative to Post-Freeze group over each period								
Pre-Freeze×Yeartrend$_{30\text{-}35}$		24.2		0.011		-0.012	-0.00086	-0.00086
		(22.9)		(0.031)		(0.021)	(0.023)	(0.025)
Pre-Freeze×Yeartrend$_{35\text{-}39}$		11.4		0.032		0.00029	0.0033	0.0033
		(18.1)		(0.019)		(0.017)	(0.025)	(0.028)
Pre-Freeze×Yeartrend$_{39\text{-}48}$		-35.5	0.024	0.0042		0.013	0.018	0.018
		(26.9)	(0.056)	(0.024)		(0.015)	(0.017)	(0.020)
Pre-Freeze×Yeartrend$_{48\text{-}54}$		61.9	0.041	0.047		0.030	0.032	0.032
		(27.6)	(0.020)	(0.023)		(0.016)	(0.016)	(0.018)
Pre-Freeze×Yeartrend$_{54\text{-}58}$		-26.0	-0.011	0.020		0.0030	0.0059	0.0059
		(18.1)	(0.0082)	(0.013)		(0.0099)	(0.028)	(0.032)
Differences in Construction Permit-holders (CP) growth relative to Pending Applicants group over each period								
CP-holders×Yeartrend$_{30\text{-}35}$					0.022			
					(0.030)			
CP-holders×Yeartrend$_{35\text{-}39}$					0.0063			
					(0.026)			
CP-holders×Yeartrend$_{39\text{-}48}$					0.0081			
					(0.022)			
CP-holders×Yeartrend$_{48\text{-}54}$					0.037			
					(0.022)			
CP-holders×Yeartrend$_{54\text{-}58}$					-0.0012			
					(0.011)			
Demographic controls	✓	✓	✓	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	17822	17822	11816	7853	7988	7499	17822	17822
Counties	2974	2974	2956	1310	1332	1250	2981	2981
Media markets (DMA)	199	199	199	67	76	187	199	199
R^2	0.83	0.81	0.77	0.83	0.84	0.86	0.83	0.83

This table shows various robustness checks. The dependent variable in all columns is log retail sales, except in Columns 2-3. Column 1 parametrically estimates the effect of TV and allows the effect to increase linearly in the number of years that a local broadcasting TV station has been on air. Columns 2-3 estimate the effect of TV on retail sales per household and on retail sales over income. The number of households between decennial census years is interpolated assuming a constant growth rate. Total county income is approximated as the number of households multiplied by the median family income, with both the number of households and the median family income interpolated between decennial census years assuming a constant growth rate. Since the median family income is only available from 1940, Column 3 excludes years before 1939. Columns 4-6 compare the relative trends in retail sales between subsamples of the Pre- and Post-Freeze groups. Column 4 tests the distraction hypothesis (i.e., that TV undermined retail sales in counties when they first gained TV access) and only includes the subsample of the Pre-Freeze counties that gained access before or in 1948 and the Post-Freeze counties that gained access at least half a year before 1954. Column 5 compares the Construction Permit-holders and Pending Applicants (see Online Appendix Section B for details). Column 6 uses only the subset of counties shown in Figure A.4b that have highway data available from Baum-Snow (2007). In Columns 1-6, standard errors are clustered by DMA. Columns 7-8 use Conley standard errors under varying distance thresholds. Standard errors are shown in parentheses.

Table A.3: Pre-trends in median occupational income score or family income

	Occupational score (1930-40)		Family income (1940-50)	
	(1)	(2)	(3)	(4)
Pre-Freeze group	0.90		101.3	
	(0.31)		(41.4)	
Difference-in-differences relative to Post-Freeze group				
Pre-Freeze×$\mathbf{1}\{t = 1940\}$	0.17	0.17		
	(0.22)	(0.22)		
Pre-Freeze×$\mathbf{1}\{t = 1950\}$			4.40	0.51
			(82.3)	(82.0)
$\mathbf{1}\{t = 1940\}$	2.20	2.20		
	(0.13)	(0.13)		
$\mathbf{1}\{t = 1950\}$			1565.8	1566.7
			(55.2)	(54.9)
Constant	15.1	15.4	655.4	691.0
	(0.14)	(0.053)	(20.5)	(20.6)
County FE		✓		✓
Observations	5945	5945	5893	5893
Counties	2974	2974	2976	2976
Media markets (DMA)	199	199	199	199
R^2	0.11	0.39	0.63	0.89

This table shows tests for differences in pre-trends between Pre- and Post-Freeze groups in IPUMS occupational income scores (OCCSCORE) from 1930 to 1940 and in median family income from 1940 to 1950. 1940 is the earliest data for which median family income data are available at the county level. IPUMS occupational scores range from 1 to 80 and represent the median total income (in hundreds of 1950 dollars) of all people with that same occupation in 1950. Estimates are from OLS regressions. Standard errors shown in parentheses are clustered by DMA.

Table A.4a: Topography summary statistics

	Pre-Freeze group		Post-Freeze group			
	Counties	DMA	Counties	DMA		
	(1)	(2)	(3)	(4)	(1)-(3)	(2)-(4)
Average elevation (m)	373.57	345.41	487.82	479.49	-114.25	-134.08
	[463.64]	[369.41]	[531.66]	[510.18]	(99.84)	(65.88)
Area (mi.^{2})	871.59	17037.09	1014.34	13512.98	-142.75	3524.11
	[1421.33]	[25606.48]	[1174.98]	[14860.04]	(177.02)	(3711.97)
N	1036	53	1945	146		

This table shows the average elevation and area for Pre- and Post-Freeze counties and DMAs. Standard deviations are provided in brackets. The last two columns show differences with standard errors clustered by DMA in parentheses.

Table A.4b: TV access, topography, and bordering counties

Dep. var.: Log(retail sales)	(1) Baseline	(2)	(3)	(4)
Differences in growth relative to Post-Freeze group over each period				
Pre-Freeze×Yeartrend$_{30\text{-}35}$	-0.001	0.000	-0.001	0.000
	(0.022)	(0.022)	(0.022)	(0.022)
Pre-Freeze×Yeartrend$_{35\text{-}39}$	0.003	0.003	0.003	0.003
	(0.016)	(0.015)	(0.016)	(0.015)
Pre-Freeze×Yeartrend$_{39\text{-}48}$	0.018	0.016	0.018	0.016
	(0.016)	(0.015)	(0.016)	(0.015)
Heterogeneity over Freeze period:				
Pre-Freeze×Yeartrend$_{48\text{-}54}$	0.032	0.030	0.056	0.056
	(0.016)	(0.015)	(0.018)	(0.018)
Border county×Yeartrend$_{48\text{-}54}$			-0.005	-0.004
			(0.018)	(0.017)
Pre-Freeze×Border × Yeartrend$_{48\text{-}54}$			-0.042	-0.045
			(0.025)	(0.024)
Pre-Freeze×Yeartrend$_{54\text{-}58}$	0.006	0.006	0.006	0.006
	(0.009)	(0.008)	(0.009)	(0.008)
Demographic controls	✓	✓	✓	✓
County FE	✓	✓	✓	✓
Topographic controls		✓		✓
Observations	17822	17666	17822	17666
Counties	2974	2948	2974	2948
Media markets (DMA)	199	199	199	199
R^2	0.83	0.84	0.83	0.84

This table compares the relative trends in log retail sales between the Pre- and Post-Freeze groups accounting for topographical features. Column 1 replicates the full specification from Column 4 of Table 2.3. (See notes in Table 2.3 for details on the variables and the demographic controls.) "Topographical controls" include the log(1+average county elevation) and the log(county area) interacted with an indicator for each year in the panel. "Border counties" are those in the Pre-Freeze (Post-Freeze) group that are adjacent to a county in the Post-Freeze (Pre-Freeze/Pre-war) group. Standard errors shown in parentheses are clustered by DMA.

Table A.5a: Top 10 evening series for 1950-54 seasons by audience size (Nielsen ratings)

Season	Rank	Series	Network	Genre	Rating	On air
1950-51	1	Texaco Star Theater	NBC	Comedy/Variety	61.6	1948-1967
1950-51	2	Fireside Theatre	NBC	Dramatic Anthology	52.6	1949-1963
1950-51	3	Philco TV Playhouse	NBC	Dramatic Anthology	45.3	1948-1955
1950-51	4	Your Show of Shows	NBC	Comedy/Variety	42.6	1950-1954
1950-51	5	The Colgate Comedy Hour	NBC	Variety	42	1950-1955
1950-51	6	Gillette Cavalcade of Sports	NBC	Sports (Boxing)	41.3	1946-1964
1950-51	7	The Lone Ranger	ABC	Western	41.2	1949-1957
1950-51	8	Arthur Godfrey's Talent Scouts	CBS	Talent	40.6	1948-1958
1950-51	9	Hopalong Cassidy	NBC	Western	39.9	1949-1951
1950-51	10	Mama	CBS	Comedy/Drama	39.7	1949-1956
1951-52	1	Arthur Godfrey's Talent Scouts	CBS	Talent	53.8	1948-1958
1951-52	2	Texaco Star Theater	NBC	Comedy/Variety	52	1948-1967
1951-52	3	I Love Lucy	CBS	Situation Comedy	50.9	1951-1957
1951-52	4	The Red Skelton Show	NBC	Comedy/Variety	50.2	1951-1971
1951-52	5	The Colgate Comedy Hour	NBC	Variety	45.3	1950-1955
1951-52	6	Arthur Godfrey and His Friends	CBS	Musical Variety	43.3	1949-1959
1951-52	7	Fireside Theatre	NBC	Dramatic Anthology	43.1	1949-1963
1951-52	8	Your Show of Shows	NBC	Comedy/Variety	43	1950-1954
1951-52	9	The Jack Benny Show	CBS	Comedy	42.8	1950-1977
1951-52	10	You Bet Your Life	NBC	Quiz	42.1	1950-1961
1952-53	1	I Love Lucy	CBS	Situation Comedy	67.3	1951-1957
1952-53	2	Arthur Godfrey's Talent Scouts	CBS	Talent	54.7	1948-1958
1952-53	3	Arthur Godfrey and His Friends	CBS	Musical Variety	47.1	1949-1959
1952-53	4	Dragnet	NBC	Police Drama	46.8	1952-1970
1952-53	5	Texaco Star Theater	NBC	Comedy/Variety	46.7	1948-1967
1952-53	6	The Buick Circus Hour	NBC	Musical Drama	46	1952-1953
1952-53	7	The Colgate Comedy Hour	NBC	Variety	44.3	1950-1955
1952-53	8	Gangbusters	NBC	Police Anthology	42.4	1952-1952
1952-53	9	You Bet Your Life	NBC	Quiz	41.6	1950-1961
1952-53	10	Fireside Theatre	NBC	Dramatic Anthology	40.6	1949-1963
1953-54	1	I Love Lucy	CBS	Situation Comedy	58.8	1951-1957
1953-54	2	Dragnet	NBC	Police Drama	53.2	1952-1970
1953-54	3	Arthur Godfrey's Talent Scouts	CBS	Talent	43.6	1948-1958
1953-54	4	You Bet Your Life	NBC	Quiz	43.6	1950-1961
1953-54	5	The Milton Berle Show	NBC	Comedy/Variety	40.2	1948-1967
1953-54	6	Arthur Godfrey and His Friends	CBS	Musical Variety	38.9	1949-1959
1953-54	7	Ford Theatre	NBC	Dramatic Anthology	38.8	1949-1957
1953-54	8	The Jackie Gleason Show	CBS	Comedy/Variety	38.1	1952-1970
1953-54	9	Fireside Theatre	NBC	Dramatic Anthology	36.4	1949-1963
1953-54	10	The Colgate Comedy Hour	NBC	Variety	36.2	1950-1955

Source: Brooks, T. and Marsh, E. (2003). *The Complete Directory to Prime Time Network and Cable TV 1946-Present* (8th ed.). NY: Ballantine Books.

TV seasons run from October to April. Shaded titles indicate that the series was also in the top ten for the preceding season.

Table A.5b: Top 10 evening series for 1954-58 seasons by audience size (Nielsen ratings)

Season	Rank	Series	Network	Genre	Rating	On air
1954-55	1	I Love Lucy	CBS	Situation Comedy	49.3	1951-1957
1954-55	2	The Jackie Gleason Show	CBS	Comedy/Variety	42.4	1952-1970
1954-55	3	Dragnet	NBC	Police Drama	42.1	1952-1970
1954-55	4	You Bet Your Life	NBC	Quiz	41	1950-1961
1954-55	5	The Ed Sullivan Show	CBS	Variety	39.6	1948-1971
1954-55	6	Disneyland	ABC	Anthology	39.1	1954-1990
1954-55	7	The Jack Benny Show	CBS	Comedy	38.3	1950-1977
1954-55	8	The George Gobel Show	NBC	Comedy/Variety	35.2	1954-1960
1954-55	9	Ford Theatre	NBC	Dramatic Anthology	34.9	1949-1957
1954-55	10	December Bride	CBS	Situation Comedy	34.7	1954-1961
1955-56	1	The $64,000 Question	CBS	Quiz/Audience	47.5	1955-1958
1955-56	2	I Love Lucy	CBS	Situation Comedy	46.1	1951-1957
1955-56	3	The Ed Sullivan Show	CBS	Variety	39.5	1948-1971
1955-56	4	Disneyland	ABC	Anthology	37.4	1954-1990
1955-56	5	The Jack Benny Show	CBS	Comedy	37.2	1950-1977
1955-56	6	December Bride	CBS	Situation Comedy	37	1954-1961
1955-56	7	You Bet Your Life	NBC	Quiz	35.4	1950-1961
1955-56	8	Dragnet	NBC	Police Drama	35	1952-1970
1955-56	9	The Millionaire	CBS	Dramatic Anthology	33.8	1955-1960
1955-56	10	I've Got a Secret	CBS	Quiz/Panel	33.5	1952-1976
1956-57	1	I Love Lucy	CBS	Situation Comedy	43.7	1951-1957
1956-57	2	The Ed Sullivan Show	CBS	Variety	38.4	1948-1971
1956-57	3	General Electric Theater	CBS	Dramatic Anthology	36.9	1953-1962
1956-57	4	The $64,000 Question	CBS	Quiz/Audience	36.4	1955-1958
1956-57	5	December Bride	CBS	Situation Comedy	35.2	1954-1961
1956-57	6	Alfred Hitchcock Presents	CBS	Suspense Anthology	33.9	1955-1986
1956-57	7	I've Got a Secret	CBS	Quiz/Panel	32.7	1952-1976
1956-57	8	Gunsmoke	CBS	Western	32.7	1955-1975
1956-57	9	The Perry Como Show	NBC	Musical Variety	32.6	1948-1963
1956-57	10	The Jack Benny Show	CBS	Comedy	32.3	1950-1977
1957-58	1	Gunsmoke	CBS	Western	43.1	1955-1975
1957-58	2	Make Room for Daddy	CBS	Situation Comedy	35.3	1953-1971
1957-58	3	Tales of Wells Fargo	NBC	Western	35.2	1957-1962
1957-58	4	Have Gun Will Travel	CBS	Western	33.7	1957-1963
1957-58	5	I've Got a Secret	CBS	Quiz/Panel	33.4	1952-1976
1957-58	6	The Life and Legend of Wyatt Earp	ABC	Western	32.6	1955-1961
1957-58	7	General Electric Theater	CBS	Dramatic Anthology	31.5	1953-1962
1957-58	8	The Restless Gun	NBC	Western	31.4	1957-1959
1957-58	9	December Bride	CBS	Situation Comedy	30.7	1954-1961
1957-58	10	You Bet Your Life	NBC	Quiz	30.6	1950-1961

Source: Brooks, T. and Marsh, E. (2003). *The Complete Directory to Prime Time Network and Cable TV 1946-Present* (8th ed.). NY: Ballantine Books.

TV seasons run from October to April. Shaded titles indicate that the series was also in the top ten for the preceding season.

Table A.6: Raw decomposition of difference in retail sales growth over 1948-54 between Pre- and Post-Freeze groups

	Pre-Freeze group		Post-Freeze group		Diff. in $\sum_j \overline{\text{share}_{jt_{48}}} \times \overline{\Delta_{48-54} \log(\text{sales}_j)}$ due to:	
	$\overline{\Delta_{48-54} \log(\text{sales})}$	$\overline{\text{Share}_{48}}$	$\overline{\Delta_{48-54} \log(\text{sales})}$	$\overline{\text{Share}_{48}}$	$\overline{\Delta_{48-54} \log(\text{sales})}$	$\overline{\text{Share}_{48}}$
Category	(1)	(2)	(3)	(4)	(5)	(6)
Food	0.108 (0.009)	0.242 (0.002)	0.071 (0.006)	0.238 (0.002)	0.009 (0.004)	0.000 (0.000)
General stores	-0.001 (0.013)	0.116 (0.002)	-0.041 (0.009)	0.113 (0.002)	0.005 (0.003)	-0.000 (0.000)
Apparel	0.031 (0.012)	0.045 (0.001)	0.039 (0.009)	0.043 (0.001)	-0.000 (0.001)	0.000 (0.000)
Furniture & appliances	0.024 (0.011)	0.046 (0.001)	0.002 (0.009)	0.044 (0.000)	0.001 (0.001)	0.000 (0.000)
Automobile	0.281 (0.010)	0.169 (0.002)	0.210 (0.008)	0.172 (0.001)	0.012 (0.004)	-0.001 (0.001)
Gasoline	0.312 (0.011)	0.073 (0.001)	0.273 (0.009)	0.078 (0.001)	0.003 (0.003)	-0.001 (0.001)
Building & hardware	-0.003 (0.011)	0.134 (0.002)	-0.061 (0.008)	0.156 (0.002)	0.008 (0.004)	0.001 (0.001)
Eating & drinking	0.084 (0.010)	0.065 (0.001)	0.047 (0.009)	0.065 (0.001)	0.002 (0.002)	0.000 (0.000)
Drug stores	0.107 (0.007)	0.028 (0.000)	0.091 (0.005)	0.029 (0.000)	0.000 (0.000)	-0.000 (0.000)
Other	0.154 (0.016)	0.099 (0.002)	0.187 (0.012)	0.086 (0.001)	-0.003 (0.004)	0.002 (0.001)
Total (weighted average)	0.122 (0.012)		0.084 (0.010)		0.036 (0.015)	0.002 (0.002)

This table presents the decomposition of the difference between Pre- and Post-Freeze groups in the weighted average of the mean log growth in each sector over 1948-54 where the weights are the mean share of each sector in 1948. Columns 1 and 3 report the average log growth within each sector over 1948-54 for the Pre- and Post-Freeze groups respectively. Columns 2 and 4 show each sector's average share of total retail sales in 1948. Column 5 calculates $\overline{\text{share}_{j48}^{\text{Pre-Fz.}}} \times \overline{\Delta_{48-54} \log(\text{sales}_j^{\text{Pre-Fz.}})} - \overline{\text{share}_{j48}^{\text{Pre-Fz.}}} \times \Delta_{48-54} \log(\text{sales}_j^{\text{Post-Fz.}})$ where j indexes the sectors, and Column 6 reports $\overline{\text{share}_{j48}^{\text{Pre-Fz.}}} \times \overline{\Delta_{48-54} \log(\text{sales}_j^{\text{Post-Fz.}})} - \overline{\text{share}_{j48}^{\text{Post-Fz.}}} \times \overline{\Delta_{48-54} \log(\text{sales}_j^{\text{Post-Fz.}})}$. Standard errors shown in parentheses are clustered by DMA.

Table A.7: TV access year and highway miles in 1947 plan

Dep. var.: Miles in 1947 plan	(1)	(2)	(3)	(4)
TV access year		-0.080	0.19	-0.56
		(0.47)	(0.52)	(0.26)
Log(1940 population)			2.77	15.5
			(0.93)	(2.41)
Constant	26.2	183.2	-374.2	968.5
	(1.11)	(907.8)	(1020.3)	(504.7)
All demographic controls				✓
Observations	1238	1238	1238	1238
Counties	1238	1238	1238	1238
Media markets (DMA)	181	181	181	181
R^2	0.00	0.00	0.02	0.39

This table shows the relationship between the year of gaining TV access in a county and the number of highway miles allocated to it in the 1947 plan. "All demographic controls" include all the base-year demographic variables listed in Table 5.2 (log population, male %, median age, median family income, log density, white %, urban %, high school educated %, all from 1940) as well as census-region fixed effects. The sample includes counties in the Pre- and Post-Freeze groups that have yearly highway construction data available from Baum-Snow (2007) and gained TV access before 1958. Standard errors shown in parentheses are clustered by DMA.

Table A.8: Access to TV and birthrate

Dep. var.: Births per 1000	(1)	(2)	(3)	(4)	(5)	(6)
Difference relative to Post-Freeze group each year						
Pre-Freeze×$\mathbf{1}\{t = 1944\}$	-0.29	0.22				
	(0.58)	(0.35)				
Pre-Freeze×$\mathbf{1}\{t = 1948\}$	0.46	0.76	0.75	0.54		
	(0.97)	(0.62)	(0.75)	(0.43)		
Pre-Freeze×$\mathbf{1}\{t = 1950\}$	1.94	1.96	2.23	1.74		
	(1.06)	(0.76)	(0.93)	(0.61)		
Pre-Freeze×$\mathbf{1}\{t = 1954\}$	1.57	1.44	1.86	1.22		
	(0.72)	(0.58)	(0.64)	(0.52)		
Pre-Freeze×$\mathbf{1}\{t = 1960\}$	0.88	1.33	1.17	1.11		
	(0.70)	(0.55)	(0.70)	(0.61)		
$\mathbf{1}\{t = 1948\}$	4.31		4.31			
	(0.37)		(0.37)			
$\mathbf{1}\{t = 1950\}$	3.03		3.03			
	(0.49)		(0.49)			
$\mathbf{1}\{t = 1954\}$	0.27		0.27			
	(0.34)		(0.34)			
$\mathbf{1}\{t = 1960\}$	-3.92		-3.92			
	(0.39)		(0.39)			
Years of TV access					0.19	0.31
					(0.078)	(0.11)
Years of TV access^2						-0.0049
						(0.0035)
Constant	21.0	44.1	20.9	43.0	3.80	14.4
	(0.31)	(6.16)	(0.23)	(9.80)	(5.98)	(10.9)
Demographic controls		✓		✓	✓	✓
County FE			✓	✓	✓	✓
Observations	14880	14870	14880	14870	14870	14870
Counties	2976	2974	2976	2974	2974	2974
Media markets (DMA)	199	199	199	199	199	199
R^2	0.10	0.27	0.21	0.32	0.31	0.31

This table compares the birthrate (births per 1000) between the Pre- and Post-Freeze counties in the years when birth data are available (1944, 1948, 1950, 1954, and 1960). Estimates for county populations are interpolated between decennial years (1940, 1950, 1960) assuming a constant growth rate. Demographic controls include all those listed in Table 5.2 (log population, male %, median age, median family income, log density, white %, urban %, high school educated %, all from 1940) with interactions for each year as well as census-region-by-year fixed effects. Region-year fixed effects absorb year fixed effects in Columns 2 and 4. Standard errors shown in parentheses are clustered by DMA.

Table A.9: Selection of Construction Permit-holders vs. Pending Applicants on retail sales before the FCC Freeze

Dep. Var.: Log(retail sales)	(1)	(2)	(3)
Construction Permit-holders group	0.26	-0.20	-0.06
	(0.17)	(0.11)	(0.03)
Log(1940 pop.}×$\mathbf{1}\{t = 1930\}$		1.12	1.08
		(0.05)	(0.05)
Log(1940 pop.}×$\mathbf{1}\{t = 1935\}$		1.16	1.08
		(0.04)	(0.04)
Log(1940 pop.}×$\mathbf{1}\{t = 1939\}$		1.18	1.07
		(0.03)	(0.03)
Log(1940 pop.}×$\mathbf{1}\{t = 1948\}$		1.16	1.07
		(0.04)	(0.05)
$\mathbf{1}\{t = 1935\}$	-0.25	-0.62	
	(0.02)	(0.22)	
$\mathbf{1}\{t = 1939\}$	-0.02	-0.64	
	(0.03)	(0.34)	
$\mathbf{1}\{t = 1948\}$	0.65	0.22	
	(0.03)	(0.15)	
Constant	8.77	-2.14	-1.44
	(0.11)	(0.53)	(1.01)
All controls			✓
Observations	5329	5328	5328
Counties	1333	1332	1332
Media markets (DMA)	76	76	76
R^2	0.07	0.83	0.95

This table compares differences in log retail sales between the Construction Permit-holders and Pending Applicants in 1930, 1935, 1939, and 1948 leading up to the FCC Freeze period (1948-52). Estimates are from OLS regressions. "All controls" include all the demographic characteristics listed in Table 5.2 (log population, male %, median age, median family income, log density, white %, urban %, high school educated %, all from 1940) with interactions for each year as well as census-region-by-year fixed effects. The indicators for each year are absorbed by the region-year fixed effects in Column 3. Standard errors shown in parentheses are clustered by DMA.

Appendix B

"Don't you plan on voting?" The motives and effects of peer pressure in voter mobilization

B.1 Additional survey details

B.1.1 Initial Survey: Non-registrants' WTP to share unregistered status

Before the end of the Initial Survey, the Non-registrants (i.e., the participants who were initially unregistered at the start of the study) took an additional section where they provided their willingness-to-pay to (not) share their registration status with other participants. Their WTP responses capture the loss of privacy and the welfare cost of revealing to others that they have not registered to vote, which could be damaging to their social reputation. They were informed that if their status was shared, they might also receive an email from another participant about the election. To reduce experimenter demand effects, the topic of the email was merely described as being "about the election" without specifying voter registration. The WTP was elicited through the same binary search procedure as for the Senders in Section 3.2.4. The Non-registrants, however, made only one WTP choice for sharing their status, as opposed to the three WTP choices made by Senders who were each assigned three potential Recipients.

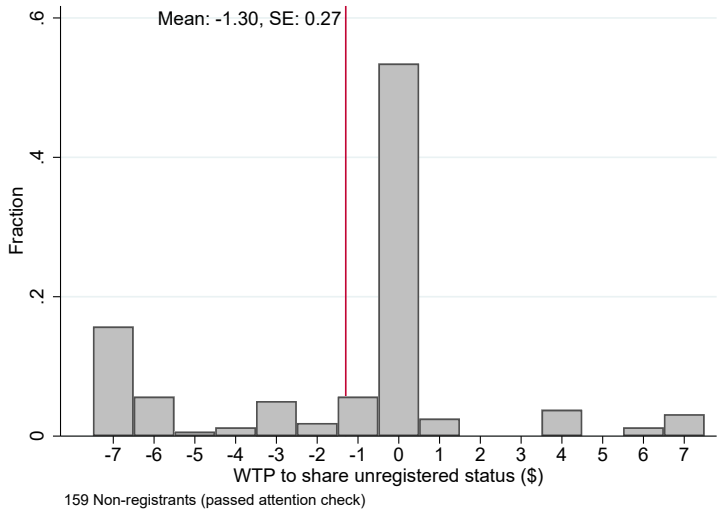
Figure B.1 shows the distribution of the WTP responses from the Non-registrants who passed the attention check preceding the WTP elicitation. The average WTP is -$1.3 (SE=0.3), and 36% of the Non-registrants strictly prefer not to share their information. Over half (53%) are indifferent, and 11% state a positive WTP. After the response of $0, the most common WTP is the negative extreme of $7, which holds 16% of the responses.

B.1.2 Verification, randomization, and matching procedure

Verification. Before the Follow-up Survey, I verified all the self-reported Registrants (i.e., the participants who were initially registered at the start of the study) against the California state voter registration records using a combination of their name, date of birth, and email address. For those whom I could not find in the database, I sent them an email asking them to forward a screenshot verifying their voter registration online within 24 hours. The 24-hour rule disqualified Registrants who registered after starting the study, as voter registration applications took 48 hours to process. This verification process was preregistered, and (`https://doi.org/10.1257/rct.6364`) only verified Registrants were allowed to continue the study.

Randomization and matching. The pre-specified randomization and matching process in this study was stratified on several levels, since there were separate treatments that needed to be randomized on both Non-registrant and Registrant sides, and the treatment for the Non-registrants required interaction from the Registrants. The process was conducted in the following steps after the Initial Survey.

Figure B.1: Non-registrants' willingness to share unregistered status



For 10% of the Non-registrants, their WTP choices on the Initial Survey for sharing their name and voter registration status were exercised (Section B.1.1); one of the rows on the multiple price list was randomly selected, and their preferred option on that row was implemented. These Non-registrants with endogenous sharing or withholding of their information with other participants are excluded from the analysis in Section 3.3.3.

The remaining 90% of the Non-registrants shared their full name and registration status with Registrants on the Follow-up Survey, regardless of their WTP choices. Roughly two-thirds were assigned to either receive a Pressure Message from a Registrant Sender (treatment group), and the remaining one-third was assigned to not receive any message (control group) using stratified randomization based on bins for their WTP, gender, race/ethnicity, and age. More Non-registrants were assigned to the treatment group to counteract attrition among the Registrant Senders.

The Non-registrants in the treatment group who were assigned to receive a Pressure Message were randomly matched to a Registrant Sender selected with uniform probability to be either (a) same gender and race/ethnicity, (b) same gender and different race/ethnicity, (c) different gender and same race/ethnicity, or (d) different gender and race/ethnicity. These randomly selected Registrant Senders were automatically assigned to the Pressure Message condition, saw their matched Non-registrant Recipient under the Direct Message condition, and had to send the Pressure Message directly to their matched Recipient to complete the study, regardless of their WTP choices.

Non-registrants in the 10% who endogenously shared their information with others were uniformly assigned to either the Direct Info, Anonymous Info, Direct Pressure, or Anonymous Pressure condition, and were matched with any remaining Registrant who had not been assigned to send a message. For these matched Registrant Senders, their WTP choices were binding on the Follow-up Survey. Therefore, for Senders in every experimental arm, there was a small chance that their WTP choices would be exercised ex ante.

The remaining Registrants who had not been selected as Senders were randomly assigned

to the Info Message arm (roughly 40%) or to the Pressure Message arm (roughly 60%). The proportion of the two treatments was chosen to balance the number of times each Non-registrant's information was shared with other participants, and to favor more power in the Pressure Message treatment. Regardless of their WTP choices, these Registrant "Senders" were always assigned to not message anyone.

All Registrant Senders randomly drew three potential Recipients (or if they had already been assigned a Recipient, they drew two more) to be shown on the Follow-up Survey. The drawing procedure ensured that each Non-registrant would be shown under both the Direct and Anonymous Pressure Message conditions to two different Registrants, and at least once under the Info Message condition (which could be Direct or Anonymous). All Registrants (regardless of whether they were assigned to the Strong or Info Message condition) saw at least one Recipient under the Direct Message condition, and at least one Recipient under the Anonymous Message condition. Registrants assigned to the Pressure Message condition only saw Non-registrants as potential Recipients, while Registrants assigned to the Info Message condition could see both Non-registrants and other Registrants as potential Recipients.

Ultimately, the randomization ensured that a Non-registrant could receive only one message (if any), and a Registrant could send only one message (if any). The 90% of Non-registrants who shared their information irrespective of their preferences were randomly assigned to receive or not receive a Direct Pressure Message exogenously, regardless of their Senders' WTP choices, which allows us to estimate the causal impact of the message on the Non-registrants' voter registration and turnout.

B.2 Pre-analysis plan details

This study was preregistered on the AEA RCT Registry (`https://doi.org/10.1257/rct.6364`). The preregistration first lists the main outcomes of interest and the covariates. On the Sender's side, the primary outcome is reported as the Sender's willingness-to-pay to send the message, and the covariates are listed as the Sender's beliefs on the instrumental and affective effects of the message, as well as the relationship with the Recipient, gender, and race. On the Recipient's side, the primary outcome is voter registration and turnout, and the key covariate is assignment to the treatment of receiving a peer pressure message from a Sender. As pre-specified, these are the outcomes and covariates that appear in the analyses.

As for the empirical specifications, the table below links the equation number in the preregistration to the respective results in the paper. Any departures from the preregistration are noted with explanations.

Table B.1: Pre-analysis plan specification and results

Equation in preregistration	Result in paper	Departures from model
1	Figure 3b, Columns 1 and 3 in Table 1	–
2	Columns 2 and 4 in Table 1	Based on feedback, the full set of interactions between the Pressure Message, the Anonymity indicator, and the two Sender's beliefs are included for completeness. The additional interaction terms do not meaningfully change the estimates on the prespecified coefficients. Campus and experimental round fixed effects are also omitted in order to present the baseline average in the constant, but including them also does not alter the results.
3	Column 3 in Table A.3a	See note on additional interactions for Equation 2.
4	Figure 5	–
5, 6	–	This result is same in essence to that of Equation 4. Omitted for conciseness based on feedback.
7	Columns 7 and 8 in Table 2	See note on additional interactions for Equation 2. The category "Friends" includes all Recipients who are not Strangers, since there are a very limited number of non-Strangers.
8	Columns 1-6 in Table 2	See note on additional interactions for Equation 2.
9	Figure 7	–
10	–	Unable to perform due to limited number of non-Stranger Recipients.

B.3 Appendix figures and tables

Figure B.2: Within-Sender assignment of Direct and Anonymous Messages

If you're selected to message a participant, it could be 1 of 2 types:

1. **Direct Message**: You'll have to email your message to the participant directly from your lionelmessi@ucla.edu account.

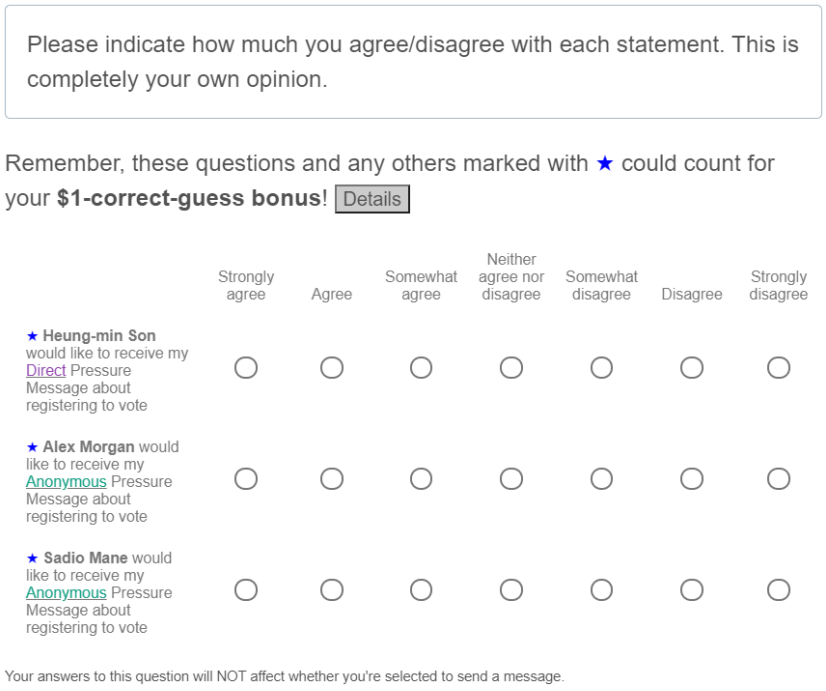
2. **Anonymous Message**: We'll email your message on your behalf, say it's "from another participant", and won't mention your name.

The type of message you'll have to send is different for each participant:

Participant	Message type
Heung-min Son	Direct
Alex Morgan	Anonymous
Sadio Mane	Anonymous

Figure B.3: Eliciting the Sender's beliefs on the affective effect (Recipient's (dis)like of the message)

(a) Pressure Message



(b) Info Message

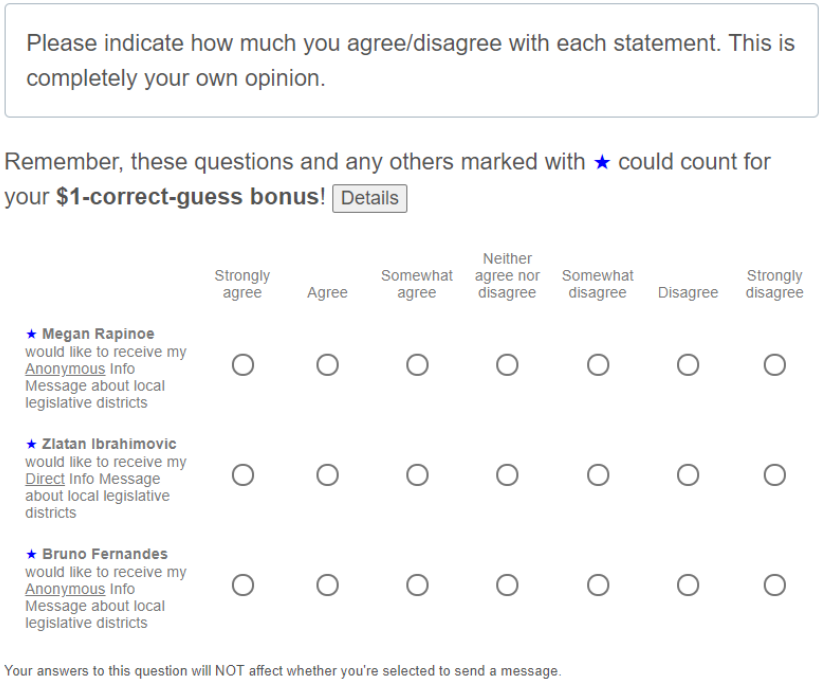
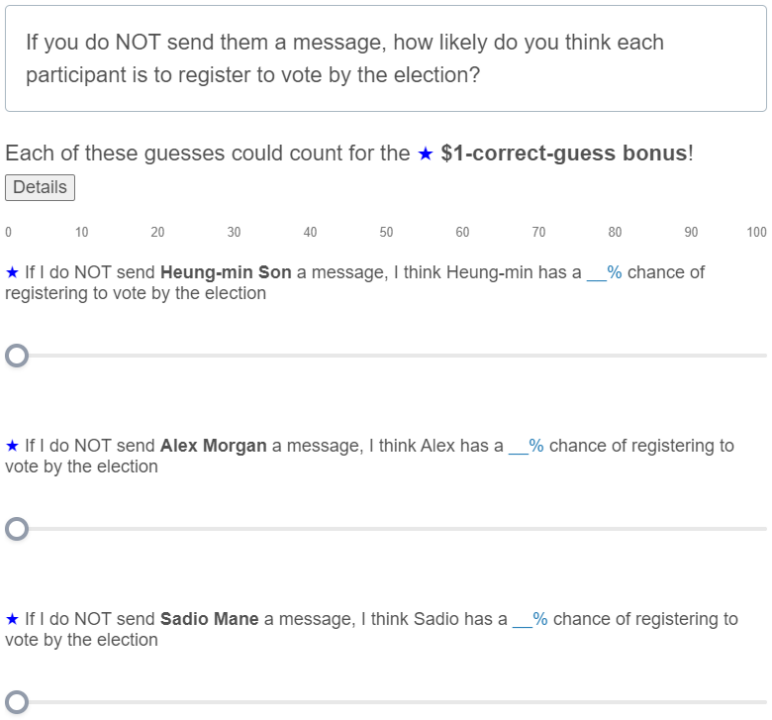


Figure B.4: Eliciting the Sender's beliefs on the instrumental effect (likelihood of persuading the Recipient)

(a) Without Message



(b) With Message

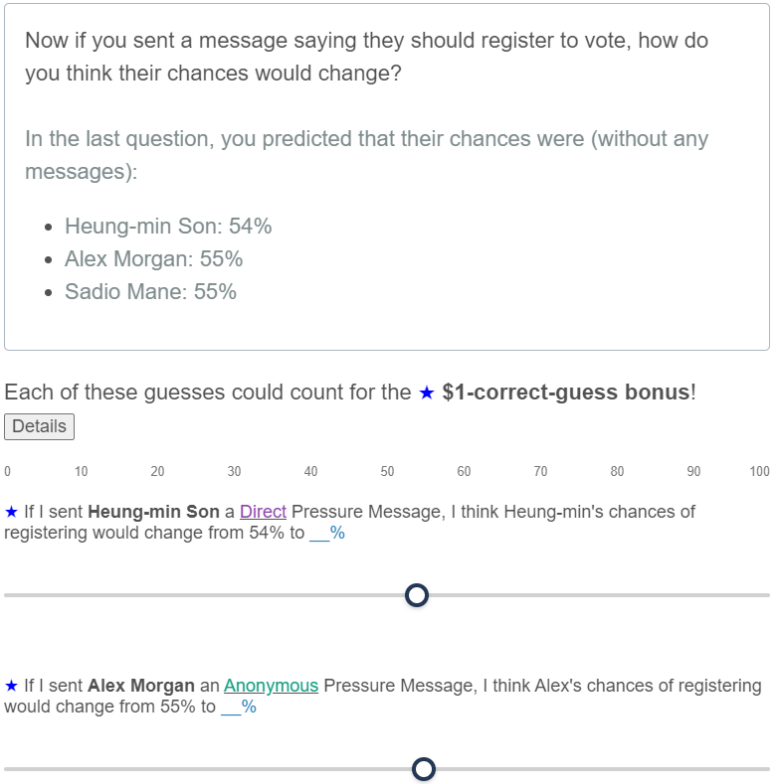


Figure B.5a: Eliciting the Sender's willingness-to-pay to send the message

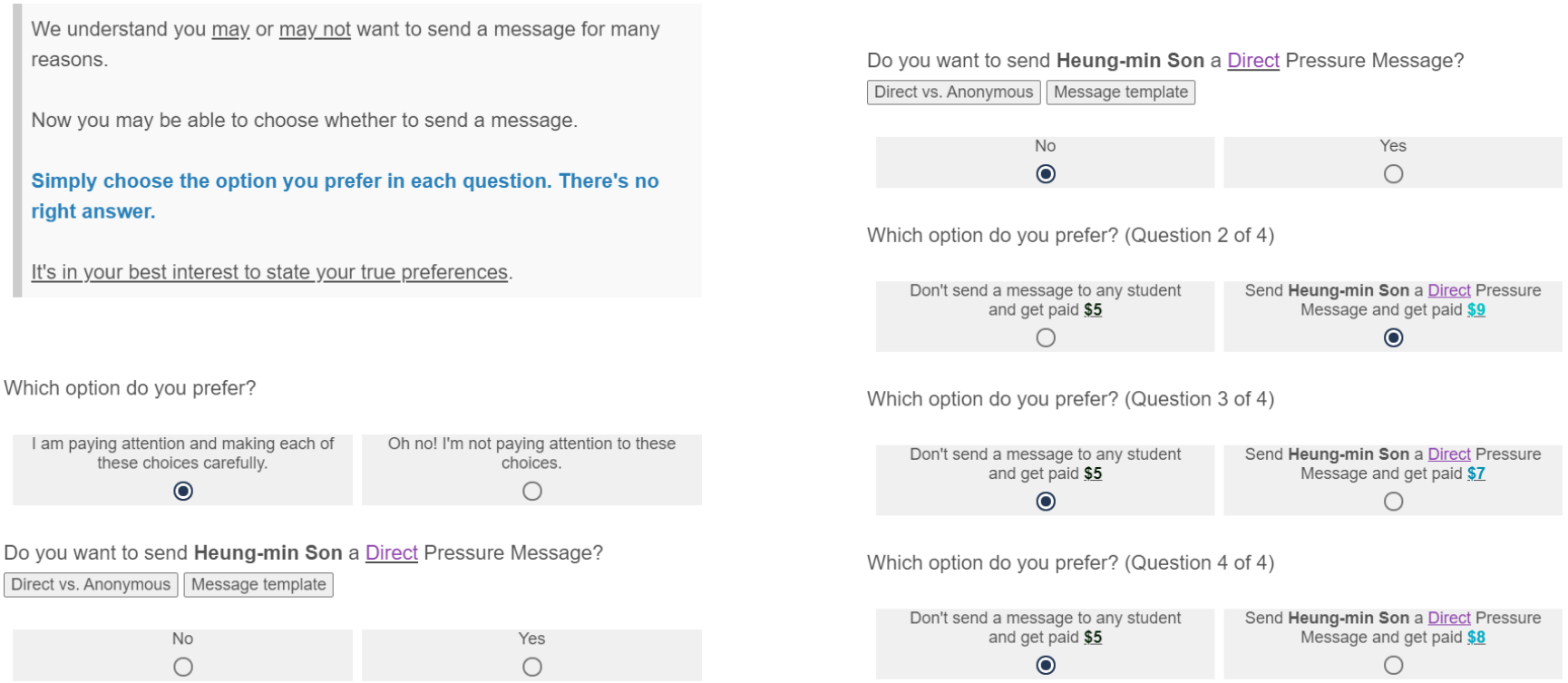
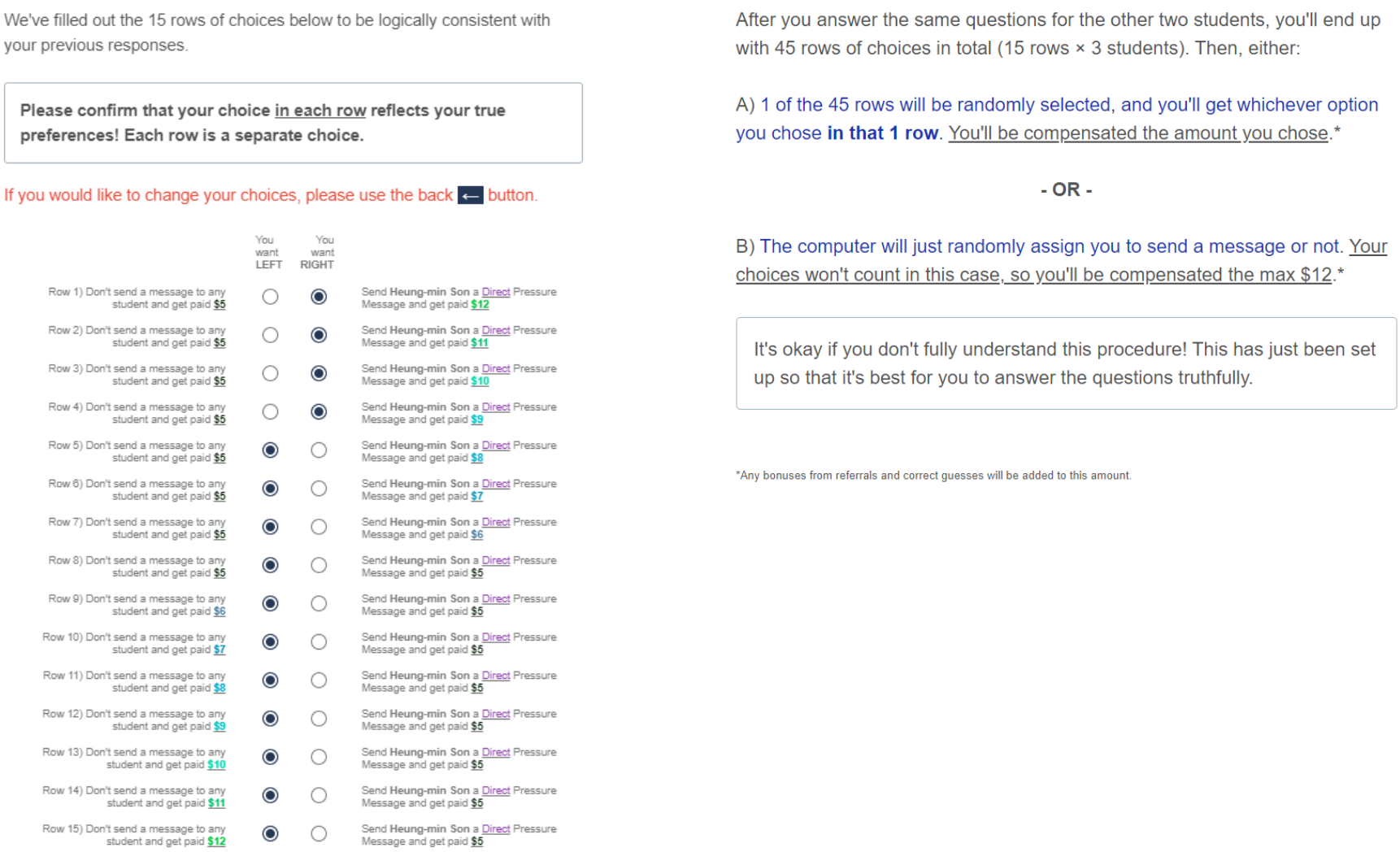


Figure B.5b: Eliciting the Sender's willingness-to-pay to send the message



We've filled out the 15 rows of choices below to be logically consistent with your previous responses.

Please confirm that your choice in each row reflects your true preferences! Each row is a separate choice.

If you would like to change your choices, please use the back ← button.

	You want LEFT	You want RIGHT	
Row 1)	○	●	Send **Heung-min Son** a Direct Pressure Message and get paid $12
Row 2)	○	●	Send **Heung-min Son** a Direct Pressure Message and get paid $11
Row 3)	○	●	Send **Heung-min Son** a Direct Pressure Message and get paid $10
Row 4)	○	●	Send **Heung-min Son** a Direct Pressure Message and get paid $9
Row 5)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $8
Row 6)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $7
Row 7)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $6
Row 8)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 9)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 10)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 11)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 12)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 13)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 14)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5
Row 15)	●	○	Send **Heung-min Son** a Direct Pressure Message and get paid $5

Next, we'll explain what will happen with these choices. You'll be able to come back to change your choices if you'd like.

After you answer the same questions for the other two students, you'll end up with 45 rows of choices in total (15 rows × 3 students). Then, either:

A) 1 of the 45 rows will be randomly selected, and you'll get whichever option you chose **in that 1 row**. You'll be compensated the amount you chose.*

- OR -

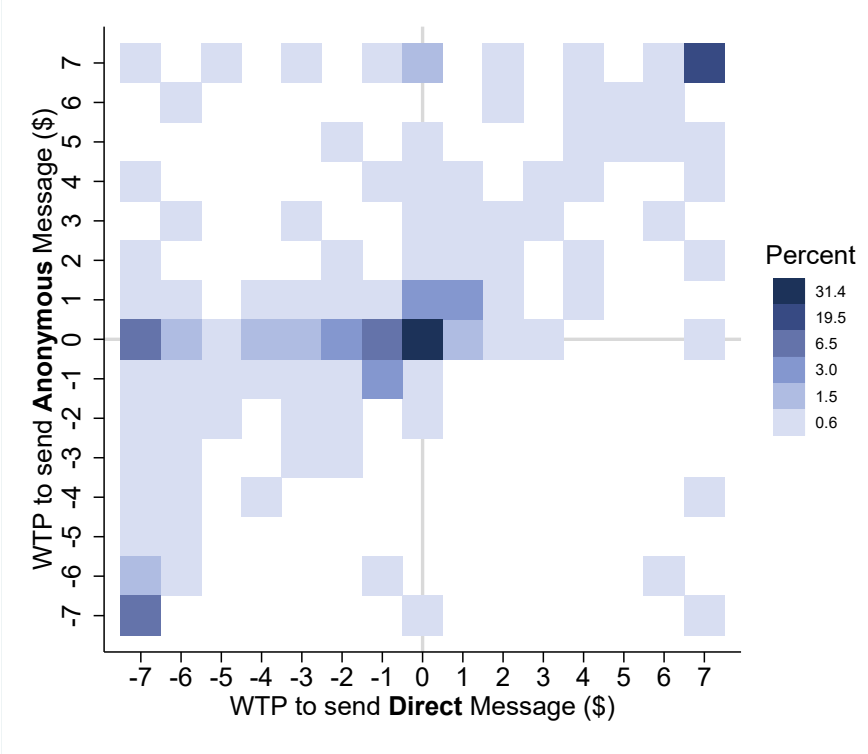
B) The computer will just randomly assign you to send a message or not. Your choices won't count in this case, so you'll be compensated the max $12.*

It's okay if you don't fully understand this procedure! This has just been set up so that it's best for you to answer the questions truthfully.

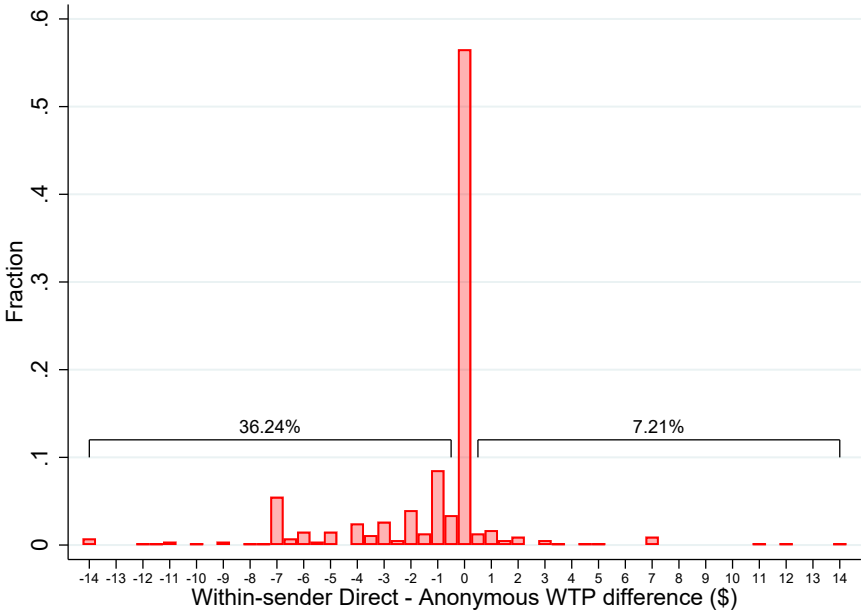
*Any bonuses from referrals and correct guesses will be added to this amount.

Figure B.6: Sender-level WTP distributions for Pressure Message

(a) Joint distribution of Direct and Anonymous



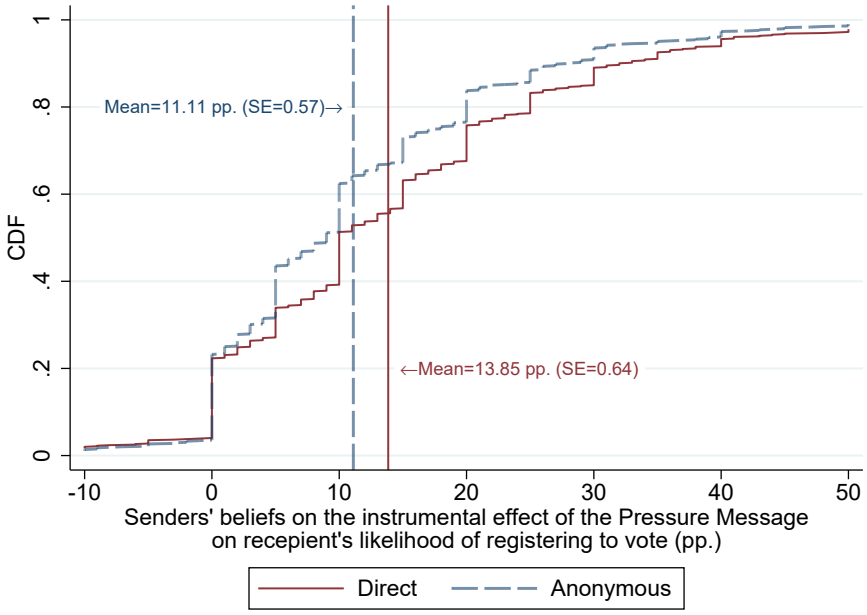
(b) Within-sender difference in Direct vs. Anonymous



Each Sender is assigned 3 potential Recipients. At least 1 of the 3 is assigned to be messaged directly, and at least another 1 of the 3 to be messaged anonymously. Consequently, 2 of the 3 potential Recipients have the same mode of delivery (Direct/Anonymous). For these 2 potential Recipients, the WTP is averaged within-Sender for these figures.

Figure B.7: Distribution of Senders' beliefs on the instrumental effect

(a) Pressure Message: Direct vs. Anonymous



(b) Pressure vs. Info Messages

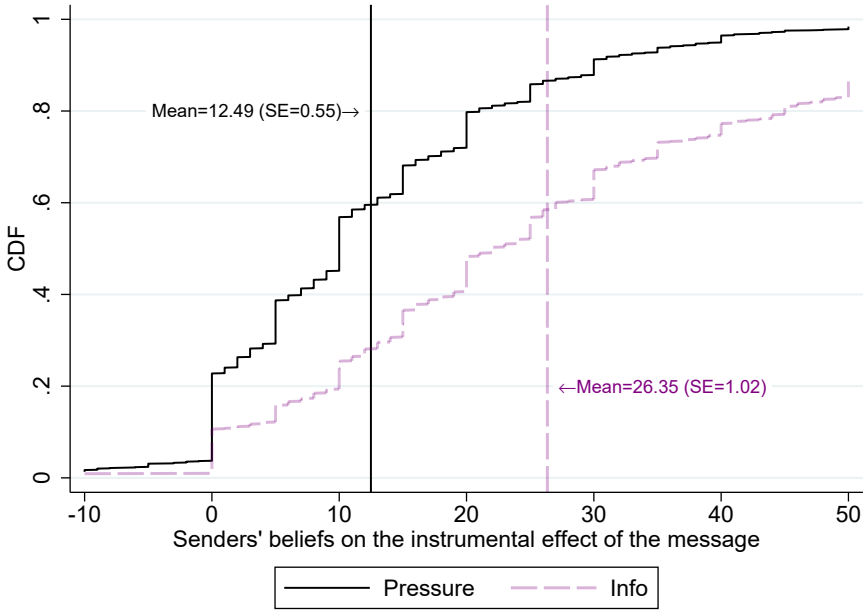
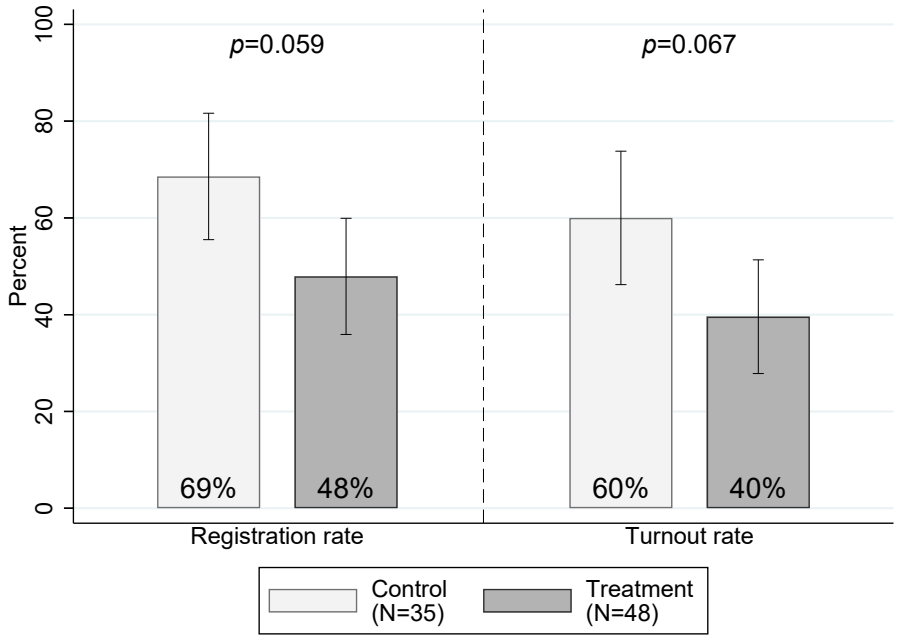


Figure B.8: Intend-to-treat effects of receiving peer pressure on electoral participation (excluding "pre-registered")



95% confidence intervals are shown with robust standard errors.

Table B.2a: Balance: Senders

	Pressure Msg.	Info Msg.	Difference (SE)	Diff. p-value
Gender				
Male	0.31 [0.46]	0.31 [0.46]	0.00 (0.03)	0.94
Female/other	0.69 [0.46]	0.69 [0.46]	-0.00 (0.03)	0.94
Age (years)	21.62 [3.27]	21.46 [2.81]	0.16 (0.20)	0.44
Race/ethnicity				
White	0.29 [0.46]	0.32 [0.47]	-0.03 (0.03)	0.40
Asian	0.47 [0.50]	0.42 [0.49]	0.05 (0.03)	0.11
Other	0.24 [0.43]	0.26 [0.44]	-0.03 (0.03)	0.36
Degree				
Masters/doctoral	0.14 [0.35]	0.12 [0.33]	0.02 (0.02)	0.49
Undergraduate	0.86 [0.35]	0.88 [0.33]	-0.02 (0.02)	0.49
STEM major/field	0.71 [0.46]	0.68 [0.47]	0.02 (0.03)	0.43
Year degree started	2018.09 [1.23]	2018.08 [1.28]	0.01 (0.09)	0.88
Current reg. rate (%)	60.53 [16.81]	61.63 [17.03]	-1.10 (1.15)	0.34
Predictions				
Final reg. rate (%)	74.20 [14.65]	75.13 [14.33]	-0.93 (0.98)	0.35
Recipients characteristics				
Mixed gender comp.	0.67 [0.47]	0.66 [0.48]	0.02 (0.03)	0.60
Mixed racial comp.	0.85 [0.36]	0.87 [0.33]	-0.03 (0.02)	0.28
All 3 Recipients are strangers	0.96 [0.19]	0.96 [0.20]	0.00 (0.01)	0.83
	527	367		0.86 (joint F-test)

This table compares the average characteristics of the registered Senders assigned to the Pressure Message condition versus the Info Message condition. Standard deviations are shown in brackets, and robust standard errors in parentheses.

Table B.2b: Balance: Non-registered Recipients

	Control (No Msg.)	Treatment (Direct Pressure Msg.)	Difference (SE)	Diff. p-value
Gender				
Male	0.39 [0.49]	0.40 [0.49]	-0.01 (0.08)	0.90
Female/other	0.61 [0.49]	0.60 [0.49]	0.01 (0.08)	0.90
Age (years)	21.06 [2.50]	20.67 [2.00]	0.40 (0.37)	0.29
Race/ethnicity				
White	0.34 [0.48]	0.33 [0.47]	0.01 (0.08)	0.91
Asian	0.38 [0.49]	0.47 [0.50]	-0.09 (0.08)	0.27
Other	0.28 [0.45]	0.20 [0.40]	0.08 (0.07)	0.24
Degree				
Masters/doctoral	0.06 [0.24]	0.05 [0.23]	0.01 (0.04)	0.83
Undergraduate	0.94 [0.24]	0.95 [0.23]	-0.01 (0.04)	0.83
STEM major/field	0.73 [0.45]	0.73 [0.45]	-0.00 (0.08)	0.99
Year degree started	2018.26 [1.19]	2018.42 [1.05]	-0.16 (0.19)	0.41
Preference for sharing status				
Indifferent	0.54 [0.50]	0.49 [0.50]	0.05 (0.09)	0.59
Prefer not to share	0.37 [0.48]	0.38 [0.49]	-0.01 (0.08)	0.87
Prefer to share	0.09 [0.29]	0.13 [0.34]	-0.03 (0.05)	0.55
WTP to share status ($)	-1.23 [3.22]	-1.38 [3.70]	0.15 (0.60)	0.80
Predictions				
Current reg. rate (%)	56.54 [15.75]	53.29 [18.20]	3.25 (2.93)	0.27
Final reg. rate (%)	70.58 [15.07]	67.67 [17.84]	2.91 (2.86)	0.31
	95	55		0.91 (joint F-test)

This table compares the average characteristics of the non-registered Recipients assigned to the treatment and control groups. Standard deviations are shown in brackets, and robust standard errors in parentheses. This table exlcudes 15 non-registered participants who received or did not receive a message endogenously depending on their or their Sender's WTP choices.

Table B.2c: Balance: Recipient relation to assigned Sender

	Same gender & race	Same gender, diff. race	Diff. gender, same race	Diff. gender & race	Referral
White male	5	3	2	1	2
White female	3	3	3	5	5
Asian male	2	3	4	2	2
Asian female	5	6	5	5	2
Other race male	2	3	3	2	1
Other race female	4	2	2	3	5

$\chi^2 = 11.09$ (p=0.94)

Table B.2d: Sample attrition

	Non-registered	Registered		
		(1) Info Message	(2) Pressure Message	(1)-(2) p-val.
Finished Initial Survey	165	461	657	
Finished Follow-up Survey	144	84% (386/461)	84% (549/657)	0.94
Sent required email	NA	99% (383/386)	99% (541/549)	0.32
Sent required confirmation email	NA	99% (382/385)	99% (471/474)	0.80
Sent required Direct email	NA	100% (1/1)	93% (70/75)[A]	1.00[C]
Passed attention check	138	96% (367/383)	98% (529/541)	0.10
Completed study (**final sample**)	138	**367**	**527**[B]	

This table shows the frequency of participants remaining after each sample restriction. The frequency is shown for the Senders separately by the Pressure and Info Message conditions. The last column shows the p-value of the difference in the frequency at each restriction.

[A]As pre-registered, 1 participant who finished the Follow-up survey, stated a WTP to send the Direct Message of less than $12, and did not send the Direct Message is considered to have fulfilled the "Sent required Direct Message" condition.

[B]2 participants stated that they did not know how to fill out the WTP responses in the open-ended feedback box. Consequently, although they completed all the study requirements, their WTP responses are not included in the analysis.

[C]p-value calculated using Fisher's exact test given the low counts in some of the 2x2 cells.

Table B.3: Sample characteristics

	Registered	Non-registered	Difference (SE)	Diff. p-value
Gender				
Male	0.31 [0.46]	0.40 [0.49]	-0.09 (0.04)	0.03
Female/other	0.69 [0.46]	0.60 [0.49]	0.09 (0.04)	0.03
Age (years)	21.55 [3.09]	21.01 [2.52]	0.54 (0.22)	0.01
Race/ethnicity				
White	0.30 [0.46]	0.32 [0.47]	-0.02 (0.04)	0.65
Asian	0.45 [0.50]	0.41 [0.49]	0.04 (0.04)	0.31
Other	0.25 [0.43]	0.27 [0.45]	-0.02 (0.04)	0.52
Degree				
Masters/doctoral	0.13 [0.34]	0.07 [0.25]	0.07 (0.02)	0.00
Undergraduate	0.87 [0.34]	0.93 [0.25]	-0.07 (0.02)	0.00
STEM major/field	0.70 [0.46]	0.72 [0.45]	-0.03 (0.04)	0.51
Year degree started	2018.09 [1.25]	2018.32 [1.13]	-0.23 (0.10)	0.02
Predictions				
Current reg. rate (%)	60.98 [16.90]	55.62 [16.63]	5.35 (1.41)	0.00
Final reg. rate (%)	74.58 [14.52]	69.69 [16.25]	4.89 (1.35)	0.00
	894	165		

This table shows averages with standard deviations in brackets. The "reg. rate" predictions refer to the forecasts for the voter registration rates on the participant's campus.

Table B.4a: Senders' WTP and beliefs across experimental conditions: Alternate specifications and robustness

	Pressure Message				Info Message									
Dep. Var.: WTP to send message ($)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(2)-(6)	(3)-(7)	(1)-(9)	(9)-(10)
Sender's beliefs on the effects of the message:														
Instrumental effect (10 pp.)	0.52	0.52	0.08	0.65	-0.03	-0.03	-0.09	0.13	-0.16	-0.02	0.54	0.17	0.69	-0.15
	(0.11)	(0.11)	(0.15)	(0.22)	(0.07)	(0.07)	(0.07)	(0.18)	(0.22)	(0.08)	(0.13)	(0.17)	(0.24)	(0.22)
Instrumental effect^2				-0.04				-0.02						
				(0.04)				(0.02)						
Affective effect [-2, 2]	0.99	0.97	0.64	0.95	0.55	0.54	0.17	0.52	0.50	0.56	0.44	0.46	0.49	-0.06
	(0.12)	(0.12)	(0.10)	(0.13)	(0.12)	(0.11)	(0.10)	(0.11)	(0.31)	(0.12)	(0.17)	(0.14)	(0.32)	(0.32)
Affective effect^2				-0.10				0.11						
				(0.10)				(0.10)						
Anonymous Message	0.90	0.91	0.90	0.81	0.93	0.98	0.74	1.13	0.97	0.95	-0.07	0.16	-0.07	0.02
	(0.24)	(0.24)	(0.18)	(0.36)	(0.26)	(0.24)	(0.19)	(0.39)	(0.93)	(0.25)	(0.34)	(0.26)	(0.94)	(0.92)
Anonymous×Instrumental effect	-0.04	-0.04	-0.05	0.15	0.09	0.07	0.09	-0.03	0.39	0.04	-0.10	-0.14	-0.43	0.35
	(0.12)	(0.12)	(0.09)	(0.25)	(0.09)	(0.08)	(0.07)	(0.18)	(0.29)	(0.09)	(0.15)	(0.11)	(0.30)	(0.30)
Anonymous×Instrumental effect^2				-0.04				0.02						
				(0.05)				(0.03)						
Anonymous×Affective effect	-0.68	-0.65	-0.38	-0.65	-0.30	-0.30	-0.06	-0.27	-0.23	-0.29	-0.36	-0.31	-0.44	0.06
	(0.14)	(0.14)	(0.09)	(0.14)	(0.14)	(0.14)	(0.10)	(0.14)	(0.32)	(0.15)	(0.20)	(0.14)	(0.34)	(0.34)
Anonymous×Affective effect^2				0.01				-0.04						
				(0.12)				(0.11)						
Constant	-0.60	-0.12		-0.44	-0.89	-0.96		-1.27	-0.64	-0.91				
	(0.24)	(0.62)		(0.35)	(0.27)	(0.67)		(0.40)	(0.61)	(0.29)				
Sender controls		✓				✓					✓			
Sender fixed effects			✓				✓					✓		
Recipients														
Registered participants only	✓	✓	✓	✓					✓				✓	
Non-registered participants only										✓				
Observations (Responses)	1581	1581	1581	1581	1100	1100	1100	1100	97	1003	2681	2681	1678	1100
Senders	527	527	527	527	367	367	367	367	80	366	894	894	607	367
R^2	0.14	0.15	0.85	0.15	0.07	0.08	0.82	0.07	0.12	0.07	0.13	0.84	0.15	0.08

Coefficient estimates are from linear regressions with standard errors clustered by Sender shown below in parentheses. The dependent variable is the willingness-to-pay (WTP) to email the message. Sender controls include gender, standardized age, age squared, race/ethnicity, undergraduate degree, and STEM major/field. See notes in Table 5.3.

Table B.4b: Senders' WTP and beliefs across experimental conditions (uncensored WTP)

	Pressure Msg.	Info Msg.	
Dep. Var.: Uncensored WTP to send message ($)	(1)	(2)	(1)-(2)
Sender's beliefs on the effects of the message:			
Instrumental: Persuasive effect of message (10 pp.)	1.23	-0.04	1.28
	(0.33)	(0.13)	(0.35)
Affective: Recipient's (dis)like of message [-2, 2]	2.21	0.83	1.38
	(0.32)	(0.19)	(0.38)
Anonymous Message	1.90	1.39	0.51
	(0.62)	(0.46)	(0.77)
Anonymous×Instrumental belief	-0.17	0.17	-0.33
	(0.33)	(0.16)	(0.36)
Anonymous×Affective belief	-1.41	-0.42	-0.98
	(0.43)	(0.23)	(0.49)
Constant	-0.11	-1.41	1.31
	(0.66)	(0.49)	(0.82)
Observations (Responses)	1581	1100	2681
Senders	527	367	894
R^2	0.08	0.06	0.10

OLS estimates with standard errors clustered by Sender in parentheses. The dependent variable is the *uncensored* willingness-to-pay (WTP) to send the message assigned to the Sender. For Senders who gave a WTP at the extremes (-$7 or $7), the survey asked them to provide their hypothetical WTP beyond these bounds. For these Senders, I use their hypothetical responses winsorized at the 10th and 90th percentiles, which correspond to -$20 and $40 for the Pressure Message and -$15 to $15 for the Info Message. See notes from Table 5.3.

Table B.5: Effect of receiving a Direct Pressure Message on rates of electoral participation

	Voter registration			Voter turnout		
	ITT	TOT		ITT	TOT	
	(1)	(2)	(3)	(4)	(5)	(6)
Receiving Direct Pressure Message	-0.10	-0.19		-0.05	-0.11	
	(0.06)	(0.11)		(0.08)	(0.14)	
Sender's beliefs on the effects of the message:						
Instrumental: Persuasive effect of message (standardized)			0.00			-0.00
			(0.06)			(0.06)
Affective: Recipient's (dis)like of message (standardized)			0.03			-0.01
			(0.06)			(0.06)
Sender's WTP to send message (standardized)			0.02			-0.00
			(0.06)			(0.06)
Constant	0.87	0.88	0.78	0.75	0.76	0.72
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
Treatment group only			✓			✓
N	150	130	74	150	130	74
K-P F-stat		116.01			116.01	

This table compares the voter registration and turnout rates between non-registered participants in the treatment group (assigned to receive a Direct Pressure Message) and in the control group (assigned to receive no message). Columns 1 and 4 show the intend-to-treat (ITT) results, and Columns 2 and 5 the treatment-on-treated (TOT) results. Columms 4 and 6 predict the effect of the message on the treatment group using the Sender's beliefs on the instrumental and affective effects of the message on the Recipient as well as the Sender's WTP to send the message. The Sender's beliefs and WTP have been standardized. Robust standard errors are displayed in parentheses.

Table B.6: Non-registered Recipients: Experimental outcomes

	Response in Follow-up Survey			
	Received email	Did not receive email	No response	Total
Panel A. Control group				
Did not register	0 [0%(↓) 0%(→)]	6 [12%(↓) 86%(→)]	1 [17%(↓) 14%(→)]	7 [13%(↓) 100%(→)]
Registered but did not vote	0 [0%(↓) 0%(→)]	6 [12%(↓) 86%(→)]	1 [17%(↓) 14%(→)]	7 [13%(↓) 100%(→)]
Registered and voted	0 [0%(↓) 0%(→)]	37 [76%(↓) 90%(→)]	4 [67%(↓) 10%(→)]	41 [75%(↓) 100%(→)]
Total	0 [0%(↓) 0%(→)]	49 [100%(↓) 89%(→)]	6 [100%(↓) 11%(→)]	55 [100%(↓) 100%(→)]
Panel B. Treatment group				
Did not register	10 [21%(↓) 45%(→)]	9 [27%(↓) 41%(→)]	3 [21%(↓) 14%(→)]	22 [23%(↓) 100%(→)]
Registered but did not vote	4 [8%(↓) 57%(→)]	2 [6%(↓) 29%(→)]	1 [7%(↓) 14%(→)]	7 [7%(↓) 100%(→)]
Registered and voted	34 [71%(↓) 52%(→)]	22 [67%(↓) 33%(→)]	10 [71%(↓) 15%(→)]	66 [69%(↓) 100%(→)]
Total	48 [100%(↓) 51%(→)]	33 [100%(↓) 35%(→)]	14 [100%(↓) 15%(→)]	95 [100%(↓) 100%(→)]

Format: Count [column frequency%(↓) row frequency%(→)]

This table tabulates non-registered participants by the post-election experimental outcome (did not register, registered but did not vote, registered and voted) and whether they indicated in the Follow-up Survey that they had received an email from another participant in the study about registering to vote. This table excludes 15 non-registered participants who received or did not receive a message endogenously depending on their or their Sender's WTP choices.

Appendix C

Appendix for Policy diffusion and polarization across U.S. states

C.1 Alternate measures of correlated adoptions

In the hazard model analysis (Section 4.4.2), we use a two-sided "likelihood" as the baseline measure of how concentrated the adoption of a policy has been among states that are (dis)similar in each dimension. We tried other measures that may be simpler but did not perform as well in specification checks. In this section, we define three alternate measures and discuss their shortcomings. Reassuringly, as shown in Table A.6, we find that the dynamic patterns of policy diffusion remain similar regardless of the measure used.

We assess two attributes of each measure. First, we consider its range of possible values as a function of the number of total adopters. Drastic variation in the range may lead to mis-specification when entering the measure as a linear term in the logit, as done in the main analysis of Table 3, since this assumes that the same coefficient applies to early as well as to late adopters of the policy. In the second assessment, we check directly for this mis-specification by allowing the coefficient on the measure to vary by the number of total adopters so far: for the first five adopters (1-5), the second five adopters (6-10), the third five adopters (11-15), and the later adopters (>15). Stable coefficients are encouraging, but coefficients that systematically differ between the early and the later adopters indicate that the estimates from the model under- or over-estimate the responsiveness to adoption among similar states at some stage of the policy's life-cycle.

To start with the baseline likelihood measure, Figure A.11a shows that its range goes from -1 to 1 and is fairly consistent across the domain of total adopters. Figure A.11e then plots the coefficients on the first three groups of 5 adopters and on the following adopters for the two dimensions of interest, distance and Republican vote-share. There does not appear to be any systematic ordering or reversals in the coefficients across the bins, and the coefficients generally remain within each other's confidence intervals. These checks return a favorable evaluation of the baseline measure.

Now for the three alternate measures below, we notate $a^k \in \{0, 1, ..., 15\}$ as the number of adopters among the 15 states that compose the closest third in dimension k, and $A \in \{0, 1, ..., 47\}, A \geq a^k$, as the number of adopters among all other 47 contiguous states.

Proportion of states in the closest third that are adopters ($a^k/15$)**.** As Figure A.11b shows, the range of this measure is limited in both the early and late stages of a policy's life-cycle. For instance, if there are 5 total adopters of the policy, then the measure can range only from 0/15 to 5/15. From 16 to 32 total adopters, the measure ranges from 0 to 1. After 32 total adopters, the range shrinks toward the upper region. Another downside is that this measure does not incorporate information about the total number of adopters, though intuitively, we should consider a case when there are 10 total adopters of a policy and all 10 are in the closest third as a stronger sign of correlated adoptions than the case when

there are 30 total adopters of which 10 are in the closest third. In light of these drawbacks, Figure A.11f finds that the coefficients on the first bin of 5 adopters are significantly lower, and even become negative, compared to the coefficients for the rest of the bins. Hence using this measure in the main specification would lead to a poor fit of the early stage diffusion process.

Proportion of all adopters that are in the closest third. (a^k/A).[72] As shown in Figure A.11c, this measure ranges from 0 to 1 until there are 16 total adopters, at which point there must be more total adopters than adopters in the closest third and thus the upper bound of the range decreases. From 33 total adopters, the lower bound of the range becomes strictly positive, since there must be at least one adopter in the closest third, and continues to increase. Given this narrowing range, similar concerns arise as with the previous measure. Figure A.11g confirms these issues, and shows that the coefficients are systematically increasing in the bins. For this measure, a single coefficient in the specification would be overly sensitive for the early adopters and too unresponsive for the later adopters.

Proportion of states in the closest third that adopters minus proportion of all states that are adopters $\left(\frac{a^k}{15} - \frac{A}{47}\right)$**.** Figure A.11d plots the range of this measure. The difference between the upper and lower bounds linearly increases in the number of total adopters and is maximized at 1 while the number of total adopters is between 15 and 32. After 32 total adopters, the range begins to linearly decrease. Figure A.11h shows that this measure is not as poorly behaved as the previous two in the logit model, but the coefficients do seem to systematically decrease across the bins in the distance dimension. The pseudo-R^2 from Row 10 of Table A.6 also indicates that this measure provides a poorer fit of the data compared to the baseline measure (Row 5 of Table A.6).

C.2 Estimating voter preferences using the ANES, GSS, and measures from political science

In Section 4.5.1, we introduce two measures of the similarity between states in voter preferences and public opinion. The first uses survey data from the American National Election Studies (ANES) and the General Social Survey (GSS), which are national surveys of American voters frequently featured in political research. The second draws on existing measures of public opinion from political science. This section describes the construction of these measures in detail.

C.2.1 Measuring voter preferences using the ANES and GSS

The ANES and GSS surveys collect demographic and background information about voters, their views on societal issues, knowledge about politics, and voting behavior. We use the cumulative ANES data set including surveys conducted over 1948-2020 and the GSS data set covering surveys over 1972-2018.

[72] Another interpretation of this measure is the ratio of the proportion of states in the closest third that are adopters to the proportion of all states that are adopters, or $(a^k/15)/(A/47)$, multiplied by a constant $(15/47)$.

Question selection. We filter through all the survey items to identify 59 questions in the ANES and 468 in the GSS that asked voters about their preference for a specific policy on an ordinal response scale. For example, these include:

- (ANES: VCF0806) *There is much concern about the rapid rise in medical and hospital costs. Some people feel there should be a government insurance plan which would cover all medical and hospital expenses for everyone. Others feel that all medical expenses should be paid by individuals, and through private insurance plans like Blue Cross or other company paid plans. Where would you place yourself on this scale, or haven't you thought much about this? (7-POINT SCALE SHOWN TO RESPONDENT)*

- (ANES: VCF0837) *There has been some discussion about abortion during recent years. Which one of the opinions on this page best agrees with your view?*

    1. *By law, abortion should never be permitted.*
    2. *The law should permit abortion only in case of rape, incest, or when the woman's life is in danger.*
    3. *The law should permit abortion for reasons other than rape, incest, or danger to the woman's life, but only after the need for the abortion has been clearly established.*
    4. *By law, a woman should always be able to obtain an abortion as a matter of personal choice.*

- (ANES: VCF0877) *Do you think homosexuals should be allowed to serve in the United States Armed Forces or don't you think so? (5-POINT SCALE SHOWN TO RESPONDENT)*

- (GSS: cuthours) *Here are some things the government might do for the economy. Circle one number for each action to show whether you are in favor of it or against it. "Reducing the work week to create more jobs." (5-POINT SCALE SHOWN TO RESPONDENT)*

- (GSS: aidcol) *On the whole, do you think it should or should not be the government's responsibility to give financial assistance to college students from low-income families? (4-POINT SCALE SHOWN TO RESPONDENT)*

We then restrict the final sample to 49 questions in the ANES and 196 in the GSS that have been asked for at least 10 years, to ensure the responses reflect voter preferences on longstanding, key issues and to reduce noise from compositional changes in the sample. For example, this restriction drops a question asking whether the respondent thought the United States should cooperate more with the Soviet Union, which was asked only from 1980-88. We link questions in the GSS that are under different variable names but are qualitatively on the same policy preference. For example, the GSS variable *aidneedy* asks, *"For students whose parents have a low income, should the government provide grants that would not have to be paid back, provide loans which the student would have to pay back, or should the government not provide any financial assistance?"* This question is only asked in the 1985 survey, but we link this question to *aidcol* (shown in the last example above), which has been asked for 26 years over 1990-2016. Hence, we keep *aidneedy* as well. Overall, Figure A.7b shows that there are typically over 10 questions represented from the 1960s, and over 20 from the 1970s.

State representation. One downside of the ANES and GSS surveys, depicted in Figure A.7a, is that not all states are sufficiently represented in every wave of the survey. In fact, there are 15 states (Delaware, Idaho, Kansas, Maine, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, Vermont, West Virginia, and Wyoming) that are missing for at least a decade. To prevent the results being contaminated from the addition of these states only in certain time periods, we drop them entirely when we use the ANES and GSS measure of voter preferences. This means that whenever we include the ANES and GSS measure, we also re-calculate the measure of similarity in Equation 6 for all the other dimensions, since there are now 33 states in the sample.

Respondent weighting and measure construction. Respondents in the ANES and GSS are sampled to be representative at the national level, not at the state level. Therefore, we use propensity weighting to provide a more accurate measure of voter preferences at the state level. We reweight the respondents in each survey wave based on their age, gender, and race to match the population distribution in the official Census and American Community Survey data for their state. We then calculate state-level yearly averages of the responses to each question using these propensity weights.

Next, we standardize the state average responses for each question in each year, subtracting by the mean and dividing by the standard deviation, to bring all responses (e.g., 5-point Likert vs. 2-point Likert) to the same scale. For every pair of states in each year of the survey, we compute the absolute difference in the standardized state average response to each question, and then take the average of the absolute differences across all questions in that year. At this point, we have a measure between every pair of states for how similarly their average voter responded to the policy preference questions. To smooth the measure, we use a moving average including the previous five years and next five years of the average standardized difference between each pair of states. Finally, for each state, we consider the third of other states (i.e., 10 out of the other 32) with the smallest average standardized difference in the responses to be the closest in voter preferences for that year. Figure A.4 shows the stability of this measure, and Table A.5b lists examples of states that are close in these voter preferences.

Matching voter preference questions to policy areas. In Columns 1-3 of Table A.7b, we extend the analysis by constructing measures of voter preferences that are specific to each policy area. A team of 10 undergraduate research assistants categorized each question used in the ANES and GSS measure into one of the six policy areas (Table 1c) or as none of them, based on the set of SPID and NBER policies classified under each policy area. Each question was assigned to two research assistants to code independently. The coders agreed on the policy area for 71% of the questions on the first attempt, which indicates that most of the questions were straightforward to classify. For the remaining 29%, we discussed each question as a group before settling on the final categorization. The composition of the questions over time by policy area is shown in Figure A.7c. The questions in the "Environment and Energy" policy area are missing until the 1970s, and even then, there are no respondents for those questions in Kentucky, Mississippi, and Kentucky until the 1980s; hence, we drop policies in

"Environment and Energy" for this analysis.

For each policy, we then calculate the Equation 6 measure of adoption among states closest in voter preferences (i) for questions specifically in that policy area, as well as (ii) for the remaining questions in all the other policy areas. That is, we explore whether states are more likely to adopt Economic policies from other states with voters who express similar preferences on questions specifically related to the economy, more so than states with voters who express similar preferences in other policy areas such as Civil Rights and Public Services. We find that both measures of voter preferences—specific to the policy area and in other policy areas—are comparably predictive of policy diffusion in each time period.

Including voter sentiment. In Columns 4-6 of Table A.7b, we use a broader measure to also capture voter sentiment that could be relevant for their policy positions. We add questions such as "thermometers" about specific groups, (e.g., on a scale of 0-100, how the respondent feels about labor unions, homosexuals, or people on welfare), whether society should make sure that everyone has an equal opportunity to succeed, and whether it matters that the respondent votes or not. Figure A.7b shows the number of voter sentiment questions in this broader measure over time. Reassuringly, this broader measure finds similar results.

C.2.2 Measures of public opinion from political science research

As another measure of state-level voter policy preferences, we combine prominent measures from political science research that span policy topics and decades from at least the 1960s to the 2010s. In particular, we use:

- the mass social and economic policy liberalism scores (1936-2014) from Caughey and Warshaw (2018), which are estimated from surveys (including the ANES and GSS) as well as polling data (such as the Gallup),
- the state policy mood measure (1956-2020) from Lagodny et al. (2022), which also aggregates data from public opinion surveys, and
- the updated citizen ideology measure (1960-2016) from Berry et al. (1998), which is based on interest group ratings of state representatives in Congress and state election results.

From these measures, we construct an index in the same method as for demographics. We standardize each measure within each year across states, calculate the absolute difference in each standardized measure, and average the absolute differences to create the index. The states that have the smallest averaged absolute difference in this index are defined as the closest in public opinion.

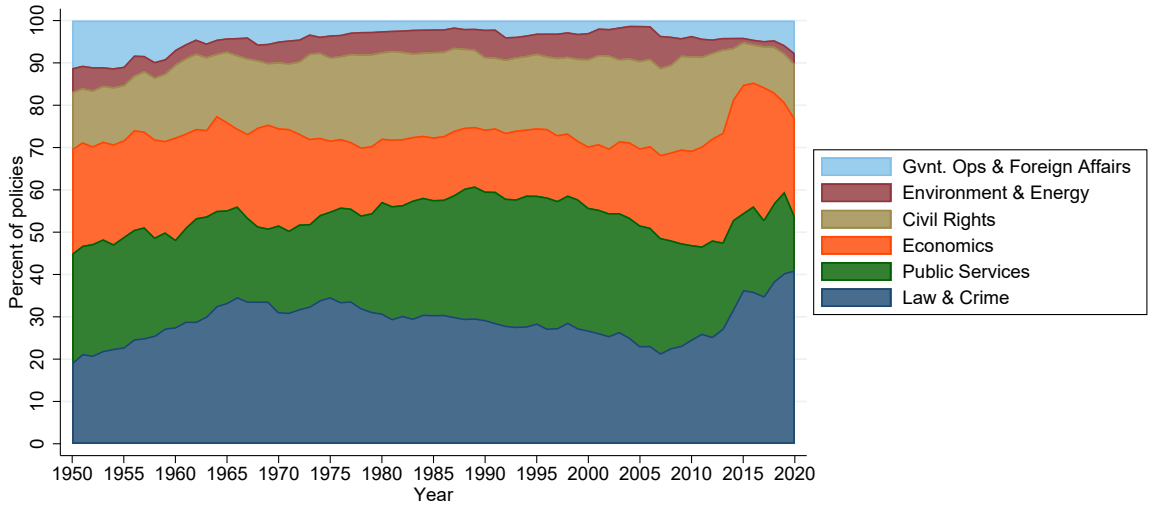
In addition to the analysis in Table 5, we also investigate the role of each measure in the index separately in Columns 4-6 of Table A.7a. For the measures that end before 2020, we extend the measure from the last year (2014 in Caughey and Warshaw, 2018, and 2016 in Berry et al., 1998) through 2020. The Berry et al. (1998) measure strongly predicts policy diffusion in the 1960-70s, but less so in the later time periods. Interestingly, the Caughey and Warshaw (2018) measures have become more predictive in recent times, and more so for

the social liberalism measure, which is consistent with the results in Caughey and Warshaw (2018). Nevertheless, the escalating role of state government partisanship far exceeds the explanatory power of any measure of voter preferences in the last two decades.

C.3 Appendix figures and tables

Figure C.1: Summary statistics: Policy area composition and adoption speed

(a) Composition of policy areas over time



(b) Speed of adoption by time period

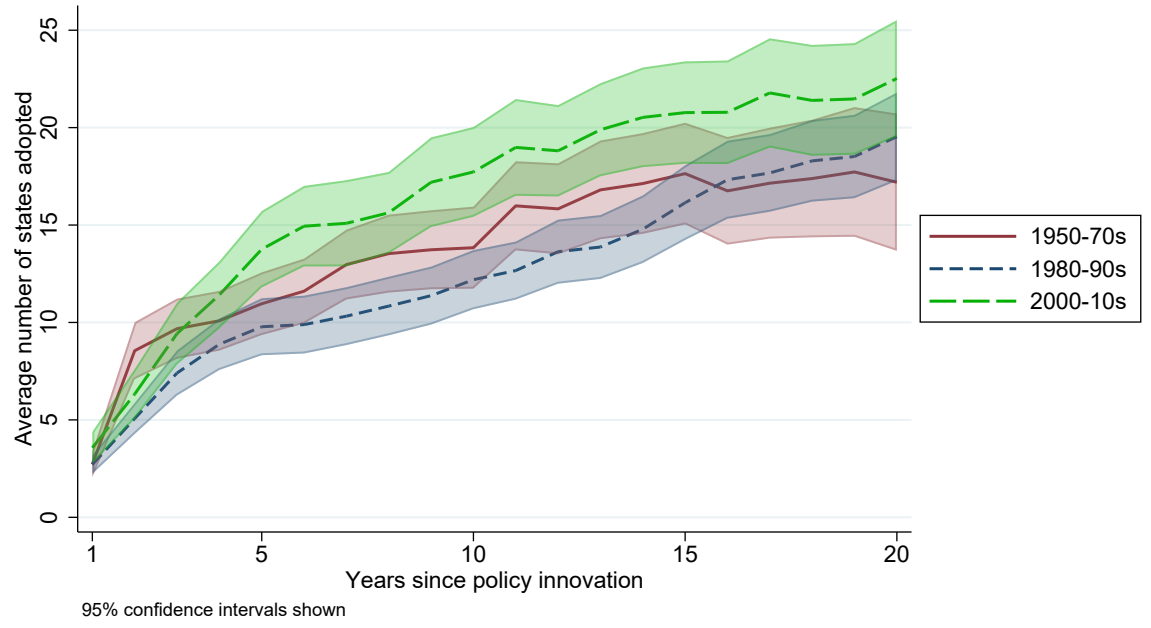
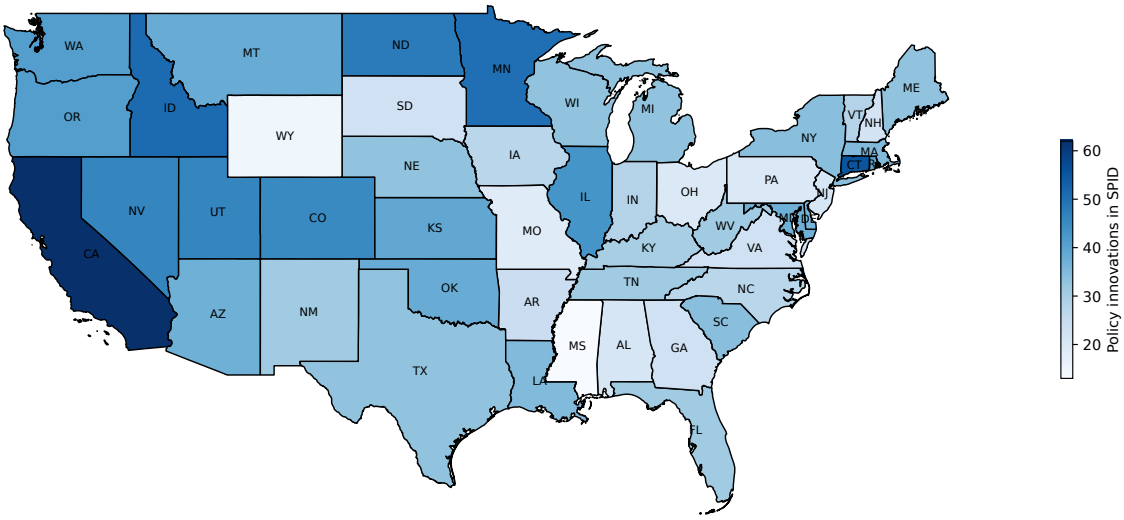


Figure C.2: Innovating states

(a) SPID policies



(b) NBER policies

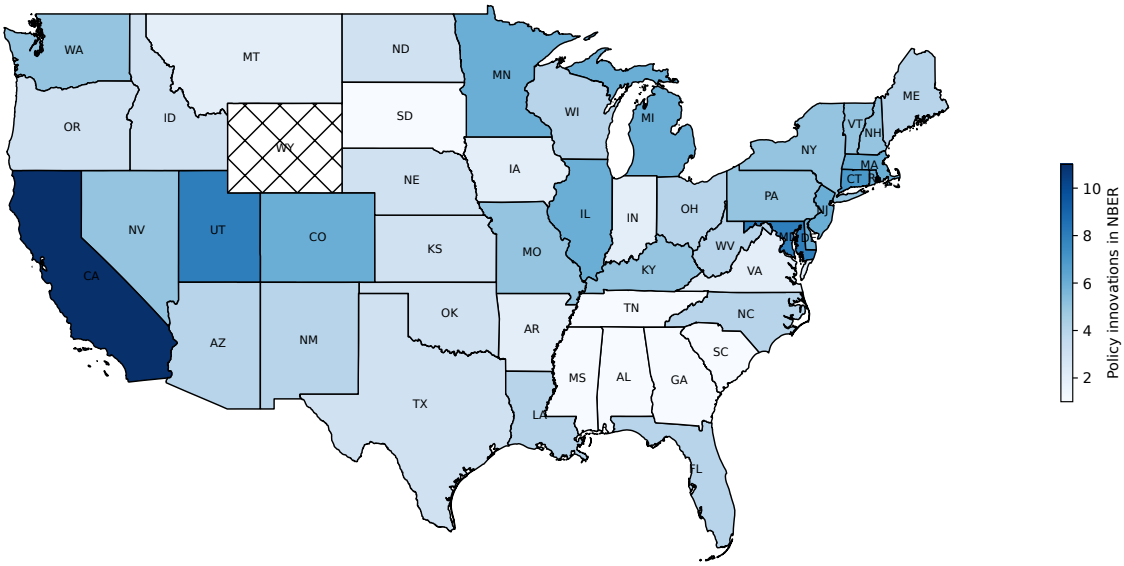
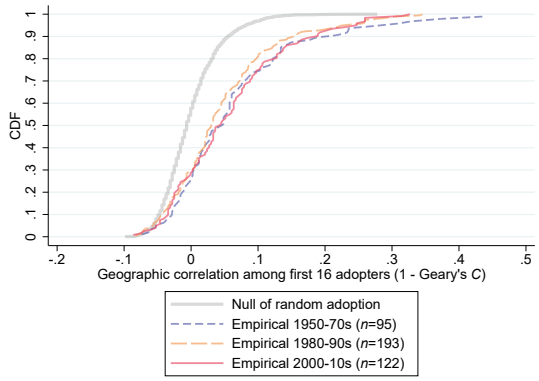
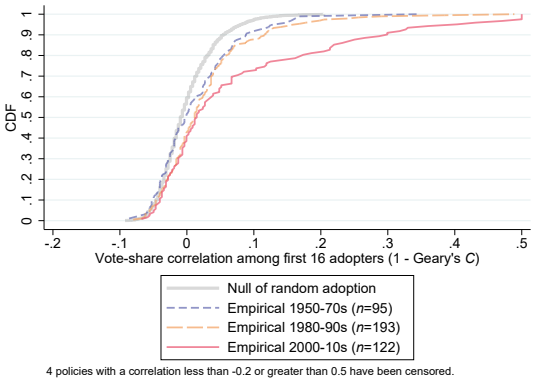


Figure C.3: Correlation in geography and politics among adopters (alternate thresholds)

(a) Correlation in geographic distance (first **16** adopters)

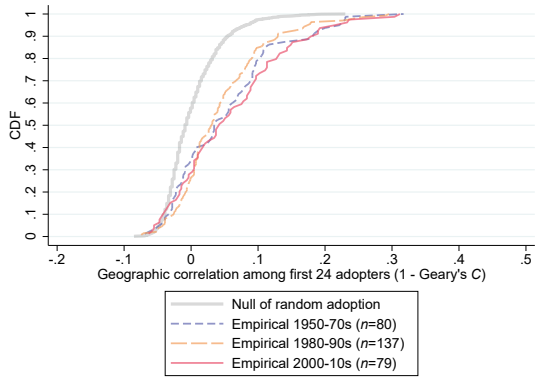


(b) Correlation in Republican vote-share (first **16** adopters)

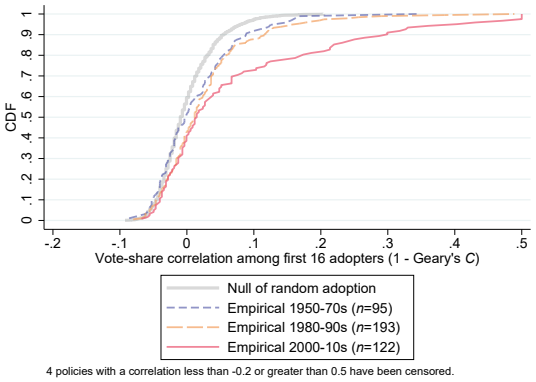


4 policies with a correlation less than -0.2 or greater than 0.5 have been censored.

(c) Correlation in geographic distance (first **24** adopters)



(d) Correlation in Republican vote-share (first **24** adopters)



4 policies with a correlation less than -0.2 or greater than 0.5 have been censored.

This figure replicates the analysis in Figures 4.5a-4.5b, but instead uses a threshold of the first 16 (Figures C.3a-C.3b) and of the first 24 (Figures C.3c-C.3d) adopters of a policy. The assignment of each policy to a decade is held constant at the year in which it reached 10 adopters. The sample of policies shrinks with higher thresholds as there are fewer policies that reach those thresholds.

Figure C.4: Stability of closest thirds in each dimension

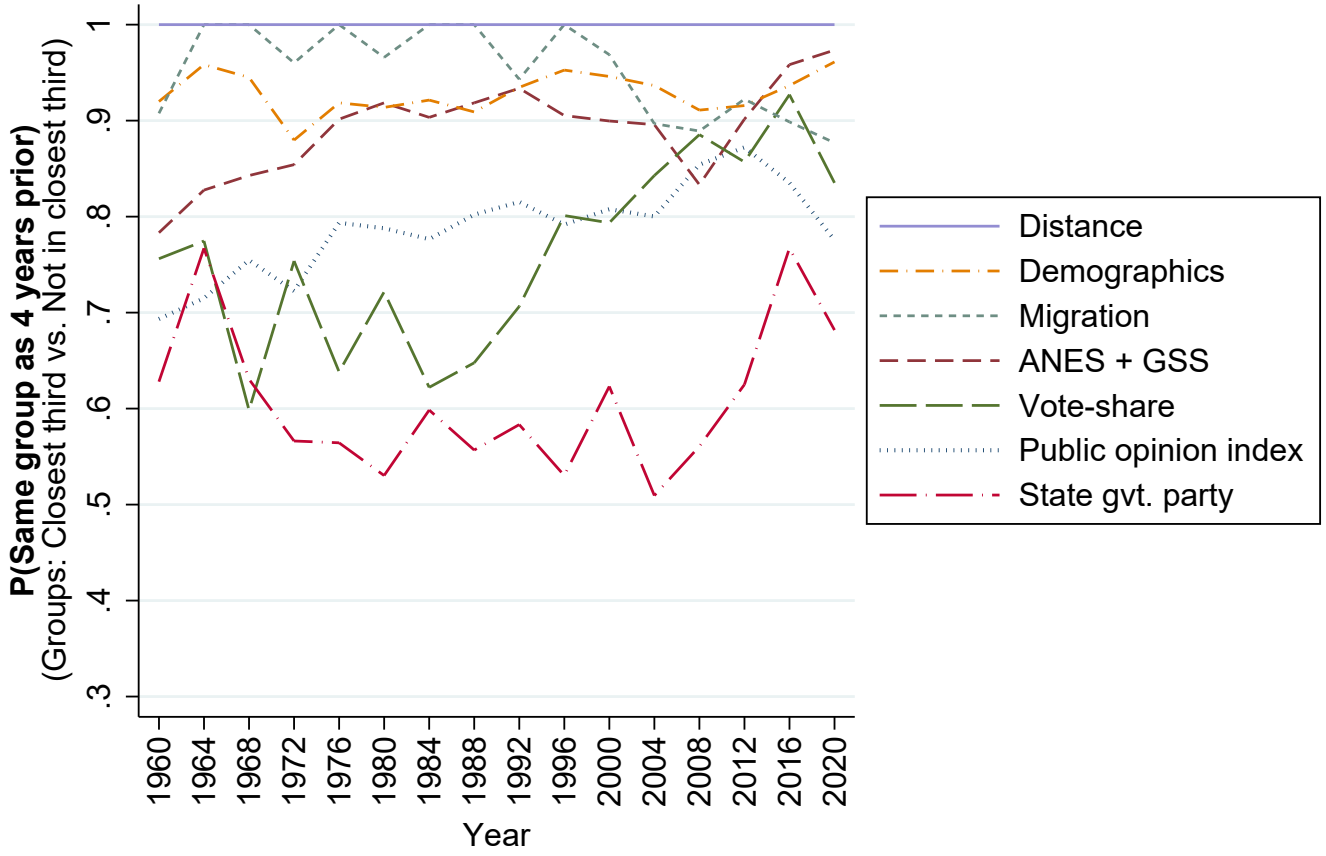
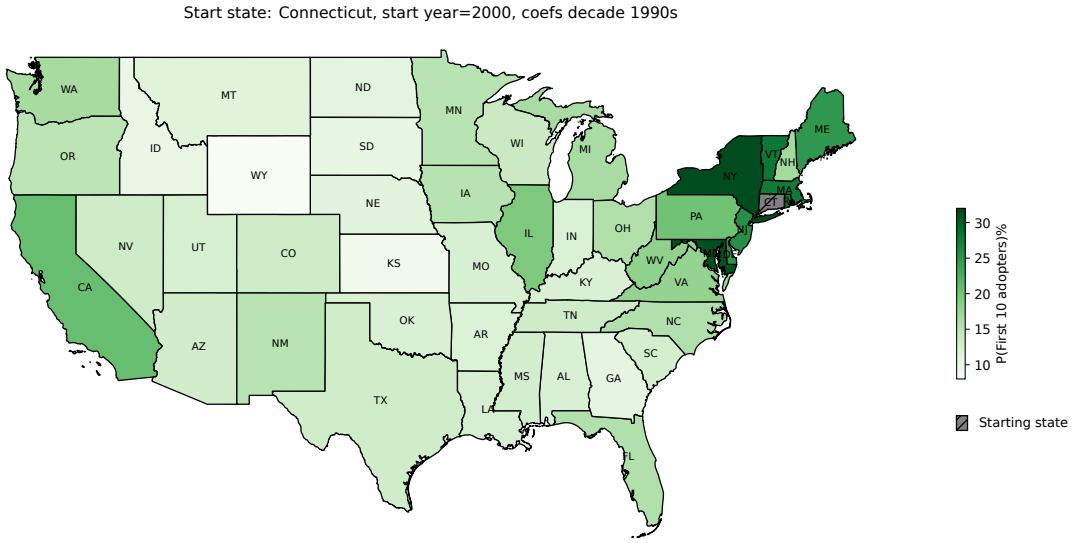


Figure C.5: Simulated policy diffusion

(a) Coefficients from 1990s (Connecticut)



(b) Coefficients from 2010s (Connecticut)

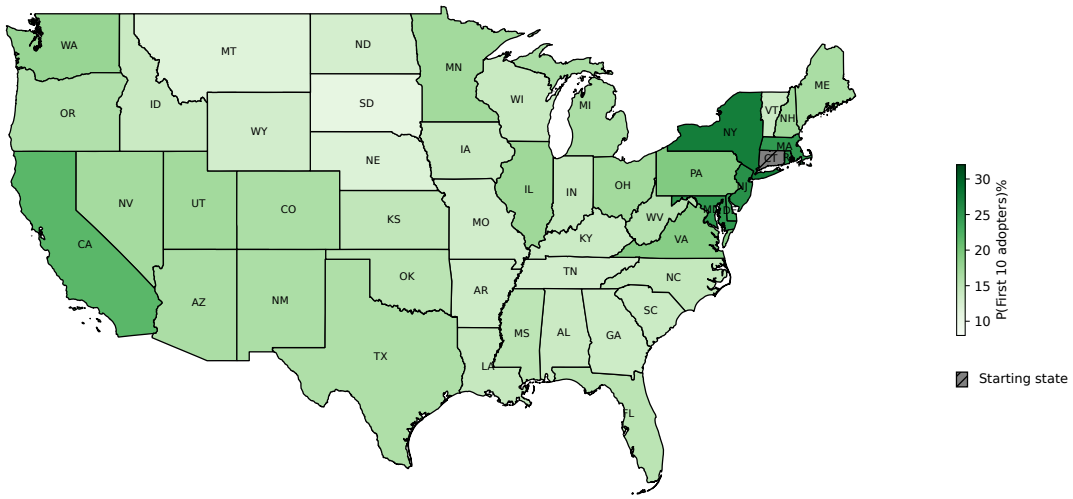
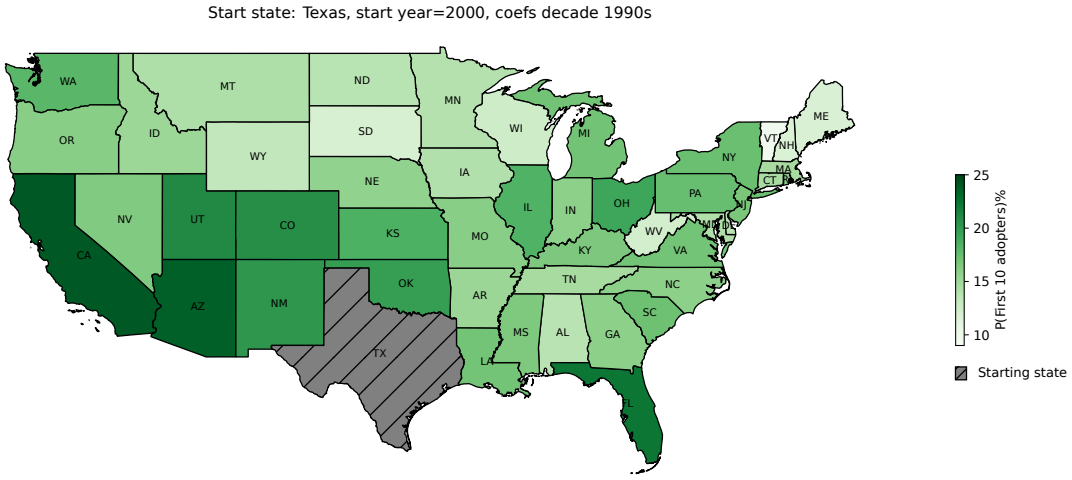


Figure C.5: Simulated policy diffusion

(c) Coefficients from 1990s (Texas)



(d) Coefficients from 2010s (Texas)

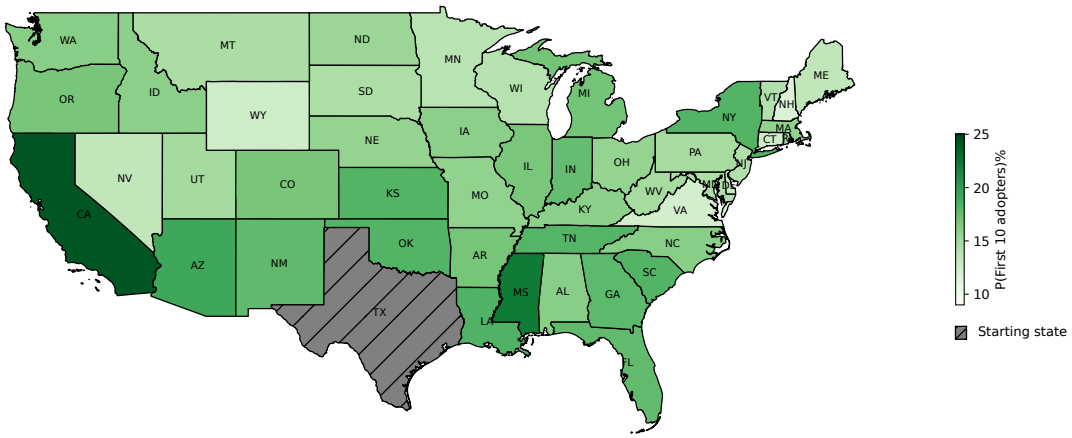
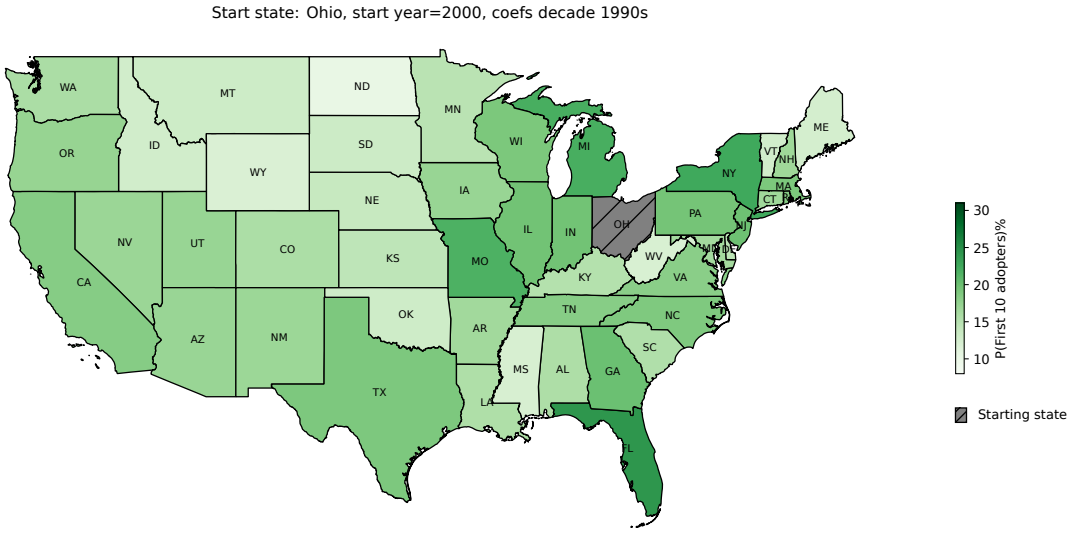


Figure C.5: Simulated policy diffusion

(e) Coefficients from 1990s (Ohio)



(f) Coefficients from 2010s (Ohio)

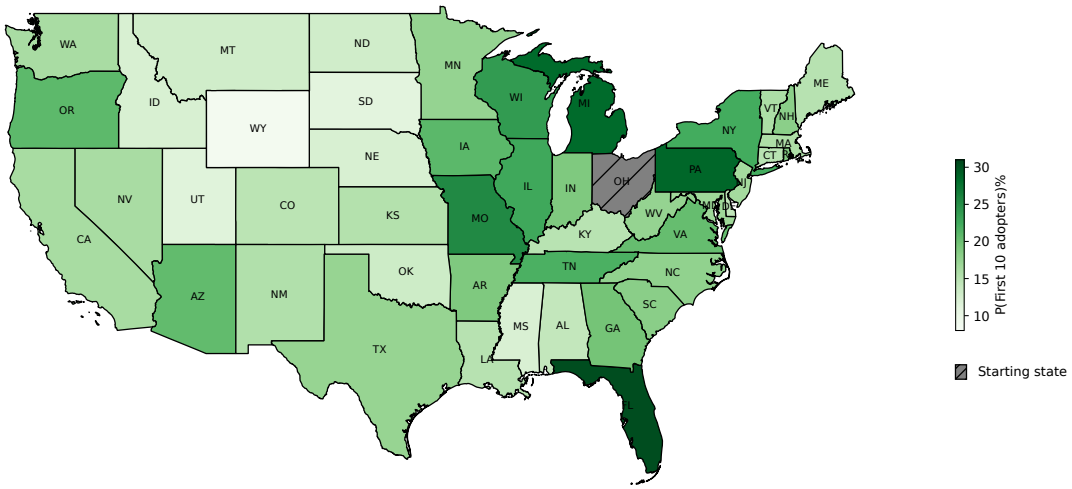
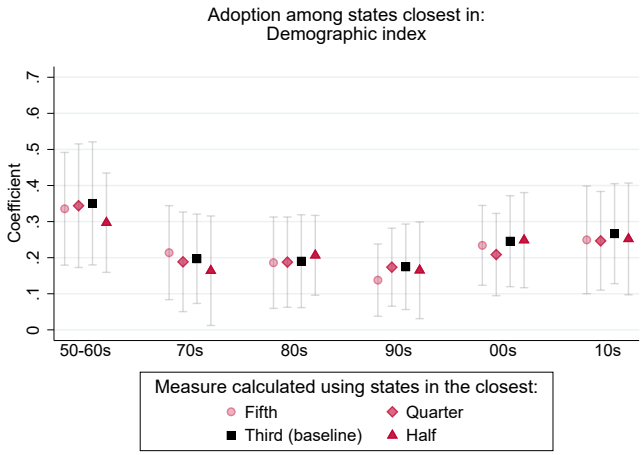
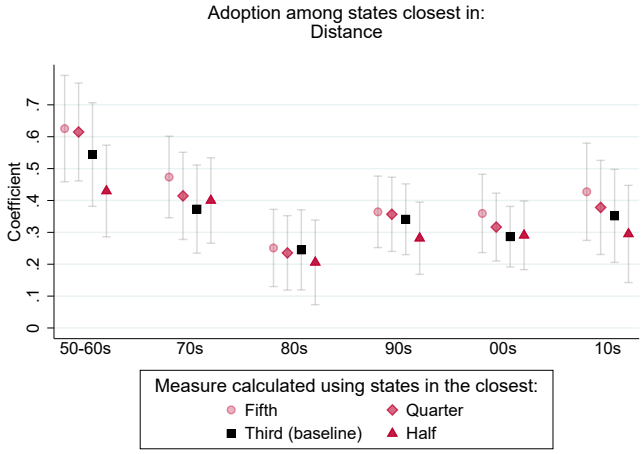


Figure C.6: Robustness checks: Threshold of closest states

(a) Dimension: Demographic index



(b) Dimension: Distance



(c) Dimension: Republican vote-share

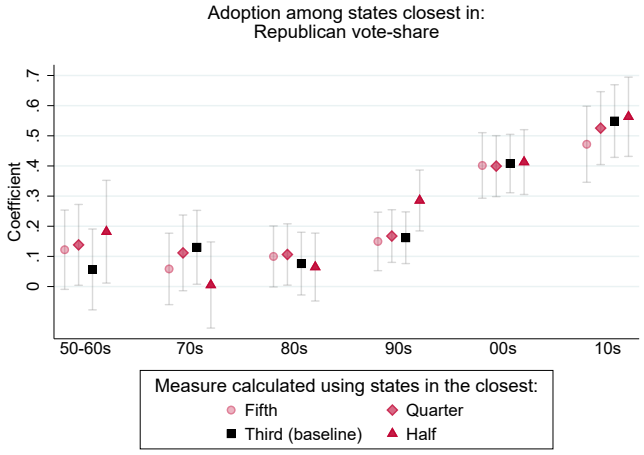
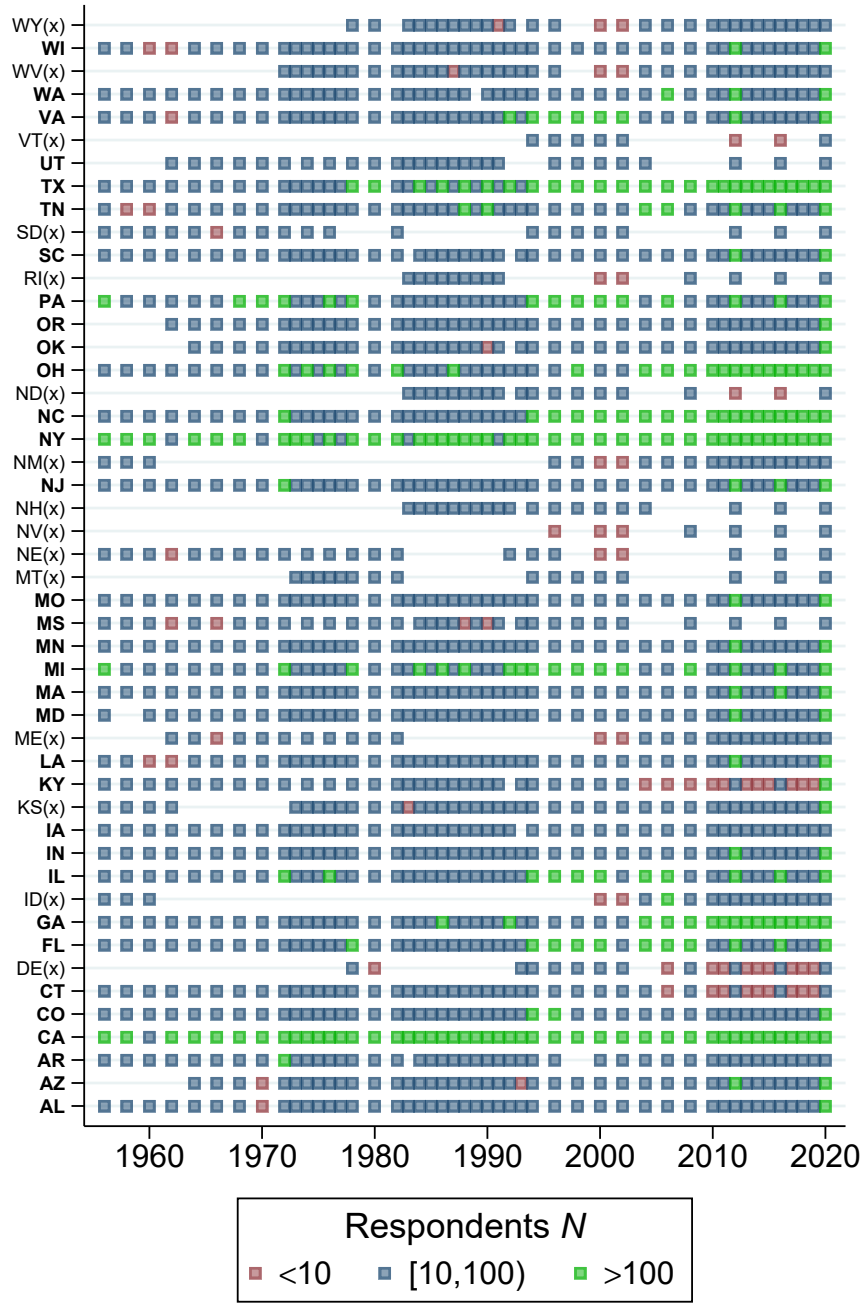


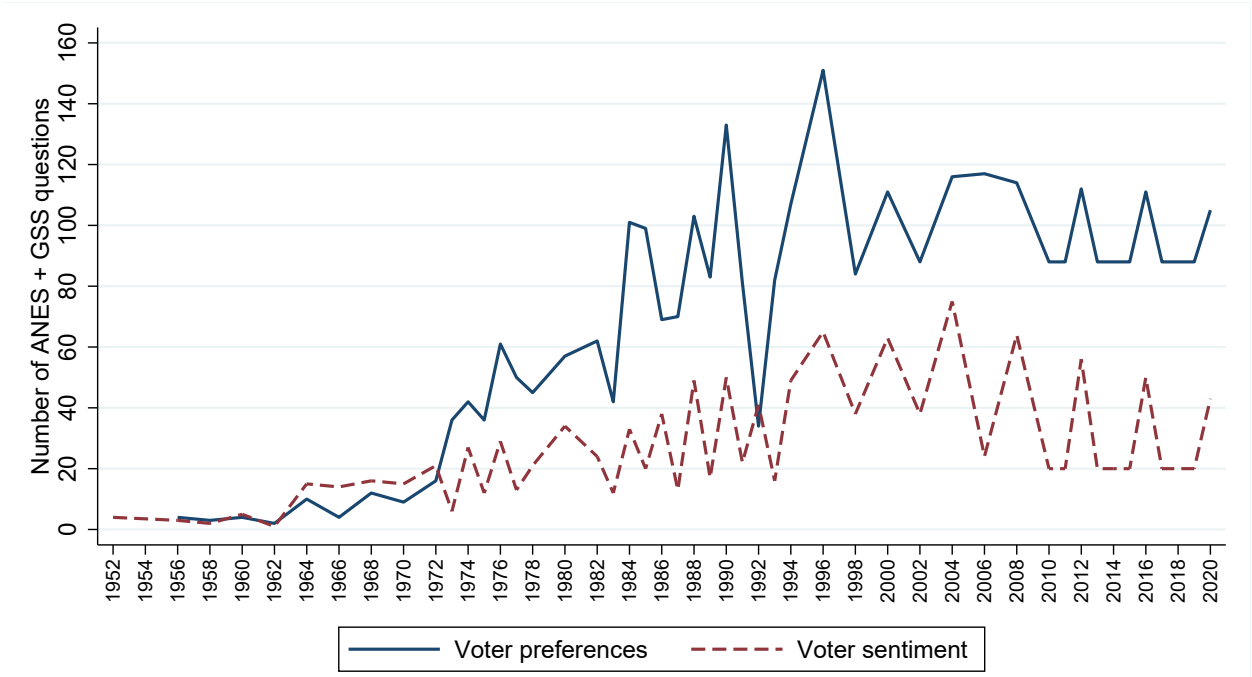
Figure C.7: Measuring voter policy preference from ANES and GSS

(a) State representation in ANES and GSS



States marked by "(x)" in Figure C.7a are excluded from the analysis using this voter preference measure due to insufficient coverage.

(b) Number of questions



(c) Questions by policy area

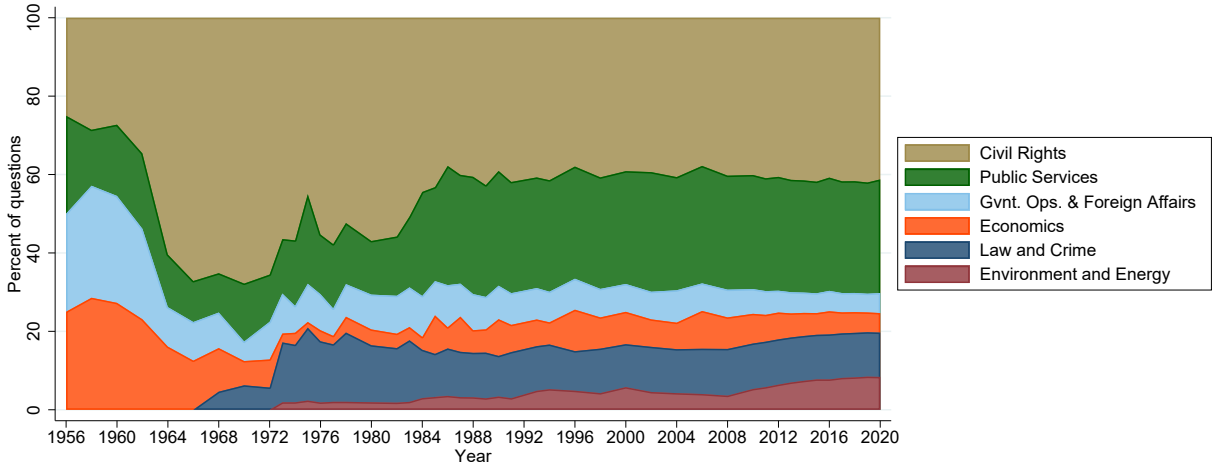
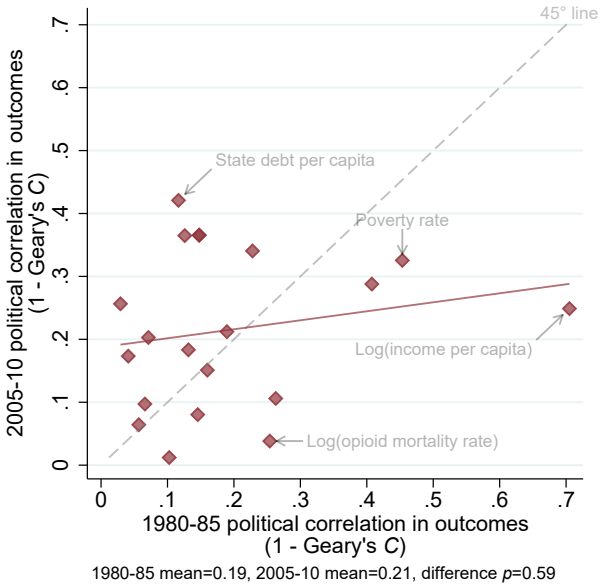


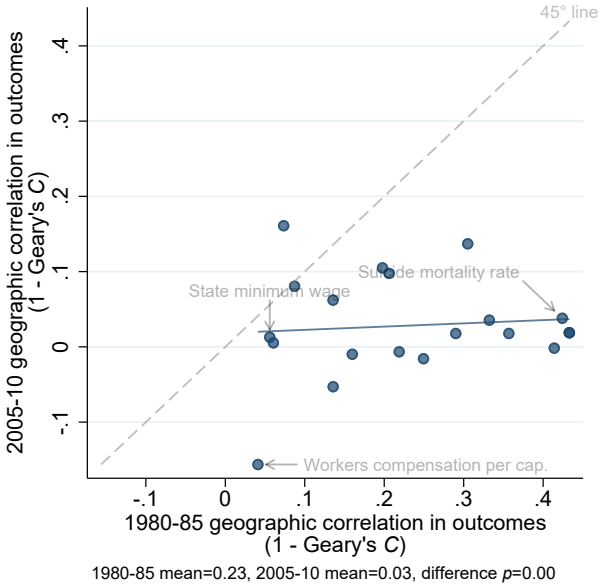
Figure C.8: Correlation of policy outcomes: 1980-85 vs. 2005-10

(a) Political correlation



1980-85 mean=0.19, 2005-10 mean=0.21, difference p=0.59

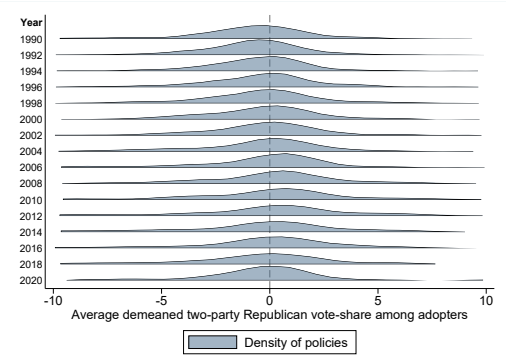
(b) Geographic correlation



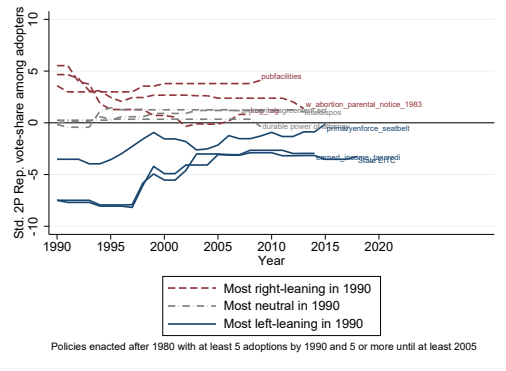
1980-85 mean=0.23, 2005-10 mean=0.03, difference p=0.00

Figure C.9: Categorizing the ideology of policies for event study analysis

(a) Distribution of policy ideologies



(b) Ideology of most extreme and neutral policies over time



(c) Number of policies under each ideology as a function of the threshold

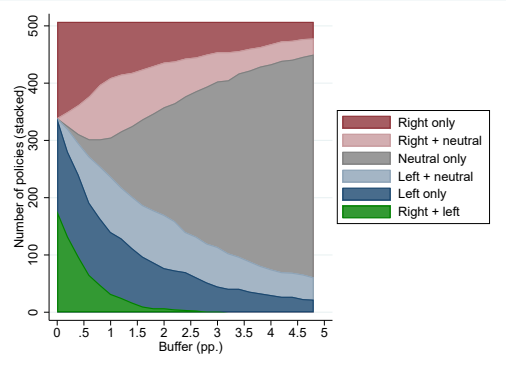
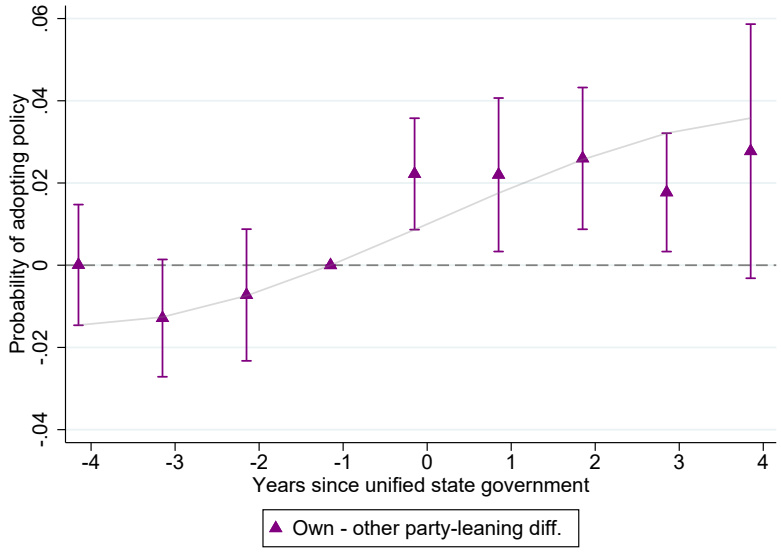
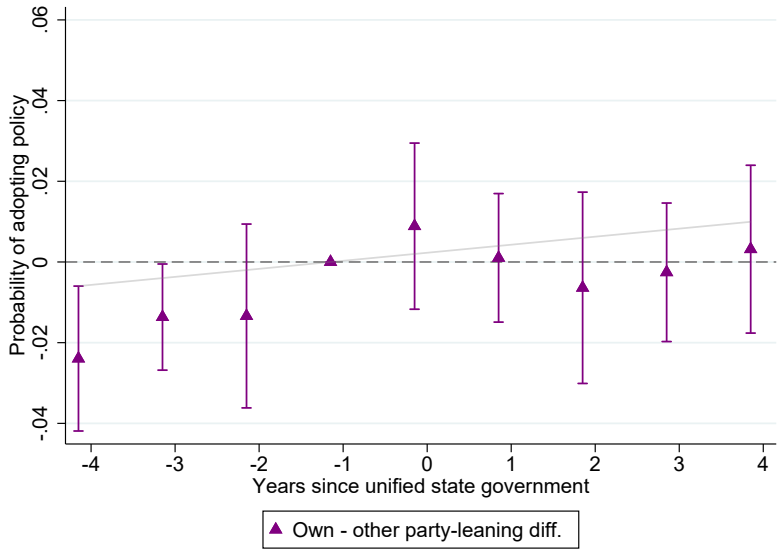


Figure C.10: Event study from switches in state government party control

(a) Events during 1990-2020



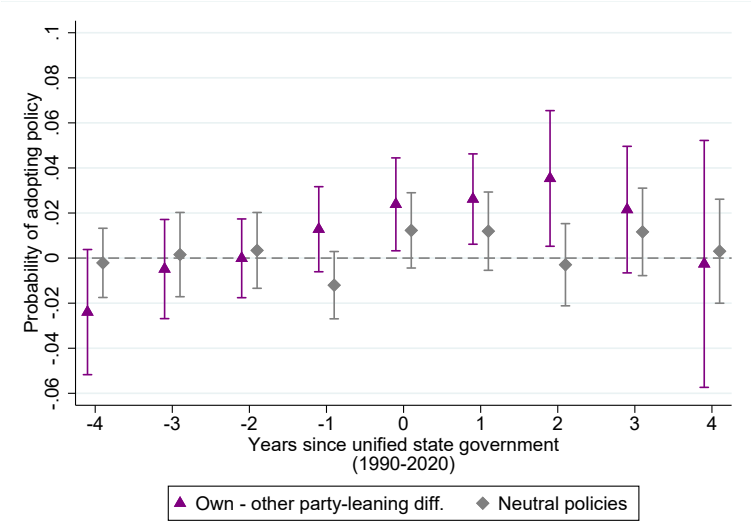
(b) Events during 1950-1989



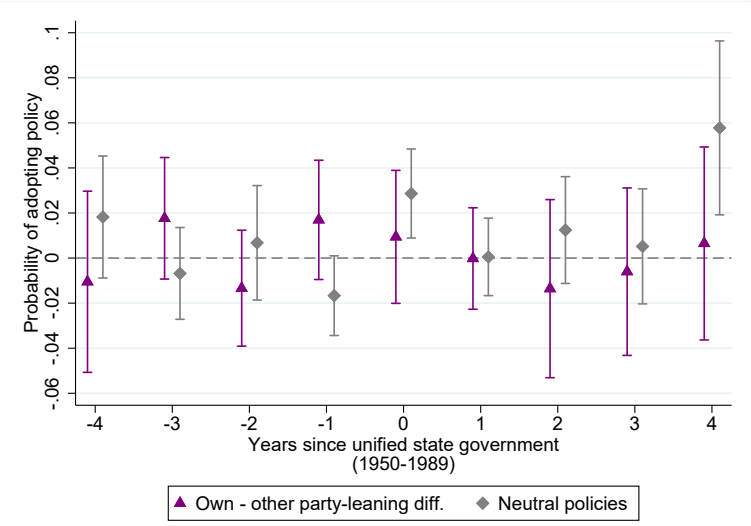
These figures show the event studies estimates from Figures 4.8a-4.8b with the most plausible confound path (Freyaldenhoven et al., forthcoming) traced in the gray curve.

Figure C.10: Event study from switches in state government party control (de Chaisemartin and D'Haultfoeuille estimator)

(c) Events during 1990-2020



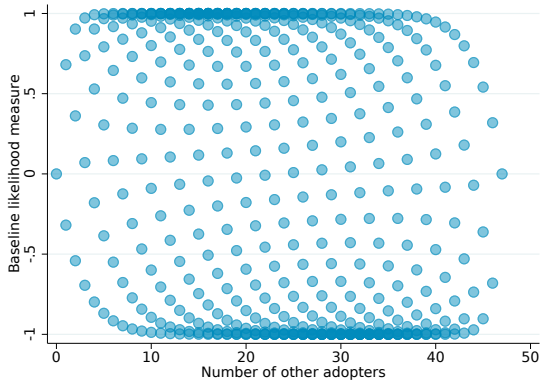
(d) Events during 1950-1989



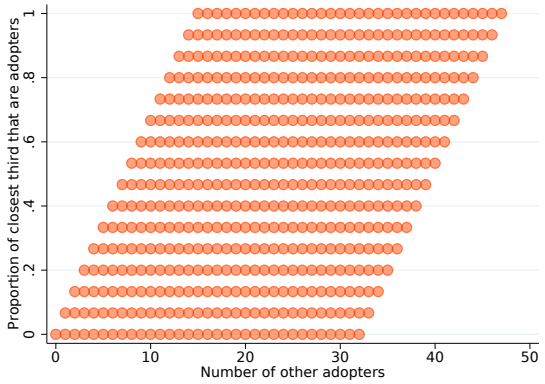
These figures show the event study estimates for switches to unified party control of state governments from the de Chaisemartin and D'Haultfoeuille (2020) estimator. Spells of unified state party control are defined as the 4 years prior to and after the switch to the unified government. During these spells for each state, policies are categorized as aligning with the party in control, not aligning, or neither, based on the procedure described in Footnote 10. For example, the policy of medical marijuana legalization is a left-learning policy based on the standardized vote-share of past adopters. In 2011, Alabama had a unified Republican state government, and thus medical marijuana legalization is categorized as a policy not in alignment with the party in control. On the other hand, Massachusetts had a unified Democratic state government, and thus the policy is categorized as being aligned with the party in control. Outside these event windows of unified state governments, all policies for that state are categorized as neutral. To run the event study using the de Chaisemartin and D'Haultfoeuille estimator, the state-policy-year panel is collapsed to the state-year average rates of adoption for aligned, not aligned, and neutral policies. The treated effects from switching to unified state governments are estimated separately for these three types of policies. 95% confidence intervals are shown from bootstrap standard errors resampling at the state level.

Figure C.11: Specification checks: Range of measure by number of adopters

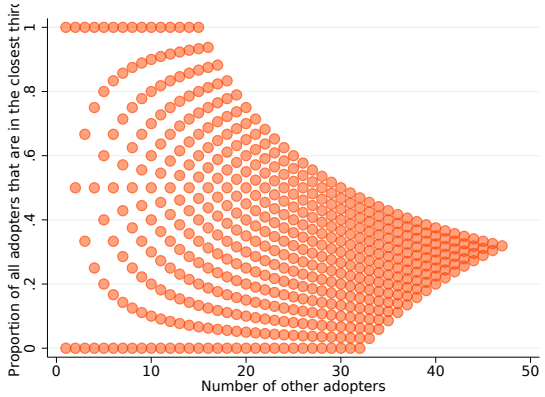
(a) Baseline two-sided likelihood measure



(b) Proportion of states in the closest third that are adopters



(c) Proportion of all adopters that are in the closest third



(d) Proportion of closest third that are adopters − Proportion of all states (excluding own) that are adopters

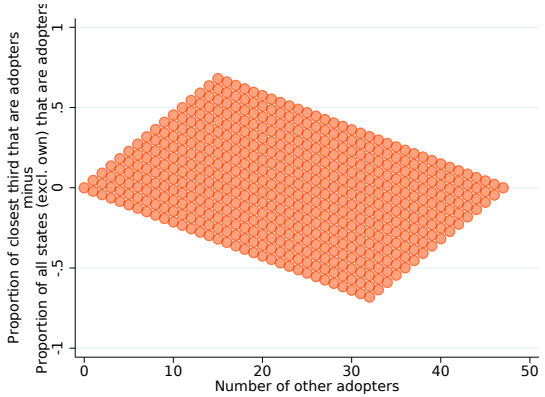
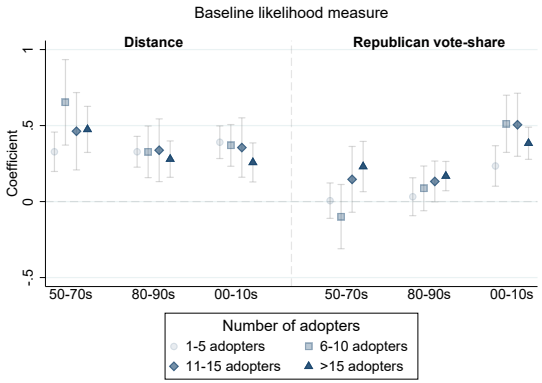
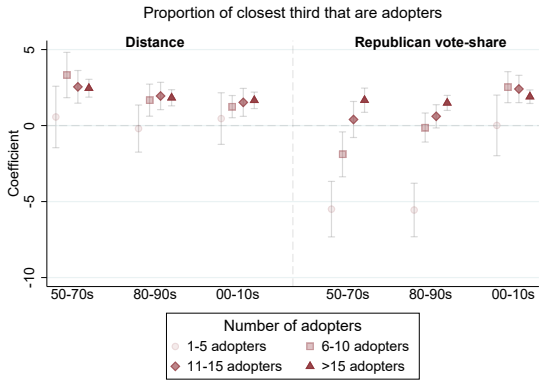


Figure C.11: Specification checks: Stability in coefficients by number of adopters

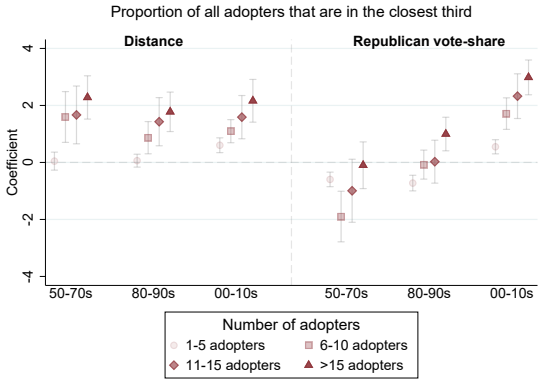
(e) Baseline two-sided likelihood measure



(f) Proportion of states in the closest third that are adopters



(g) Proportion of all adopters that are in the closest third



(h) Proportion of closest third that are adopters – Proportion of all states (excluding own) that are adopters

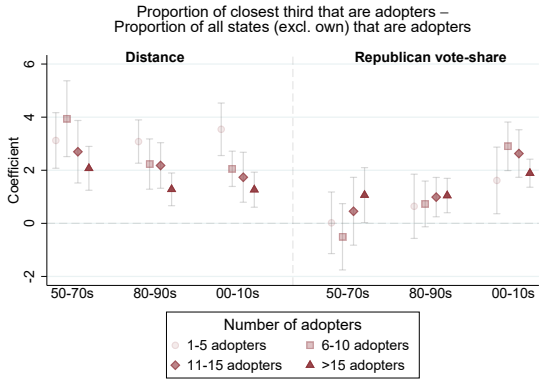


Table C.1: Summary of recent papers on policy diffusion in political science and examples in economics

Paper	(1) No. of Policies	(2) Time Period	(3) Source of Policies	(4) Methodology	(5) Determinants	(6) Quantitative Comparison across Determinants	(7) Main Results
DellaVigna and Kim (2022)	705	1950-2020	SPID, self-collected from NBER working papers	Logit hazard model	Geographic distance, Demographics, Migration, State government partisanship, Voter preferences	Yes	1950-1990s: Geographic and demographic diffusion, 2000-2020: Political diffusion
Desmarais, Harden, and Boehmke (2015)	151	1960-1999	Boehmke and Skinner (2012)	Multilevel logit model (dyadic)	Geographic contiguity, Ideology, Legislative professionalism, Diversity, Income, Population, State government partisanship	No	Diffusion predicted by: Geographic proximity; Citizen ideology; Demographics; Legislative professionalism difference; State party control
Caughey, Warshaw, and Xu (2017)	148	1936-2014	Caughey and Warshaw (2016)	Regression discontinuity, Dynamic panel analysis	State party control	No	Pre-2000: State party had little impact, Post-2000: State party control has strong impact on policy liberalism
Caughey and Warshaw (2018)	148	1936-2014	Caughey and Warshaw (2016)	Dynamic panel analysis	Mass liberalism, State party control, Suffrage, Campaign contribution limits, Reforms for citizen participation, Legislative professionalism	No	Role of public opinion has increased over time
Grumbach (2018)	135	1970-2014	Self-collected, Jordan and Grossmann (2016), Caughey and Warshaw (2016), Boehmke and Skinner (2012)	Dynamic panel analysis	State party control	No	Pre-2000: Little polarization, Post-2000: Substantial polarization in certain policy areas
Maillinson (2021)	556	1960-2014	SPID	Multilevel logit model	Geographic distance, Ideological similarity, Congressional hearings, Initiative availability, Initiative qualification difficulty, Legislative professionalism, Slack resources (per capita income and population), State party control, Salience, Policy complexity	No	Geographic diffusion has decreased over time, Role of ideology has remained stable
Case, Rosen, and Hines Jr. (1993)	1	1970-1985	Direct state expenditures per capita	Two-way fixed effects	Expenditures of similar states	Yes	Expenditure by other states similar in geography, income, and racial composition predict own state's expenditures
Strumpf and Oberholzer-Gee (2002)	1	1934-1970	Liquor control	Latent taste model for liquor controls; Probit estimation of policy choice on latent taste heterogeneity	Preference heterogeneity within state	No	Greater heterogeneity in local taste for liquor controls predicts the state adopting decentralized liquor laws

Table C.2a: SPID sample examples

Number	Source	Description	Area	Adoptions	First year	Last year
4	Boehmke-Skinner	Abortion Pre-Roe	Civil Rights	16	1966	1972
31	Uniform Law	Framework For Donation Of Organs Other Body Parts (1987 Version)	Public Services	19	1989	2003
44	Biggers	Photo Id Voting Requirement	Civil Rights	20	1997	2013
49	Walker	Architects Licensing	Economics	47	1897	1951
61	Boehmke-Skinner	Ban On Financial Incentives For Doctors To Perform Less Costly Procedures/Prescribe Less Costly Drugs	Public Services	28	1996	2001
65	Sheprd	Requiring Helmet For Riding Bike	Public Services	20	1989	2007
67	Walker	Aid To The Blind (Social Security)	Public Services	48	1936	1953
70	Boushey	Short-Term Programs For Incarcerated Youth (Similar To Military School)	Law and Crime	22	1982	1999
77	Min Hee Go	Building Code Adoption	Public Services	43	1953	2010
102	Kreitzer	No-Protest Zone Around Abortion Clinic	Civil Rights	16	1973	2005
117	Karch	Provides All The Benefit Of Adoption Subsidy Agreement, Regardless Of State	Law and Crime	21	1984	2002
157	Uniform Law	Authorizes Courts To Adjudicate Actual Controversies Concerning Legal Rights And Duties Even Though Traditional Remedies For Damages Or Equitable Relief Are Not Available.	Law and Crime	41	1922	2008
161	Lacy	Comprehensive Remediation Reform	Public Services	17	1988	2009
166	Lacy	Placement Policies (Placement Examination, Changes To Placement Criteria	Public Services	3	1997	2008
176	Other	Notification Of Sex Offenders Is At Authority'S Discretion	Law and Crime	8	1995	2006
186	Caughey-Warshaw	Is It Legal To Use Marijuana For Medical Purposes?	Public Services	35	1996	2022
213	Uniform Law	Regulating Condemnation Of Property On Behalf Of Private And Public Entities	Public Services	1	1985	1985
219	Uniform Law	Allow Conversion Of One Kind Of Business Organization To Another, Or The Merger Of Two Or More Business Organizations Into One Organization	Economics	6	2009	2017
225	Caughey-Warshaw	Does The State Have A Public Benefit Fund For Renewable Energy And Energy Efficiency?	Environment and Energy	19	1996	2005
231	Caughey-Warshaw	Is There A State Estate Tax?	Economics	23	2009	2009
288	Caughey-Warshaw	Does The State Require Background Checks For Private Handgun Sales?	Law and Crime	13	1935	1996
295	Boehmke-Skinner	Harassment Crime	Law and Crime	10	1998	2001
333	Other	Reforms The Process Defendants Use To Claim Insanity, Including The Standard Or Burden Of Proof	Law and Crime	34	1975	1998
344	Other	Internet Registry Of Sex Offenders	Law and Crime	15	1997	2004
392	Uniform Law	Regulation On Promotional Sale Of Land	Economics	7	1967	1973
425	Lacy	Merit Based Aid Program	Public Services	12	1993	2005
442	Boehmke-Skinner	Voter Registration With Driver'S License Renewal	Civil Rights	47	1976	1995
490	Boehmke-Skinner	Interstate Pest Control Compact	Economics	36	1968	2009
534	Caughey-Warshaw	Does The State Have Fair Employment Laws	Civil Rights	41	1965	1992
536	Caughey-Warshaw	Does The State Have Fair Housing Laws For Owners	Civil Rights	13	1959	1968
571	Boushey	Repeal Sodomy Law	Civil Rights	48	1962	2003
574	Uniform Law	Regulating Mortgage Holder Rights	Economics	4	2005	2014
582	Boushey	Program To Combat Organized Crime	Law and Crime	29	1972	1995
620	Uniform Law	Defines Rights Of Convicted People	Law and Crime	1	1968	1968
651	Other	State Passes Truth-In-Sentencing Law Before 1997	Law and Crime	25	1911	1996
676	Uniform Law	Governs All Unincorporated Nonprofit Associations That Are Formed Or Operate In A State	Economics	5	2009	2015
698	Caughey-Warshaw	Does The State Mandate Ten Commandments Be Posted In Public Institutions/Schools?	Civil Rights	2	1935	1978
704	Caughey-Warshaw	Does The State Ban Discrimination In Public Accommodations On The Basis Of Gender Id?	Civil Rights	16	1993	2014
706	Caughey-Warshaw	Does The State Ban Discrimination In Public Accommodations On The Basis Of Sexual Orientation?	Civil Rights	20	1989	2009
714	Karch	Provides A Means Through Which States Can Participate In A Reciprocal Program To Enforce Wildlife Citations	Environment and Energy	20	1989	2014

Table C.2b: NBER working paper sample

Number	Policy	Title	Area	Adoptions	First year	Last year
18187	Stand Your Ground laws	Stand Your Ground Laws, Homicides, and Injuries	Law and Crime	25	1994	2009
18299	Leave for state employee organ donors	Removing Financial Barriers to Organ and Bone Marrow Donation: The Effect of Leave and Tax Legislation in the U.S.	Public Services	29	1989	2007
18341	Physical education requirement	The Impact of Physical Education on Obesity among Elementary School Children	Public Services	38	1940	2007
18516	Wrongful discharge laws	Wrongful Discharge Laws and Innovation	Economics	45	1970	1998
18773	Bicycle helmet laws	Effects of Bicycle Helmet Laws on Children's Injuries	Public Services	19	1987	2006
18887	AFDC waiver	Effects of Welfare Reform on Women's Crime	Economics	27	1992	1996
18887	TANF	Effects of Welfare Reform on Women's Crime	Economics	48	1996	1998
19294	Biotech tax incentives	State Incentives for Innovation, Star Scientists and Jobs: Evidence from Biotech	Economics	7	1984	2003
19904	Community rating regulations	Regulatory Redistribution in the Market for Health Insurance	Public Services	7	1993	1997
20149	Interstate bank branching laws	Does Financing Spur Small Business Productivity? Evidence from a Natural Experiment	Economics	48	1995	1997
20565	Medical record copy fee cap	Expanding Patients' Property Rights In Their Medical Records	Public Services	42	1972	2007
20808	NOx cap-and-trade	Who Loses Under Power Plant Cap-and-Trade Programs?	Environment and Energy	20	2003	2007
21170	Commonsense Consumption Acts	Do "Cheeseburger Bills" Work? Effects of Tort Reform for Fast Food	Economics	26	2003	2013
21345	Medical marijuana laws	Do Medical Marijuana Laws Reduce Addictions and Deaths Related to Pain Killers?	Public Services	21	1996	2014
21373	Individual income tax	Broadening State Capacity	Economics	42	1911	1971
21373	Corporate income tax	Broadening State Capacity	Economics	43	1911	1971
22344	Nurse Licensure Compact	Labor Supply Effects of Occupational Regulation: Evidence from the Nurse Licensure Compact	Public Services	25	1999	2015
22899	Initial Medicaid implementation	The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes	Public Services	48	1966	1982
23171	Good Samaritan Law	With a Little Help from My Friends: The Effects of Naloxone Access and Good Samaritan Laws on Opioid-Related Deaths	Public Services	45	2007	2019
23171	Naloxone Access Law	With a Little Help from My Friends: The Effects of Naloxone Access and Good Samaritan Laws on Opioid-Related Deaths	Public Services	48	2001	2017
23313	E-cigarette minimum age law	The Effects of E-Cigarette Minimum Legal Sale Age Laws on Youth Substance Use	Public Services	48	2010	2016
23388	Substance use disorder parity laws	Health Insurance and Traffic Fatalities: The Effects of Substance Use Disorder Parity Laws	Public Services	12	1994	2009
23510	Concealed handgun carry law	Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Control Analysis	Law and Crime	41	1959	2014
23995	Smoking ban	Impact of Comprehensive Smoking Bans on the Health of Infants and Children	Public Services	34	1994	2012
24153	Interstate tax audit info sharing	Intergovernmental Cooperation and Tax Enforcement	Government Operations	5	1950	1955
24259	Right-to-work laws	From the Bargaining Table to the Ballot Box: Political Effects of Right to Work Laws	Economics	27	1943	2017
24381	Ban-the-box laws	Do Ban the Box Laws Increase Crime?	Economics	11	2009	2014
24651	Same-sex marriage	Effects of Access to Legal Same-Sex Marriage on Marriage and Health: Evidence from BRFSS	Civil Rights	33	2004	2014
24662	Merit-aid programs	State Merit Aid Programs and Youth Labor Market Attachment	Public Services	18	1988	2005
24782	Duty-to-bargain laws	The Long-run Effects of Teacher Collective Bargaining	Economics	31	1960	1987

Table C.2b: NBER working paper sample

Number	Policy	Title	Area	Adoptions	First year	Last year
24986	Community eligibility provision	School Nutrition and Student Discipline: Effects of Schoolwide Free Meals	Public Services	10	2012	2014
25209	Child gun access prevention laws	Child Access Prevention Laws and Juvenile Firearm-Related Homicides	Law and Crime	25	1989	2001
25369	Age anti-discrimination	Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment	Economics	45	1934	1997
25369	Disability anti-discrimination	Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment	Economics	46	1971	1988
25390	Wind energy incentives	Technological Spillover Effects of State Renewable Energy Policy: Evidence from Patent Counts	Environment and Energy	48	2000	2011
25758	Minor abortion parental consent	The Impact of Parental Involvement Laws on Minor Abortion	Public Services	37	1974	2013
25974	Initial prescription drug monitoring	Can Policy Affect Initiation of Addictive Substance Use? Evidence from Opioid Prescribing	Public Services	24	1988	2018
25974	Must-access prescription drug monitoring	Can Policy Affect Initiation of Addictive Substance Use? Evidence from Opioid Prescribing	Public Services	29	2007	2019
26017	E-cigarette tax	The Effects of Traditional Cigarette and E-Cigarette Taxes on Adult Tobacco Product Use	Public Services	7	2010	2017
26135	Pill mill laws	Mortality and Socioeconomic Consequences of Prescription Opioids: Evidence from State Policies	Public Services	8	2005	2014
26140	NBCCEDP cancer screenings	Effects of Direct Care Provision to the Uninsured: Evidence from Federal Breast and Cervical Cancer Programs	Public Services	48	1991	1999
26206	Strict voter ID	Strict Voter Identification Laws, Turnout, and Election Outcomes	Civil Rights	11	2004	2016
26405	State EITC	The EITC and the Extensive Margin: A Reappraisal	Economics	28	1986	2018
26500	Triplicate prescription	Origins of the Opioid Crisis and Its Enduring Impacts	Public Services	7	1939	1988
26676	E-verify for employment	States Taking the Reins? Employment Verification Requirements and Local Labor Market Outcomes	Economics	22	2006	2015
26749	Modern prescription drug monitoring	Effect of Prescription Opioids and Prescription Opioid Control Policies on Infant Health	Public Services	47	1999	2017
26777	Anti-bullying laws	Anti-Bullying Laws and Suicidal Behaviors among Teenagers	Law and Crime	48	2001	2015
26832	Mandated sick pay	Mandated Sick Pay: Coverage, Utilization, and Welfare Effects	Economics	10	2011	2018
27054	Salary history ban	Information and the Persistence of the Gender Wage Gap: Early Evidence from California's Salary History Ban	Economics	12	2017	2021
27306	Medicaid expansion	Medicaid Expansion and the Mental Health of College Students	Public Services	36	2014	2021
27520	Tramadol as Schedule IV drug	Competitive Effects of Federal and State Opioid Restrictions: Evidence from the Controlled Substance Laws	Public Services	12	2007	2014
27728	2003 standard certificate of live birth implementation	Heterogeneous Effects Of Health Insurance On Birth Related Outcomes: Unpacking Compositional Vs. Direct Changes	Public Services	48	2003	2016
27788	Paid family leave	Paid Leave Pays Off: The Effects of Paid Family Leave on Firm Performance	Economics	6	2002	2018
28173	Tobacco 21 laws	Do State Tobacco 21 Laws Work?	Public Services	15	2016	2019
28903	Right of workers to talk law	Equilibrium Effects of Pay Transparency	Economics	12	2004	2016
29087	Recreational marijuana legalization	Recreational Marijuana Laws and the Use of Opioids: Evidence from NSDUH Microdata	Public Services	17	2012	2021
29318	CPA 150-hour rule	Occupational Licensing and Accountant Quality: Evidence from the 150-Hour Rule	Economics	48	1983	2015

Table C.3a: Summary statistics: Policy outcomes from NBER papers

Outcome	Coverage	Example NBER policy	NBER WP numbers
Log(income per capita)	1950-2020	Partial paid leave for pregnancy	26416, 19294
Voter turnout rate	1980-2019	Strict voter ID	26206, 24259
Log(opioid mortality rate)	1968-2014	Naloxone Access Law	23171, 25974, 26135, 26500, 27520, 29087
Employment rate in energy-intensive industry	1975-2018	NOx cap-and-trade	20808
Private insurance coverage rate	1987-2006	Community rating regulations	19904
Log(state revenue per capita)	1950-2016	Tax audit info sharing	21373, 24153
Log(state expenditure per capita)	1950-2016	State income and corporate taxes	21373, 24153
Average BMI	1987-2020	Physical education requirements	18341, 21170
Firearm mortality rate	1968-2016	Stand Your Ground laws	18187, 23510, 25209
Alcohol-induced traffic mortality rate	1975-2015	Substance use disorder parity laws	23388

Table C.3b: COVID-19 policies

Example policy	Coverage (MM/DD/YYYY)	Num. adopted states
Modify Medicaid requirements with 1135 waivers (date of CMS approval)	3/16/2020-4/22/2020	48
SNAP Waiver - Pandemic EBT during school year 2020-2021	12/15/2020-3/23/2021	25
Late Fee Ban Start	2/29/2020-5/22/2020	11
Date K-12 school employees became eligible for COVID-19 vaccination	1/8/2021-4/5/2021	48
Date banned visitors to nursing homes	3/9/2020-8/13/2020	30
Stopped visitation in state prisons x2	7/15/2020-12/30/2020	9
Date adults ages 55+ became eligible for COVID-19 vaccination	3/1/2021-4/19/2021	48
SNAP Waiver - Emergency Allotments to Current SNAP Households	3/24/2020-4/15/2020	48
Reopened bars (x2)	8/11/2020-5/7/2021	18
Face mask mandate in public spaces	4/8/2020-12/9/2020	38
SNAP Waiver - Temporary Suspension of Claims Collection	4/2/2020-5/13/2020	24
Face mask mandate in schools for 2021-22 school year	5/1/2020-4/16/2021	15
Closed movie theaters (x2)	6/29/2020-12/12/2020	6
Closed gyms (x2)	6/29/2020-12/12/2020	7
State of emergency issued	2/29/2020-3/16/2020	48
Reopened ACA enrollment using a special enrollment period	3/10/2020-4/1/2020	11
Date closed K-12 public schools	3/16/2020-4/3/2020	47
First eviction enforcement ban start	3/16/2020-4/30/2020	27
Utilities reconnection start	3/4/2020-4/13/2020	8
Date adults ages 75+ became eligible for COVID-19 vaccination	12/23/2020-2/15/2021	48
SNAP Waiver - Pandemic EBT during school year 2019-2020	4/9/2020-8/13/2020	48
Allowed restaurants to sell takeout alcohol	3/16/2020-5/8/2020	42
Allow audio-only telehealth	1/1/2020-6/22/2020	45
Exceptions to emergency oral prescriptions	3/11/2020-4/6/2020	6
Closed restaurants except take out	3/16/2020-4/3/2020	47
Date adults ages 40+ became eligible for COVID-19 vaccination	3/16/2021-4/19/2021	48
Reopened hair salons/barber shops	4/24/2020-8/28/2020	47
Date adults ages 50+ became eligible for COVID-19 vaccination	3/3/2021-4/19/2021	48
Reopened religious gatherings	4/26/2020-6/22/2020	34
Closed gyms	3/16/2020-4/3/2020	47
Average (all 76 policies)	6/30/2020-9/27/2020	30.62

This table shows 30 randomly selected COVID-19 policies in the data set as well as the overall average. Policies are kept in data set until the first repeal. Source: COVID-19 US State Policies (CUSP)

Table C.3c: Vaccine regulations

Policy	Coverage	Num. adopted states
Hepatitis A Vaccine Mandates for Child Care	1998-2021	22
Hepatitis A Vaccine Mandates for K-12	1988-2021	15
Hepatitis B Vaccine Mandates for Child Care	1993-2018	43
Hepatitis B Vaccine Mandates for Colleges and Universities	1992-2011	15
Hepatitis B Vaccine Mandates for elementary	1994-2008	44
Hepatitis B Vaccine Mandates for secondary	1995-2014	41
Influenza Vaccine Mandates for Child Care and Pre-K	1999-2020	7
MenACWY Vaccine Mandates for Colleges and Universities	2001-2020	23
MenACWY Vaccine Mandates for Elementary and Secondary Schools	2005-2021	33
PCV Vaccine Mandates for Childcare	2001-2018	39
Rotavirus Vaccine Mandates for Child Care and Pre-K	1999-2021	8
Tdap Vaccine Mandates for Elementary and Secondary Schools	2006-2017	48
Varicella Vaccine Mandates for Child Care	1997-2016	47
Varicella Vaccine Mandates for Elementary School	1998-2015	48
Varicella Vaccine Mandates for Middle/junior/senior high	1999-2015	39
Hep B vaccine is either offered or mandated in hospitals	1993-2016	8
Hep B vaccine is either offered or mandated in long-term care facilities	1993-2018	9
Hep B vaccine is either offered or mandated in ambulatory care facilities	1993-2016	12
Any of the MMR vaccines are either offered or mandated in hospitals	1980-2014	14
Any of the MMR vaccines are either offered or mandated in long-term care facilities	1981-2020	10
Any of the MMR vaccines are either offered or mandated in ambulatory care facilities	1992-2022	12
Pertussis vaccine is either offered or mandated in hospitals	2002-2013	5
Pneumococcal vaccine is either offered or mandated in hospitals	2002-2017	13
Pneumococcal vaccine is either offered or mandated in long-term care facilities	1991-2015	26
Varicella vaccine is either offered or mandated in ambulatory care facilities	1995-2017	5
Influenza vaccine is either offered or mandated in hospitals	1995-2019	24
Influenza vaccine is either offered or mandated in long-term care facilities	1995-2020	33
Influenza vaccine is either offered or mandated in ambulatory care facilities	1998-2021	12
Average (28 policies)	1996-2017	23.39

This table lists all 28 policies in the vaccine regulations data set.

Table C.4: Highest and lowest innovators of NBER policies (20%)

	1991-2020		Difference (SE)
	(1) Top 20%	(2) Bottom 20%	(1)-(2)
Rep. two-party vote-share %	44.65	55.98	-11.33
	[8.59]	[6.54]	(2.59)
|Demeaned two-party vote-share|	9.20	6.19	3.01
	[5.05]	[4.41]	(1.49)
Unified Dem. state gvt.	0.31	0.13	0.18
	[0.46]	[0.34]	(0.07)
Unified Rep. state gvt.	0.16	0.42	-0.26
	[0.37]	[0.49]	(0.11)
Legislative professionalism	1.00	-0.84	1.84
	[2.11]	[0.45]	(0.63)
Log(population)	15.42	14.90	0.52
	[1.00]	[0.89]	(0.39)
Income per capita	40569.71	33051.67	7518.04
	[13543.22]	[11052.03]	(1987.49)
Log(income per cap.)	10.55	10.35	0.21
	[0.34]	[0.34]	(0.05)
Urban pop. %	85.05	61.06	23.99
	[7.27]	[7.83]	(2.95)
Minority %	27.91	23.25	4.66
	[12.19]	[11.45]	(4.66)
Unemployed %	5.69	5.28	0.41
	[2.09]	[1.79]	(0.40)
States	12	12	

This table compares characteristics of the states in the highest and lowest 20% for first innovations for the NBER policies. Averages are taken over the entire time period. Standard deviations are in brackets and standard errors in parentheses. Standard errors for the difference are clustered by state. Hawaii, Washington D.C., and Alaska are excluded.

Table C.5a: Examples of states in closest thirds

Decade	Demographics	Distance	Vote-share	State party	Migration
1960s	NM ← NH,	VT ← MI,	AR ← LA,	SC ← NC,	AR ← OH,
	TN ← SC,	VA ← RI,	VT ← CO,	VA ← NC,	UT ← NV,
	WV ← SD,	WI ← PA,	KS ← WY,	MS ← NC,	MD ← CA,
	WI ← CO,	NM ← ID,	CT ← NJ,	GA ← NC,	MD ← NC,
	KS ← MO	OR ← AZ	MD ← NY	GA ← TX	ND ← WA
1970s	GA ← AL,	KY ← PA,	VA ← NH,	WY ← NY,	WV ← VA,
	MO ← PA,	MT ← KS,	ID ← NE,	MD ← GA,	WA ← MN,
	NH ← WY,	AZ ← CA,	IL ← GA,	GA ← MS,	OR ← NY,
	VT ← NE,	NE ← CO,	ID ← VA,	RI ← MS,	TX ← FL,
	IL ← TX	MO ← OK	GA ← DE	OH ← ID	OR ← UT
1980s	PA ← VA,	TX ← CO,	IN ← KS,	PA ← ND,	IL ← AZ,
	NM ← KS,	MA ← NJ,	MT ← NJ,	RI ← NM,	ND ← TX,
	TN ← WI,	MD ← KY,	MA ← MN,	WY ← DE,	CT ← NY,
	NM ← MT,	MS ← OK,	CT ← OH,	CO ← NY,	ID ← NV,
	KY ← IN	IN ← MO	OR ← WI	MA ← MD	WI ← MN
1990s	MI ← MO,	AR ← LA,	NM ← MI,	TX ← OK,	ME ← CA,
	VA ← MD,	RI ← MA,	UT ← VA,	NC ← AL,	NJ ← AZ,
	MD ← CO,	IL ← KY,	MD ← WV,	DE ← NY,	NJ ← OH,
	DE ← NV,	MD ← NJ,	MA ← WV,	MT ← WY,	NY ← NJ,
	KY ← AL	MA ← PA	WA ← MD	KY ← GA	NH ← NY
2000s	ID ← KY,	VA ← DE,	VT ← DE,	ND ← NE,	MD ← GA,
	WI ← PA,	SD ← ND,	RI ← MI,	OR ← MA,	ID ← OR,
	CA ← CO,	ND ← OK,	NM ← PA,	NE ← SD,	IN ← FL,
	MA ← AZ,	MN ← OK,	NV ← NM,	FL ← ID,	MN ← AZ,
	MA ← NY	CO ← NE	DE ← MD	KS ← WY	NE ← IA
2010s	OK ← SC,	MS ← OK,	WY ← MS,	ND ← ID,	MD ← OH,
	LA ← OR,	VT ← NJ,	NV ← MN,	KS ← MI,	VT ← FL,
	DE ← NV,	MN ← KY,	AZ ← FL,	UT ← FL,	KY ← TN,
	NM ← IN,	NY ← NH,	MD ← NJ,	NE ← GA,	NJ ← GA,
	GA ← WI	NH ← ME	OR ← DE	AZ ← ID	OR ← TX

XX ← YY means state YY is most consistently in the third of states closest to state XX averaged over the decade. Ties are randomly broken.

Table C.5b: Examples of states in closest thirds of voter preferences (ANES & GSS)

Decade	All policy areas	Within policy area				
		Civic Rights	Economics	Government	Law and Crime	Public Services
1960s	NJ ← PA, IL ← CA, CA ← IL, NC ← FL, OK ← FL	WA ← CA, UT ← CA, IN ← PA, PA ← NY, NC ← KY	MS ← MA, WI ← MO, TN ← CT, MD ← MA, IN ← TX	AZ ← WA, UT ← CT, TN ← AR, KY ← OH, OK ← NJ	OR ← WA, AR ← FL, MI ← MO, LA ← TX, CO ← WA	WI ← GA, WA ← NY, KY ← PA, WA ← SC, LA ← OH
1970s	LA ← TX, AR ← VA, MI ← TX, AL ← MI, OH ← VA	AL ← NC, VA ← MI, IL ← NJ, MN ← MI, CA ← NY	TN ← IA, IA ← IN, AR ← OH, MN ← WI, KY ← MI	SC ← KY, TX ← VA, MO ← OH, IN ← VA, TN ← CA	TX ← AL, IN ← PA, MD ← PA, WI ← PA, AL ← TX	MA ← NJ, UT ← SC, NJ ← OH, NC ← MI, SC ← IL
1980s	NY ← MA, PA ← IL, CA ← NY, FL ← IL, LA ← TX	MI ← MO, NY ← CA, CO ← IL, OK ← FL, WI ← IL	CA ← PA, OR ← MI, MA ← VA, OR ← CA, IN ← CA	MI ← CA, MA ← CA, MS ← OH, MN ← NY, CO ← PA	NJ ← OH, GA ← PA, MA ← IL, OR ← CA, MI ← OH	KY ← MI, MN ← CA, IL ← CA, NY ← PA, GA ← VA
1990s	CT ← PA, FL ← TX, OR ← NY, MI ← PA, CA ← IL	KY ← NC, TX ← IN, IN ← FL, MO ← OH, KY ← OH	AR ← IN, KY ← MI, FL ← CA, FL ← OH, NY ← CA	IN ← VA, OH ← IL, FL ← PA, VA ← PA, OH ← TN	CO ← MI, AZ ← CA, IL ← NY, PA ← CA, MO ← MI	MI ← PA, MO ← CA, TX ← MI, IL ← PA, MI ← OH
2000s	IA ← FL, MD ← NJ, MA ← CA, NY ← PA, GA ← NC	MO ← NC, WA ← CA, LA ← TX, NC ← TN, NJ ← WA	WI ← MI, NC ← NY, MI ← OH, TN ← TX, GA ← FL	GA ← PA, AL ← PA, MS ← TX, IA ← GA, NY ← CA	PA ← NC, FL ← NC, IN ← FL, MA ← IL, IA ← NY	FL ← CA, NJ ← PA, CT ← MI, MD ← FL, CA ← TX
2010s	AR ← MI, LA ← PA, WI ← TX, NJ ← TX, SC ← MA	MA ← FL, OH ← NJ, AZ ← MI, SC ← IL, PA ← GA	OR ← CA, AZ ← WI, UT ← GA, AZ ← FL, SC ← TN	MA ← CO, CA ← OR, VA ← IA, MN ← CA, SC ← CO	CO ← OH, KY ← OH, OR ← UT, SC ← WI, VA ← IA	AR ← TN, MA ← MI, NY ← MI, CO ← CA, NY ← NJ

XX ← YY means state YY is most consistently in the third of states closest to state XX averaged over the decade. Ties are randomly broken.

Table C.6: Robustness checks

Distance			Republican vote-share			State gvnt. party control		
1950-70s	1980-90s	2000-10s	1950-70s	1980-90s	2000-10s	1950-70s	1980-90s	2000-10s
Dep. var.: Policy adoption (all logit except (2))								
(1) Baseline (Table 4.3) (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.43	0.31	0.30	0.13	0.13	0.38	0.06	0.06	0.51
(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)
(2) Baseline linear probability model (coefficients and SEs ×100) (R^2: 0.18, 0.18, 0.18; $N_{\text{pol.}}$: 269, 425, 379)								
1.24	1.04	1.35	0.34	0.46	1.75	0.22	0.28	2.62
(0.15)	(0.16)	(0.22)	(0.12)	(0.12)	(0.19)	(0.15)	(0.16)	(0.29)
(3) Controlling for policy area composition over decades (R^2: 0.18, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.46	0.30	0.33	0.12	0.12	0.36	0.04	0.06	0.52
(0.06)	(0.05)	(0.05)	(0.05)	(0.03)	(0.04)	(0.07)	(0.04)	(0.05)
(4) Expanded state-level controls (R^2: 0.18, 0.18, 0.20; $N_{\text{pol.}}$: 265, 420, 375)								
0.24	0.22	0.18	0.12	0.14	0.40	0.09	0.09	0.50
(0.06)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)
(5) Parsimonious model (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
0.43	0.31	0.31	0.15	0.13	0.37	0.04	0.05	0.50
(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.05)
(6) Adoption measure: Lagged by one year (R^2: 0.14, 0.15, 0.15; $N_{\text{pol.}}$: 267, 420, 375)								
0.34	0.26	0.22	0.06	0.12	0.33	0.03	0.04	0.41
(0.06)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)
(7) Adoption measure: Rank-inverse weighted average (R^2: 0.16, 0.17, 0.17; $N_{\text{pol.}}$: 267, 420, 375)								
2.59	1.85	1.74	0.25	0.77	1.62	0.89	1.26	3.39
(0.26)	(0.23)	(0.30)	(0.23)	(0.17)	(0.22)	(0.53)	(0.30)	(0.46)
(8) Adoption measure: Proportion of closest third that are adopters (R^2: 0.16, 0.17, 0.18; $N_{\text{pol.}}$: 267, 420, 375)								
2.40	1.76	1.59	0.79	0.91	2.07	0.42	0.60	1.49
(0.29)	(0.28)	(0.27)	(0.33)	(0.23)	(0.21)	(0.25)	(0.15)	(0.21)
(9) Adoption measure: Proportion of all adopters in the closest third (R^2: 0.14, 0.15, 0.15; $N_{\text{pol.}}$: 243, 400, 375)								
0.77	0.72	0.86	0.12	0.20	0.94	0.09	-0.37	1.51
(0.09)	(0.09)	(0.10)	(0.07)	(0.08)	(0.08)	(0.12)	(0.13)	(0.13)
(10) Adoption measure: P(Adopt) closest third−P(Adopt) all states (R^2: 0.15, 0.15, 0.16; $N_{\text{pol.}}$: 267, 420, 375)								
2.31	1.71	1.56	0.76	0.95	2.09	-0.33	0.02	1.92
(0.32)	(0.27)	(0.28)	(0.36)	(0.24)	(0.21)	(0.32)	(0.20)	(0.28)

This table presents results from alternate specifications of the policy diffusion model. The table shows coefficients on the measure of adopters among the "closest" states (i.e., the closest third unless otherwise noted) in terms of distance, the average Republication vote-share in the two most recent presidential election, and state government party control. Standard errors clustered by state are in parentheses. Each model is estimated over three separate time periods (1950-70s, 1980-90s, and 2000-10s). The (pseudo-)R^2 and number of policies are reported in parentheses in chronological order corresponding to the three time periods.

Baseline: replicates the specification from Table 4.3 over the longer time periods.

Baseline linear probability model: uses the same covariates in the Baseline specification but estimates the coefficients using a linear probability model.

Controlling for policy area composition over decades: reweights policies in each decade to match the composition of policy areas in the 1980s.

Expanded state-level controls: takes the specification from Table 4.3 and adds: non-white % and unemployed %; quadratic terms for the proportion of all other states adopted, Republican vote-share, log population, income per capita, urban %, non-white %, and unemployed %; adoption measures among the closest third of states in migration flows, non-white %, and unemployed %; a more flexible policy-specific baseline hazard parametrized as a step function that can vary every five years; and state fixed-effects.

Parsimonious model: includes only policy fixed effects and the proportion of adopters among all other states, and the adoption measure among the closest third of other states in the demographic index (not shown), geography, Republican vote-share in the most recent presidential election, and state government party control. (This specification is also used in Table 4.4.)

The following specifications use alternate measures of concentrated adoptions among the similar states, in place of the baseline two-sided likelihood measure. Each specification is "parsimonious" in that the only controls included are policy fixed effects and, except for specifications (8) and (9), the proportion of adopters among all other states.

Lagged by one year: uses the Parsimonious model but takes the adoption measure among the closest other states up to the *prior* (not current) year.

Rank-inverse weighted average: instead of defining the closest states as the third with smallest absolute difference, this measure weights the other states' adoptions by the inverse of their rank in absolute distance, where the closest state is is ranked 1 and the furthest state is ranked 47.

Proportion of closest third that are adopters: uses the proportion of states in the closest third that have adopted.

Proportion of all adopters in the closest third: uses the proportion of all adopters that are in the closest third of states.

P(Adopt) closest third−P(Adopt) all states: uses the proportion of states in the closest third that have adopted minus the proportion of all states (excluding one's own) that have adopted.

Table C.7a: Models of policy diffusion: Role of migration and voter preferences (expanded demographics and public opinion)

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: Policy adoption (logit)	60-70s	80-90s	00-10s	60-70s	80-90s	00-10s
Proportion of states adopted	0.83	2.27	2.37	1.33	2.34	2.49
	(0.20)	(0.14)	(0.23)	(0.28)	(0.13)	(0.20)
Divided state government	-0.00	0.02	-0.18	0.03	0.02	-0.16
	(0.10)	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)
State gvnt./legis. election year	0.16	0.02	-0.01			
	(0.12)	(0.07)	(0.06)			
Presidential election year	-0.25	-0.00	-0.20			
	(0.13)	(0.07)	(0.09)			
Measure of adoption among other states closest in:						
Demographic index				0.29	0.06	0.08
				(0.08)	(0.05)	(0.06)
Log(population)	0.05	0.04	-0.07			
	(0.07)	(0.05)	(0.04)			
Log(income per capita)	0.22	0.01	0.09			
	(0.07)	(0.05)	(0.05)			
Urban population %	0.12	0.03	0.17			
	(0.08)	(0.05)	(0.04)			
Non-white %		0.03	0.00			
		(0.03)	(0.06)			
Unemployed %		0.02	0.02			
		(0.03)	(0.05)			
Distance	0.22	0.19	0.14	0.16	0.20	0.12
	(0.10)	(0.08)	(0.07)	(0.10)	(0.08)	(0.07)
Republican vote-share	0.14	0.09	0.27	0.11	0.08	0.16
	(0.06)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)
State gvnt. partisanship	0.07	0.10	0.40	0.06	0.08	0.45
	(0.12)	(0.07)	(0.08)	(0.11)	(0.07)	(0.07)
State gvnt. partisanship×Divided gvnt.	-0.16	-0.03	-0.48	-0.16	0.01	-0.58
	(0.22)	(0.15)	(0.15)	(0.19)	(0.14)	(0.13)
Migration flows	0.14	0.01	0.12	0.14	0.02	0.15
	(0.14)	(0.10)	(0.09)	(0.14)	(0.09)	(0.08)
Voter preferences (ANES & GSS)	0.25	0.29	0.14	0.18	0.25	0.15
	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)	(0.07)
Index of public opinion measures	0.19	0.17	0.19			
	(0.06)	(0.04)	(0.05)			
Citizen ideology (Berry et al., 1998)				0.36	0.05	0.17
				(0.06)	(0.03)	(0.05)
Public policy mood (Lagodny et al., 2022)				-0.07	-0.04	-0.04
				(0.06)	(0.04)	(0.04)
Mass social liberalism (Caughey and Warshaw, 2018)				0.07	0.19	0.22
				(0.06)	(0.05)	(0.05)
Mass economic liberalism (Caughey and Warshaw, 2018)				0.06	0.05	0.11
				(0.08)	(0.04)	(0.05)
Observations	53757	107999	66721	60453	116510	70341
Policies	227	414	359	227	414	350
Pseudo R^2	0.17	0.17	0.19	0.16	0.17	0.18

This table reports the role of each individual factor included in the demographic and public opinion indices from Table 4.5 and shows the results for additional predictors, such as election years and similarity in non-white % and unemployment rates (available from the 1970s). The three factors included in the demographic index are population, income per capita, and urban %. See the notes for Table 4.5 for a description of each factor included in the public opinion index. Standard errors clustered by states are in parentheses.

Table C.7b: Models of policy diffusion: Role of migration and voter preferences (expanded ANES & GSS voter preferences)

Dep. var.: Policy adoption (logit)	(1) 60-70s	(2) 80-90s	(3) 00-10s	(4) 60-70s	(5) 80-90s	(6) 00-10s
Proportion of states adopted	1.19	2.34	2.42	1.19	2.35	2.42
	(0.29)	(0.13)	(0.21)	(0.29)	(0.13)	(0.21)
Divided state government	0.01	0.02	-0.18	0.02	0.02	-0.18
	(0.11)	(0.05)	(0.07)	(0.11)	(0.05)	(0.07)
Measure of adoption among other states closest in:						
Demographic index	0.34	0.09	0.13	0.33	0.09	0.13
	(0.08)	(0.05)	(0.06)	(0.08)	(0.05)	(0.06)
Distance	0.23	0.19	0.11	0.23	0.18	0.11
	(0.10)	(0.08)	(0.06)	(0.10)	(0.08)	(0.06)
Republican vote-share	0.15	0.09	0.27	0.13	0.09	0.26
	(0.06)	(0.04)	(0.06)	(0.06)	(0.04)	(0.06)
State gvnt. partisanship	0.06	0.10	0.48	0.07	0.09	0.47
	(0.11)	(0.06)	(0.07)	(0.11)	(0.07)	(0.07)
State gvnt. partisanship×Divided gvnt.	-0.09	-0.00	-0.61	-0.11	0.00	-0.61
	(0.19)	(0.13)	(0.13)	(0.19)	(0.13)	(0.13)
Migration flows	0.14	0.02	0.15	0.11	0.03	0.14
	(0.14)	(0.09)	(0.08)	(0.13)	(0.09)	(0.08)
Voter preferences (ANES & GSS) in policy area	0.20	0.12	0.07			
	(0.09)	(0.05)	(0.05)			
Voter preferences in other policy areas	0.11	0.16	0.08			
	(0.09)	(0.08)	(0.06)			
Voter preferences and sentiment (ANES & GSS)				0.34	0.25	0.16
				(0.10)	(0.09)	(0.08)
Index of public opinion measures	0.22	0.14	0.21	0.20	0.14	0.21
	(0.07)	(0.04)	(0.05)	(0.07)	(0.04)	(0.05)
Observations	57932	108973	67435	57932	108973	67435
Policies	216	389	338	216	389	338
Pseudo R^2	0.16	0.17	0.18	0.16	0.17	0.18

This table reports two extensions of the ANES & GSS measure of voter preferences from Table 4.5. In Columns 1-3, the ANES & GSS survey questions are categorized into the six policy areas shown in Table 4.1c. The measure of voter preferences is then calculated separately for questions in the relevant policy area and for all other questions related to the other policy areas. Policies in the "Environment and Energy" policy area are dropped due to insufficient representation of voter preferences in ANES & GSS survey questions for earlier time periods. In Columns 4-6, the set of ANES & GSS questions used to measure voter preferences is expanded to include questions about not only preferences regarding specific policies but also attitude toward policy topics more broadly. See Online Appendix Section B for details. Standard errors clustered by states are in parentheses.

Table C.8: Event studies

	Uni. st. gvnt.	Unified Republican state government				Unified Democratic state government				Loss of uni.
	(1) Diff.	(2) Right-lean. policy	(3) Left-lean. policy	(4) Diff. (2-3)	(5) Neutral policy	(6) Left-lean. policy	(7) Right-lean. policy	(8) Diff. (6-7)	(9) Neutral policy	(10) Diff.
Events during years 1950 to 1989										
4 years pre-event	-0.024 (0.009)	-0.023 (0.012)	0.024 (0.013)	-0.047 (0.023)	0.019 (0.017)	-0.006 (0.007)	0.019 (0.011)	-0.025 (0.014)	-0.008 (0.008)	-0.001 (0.008)
3 years pre-event	-0.014 (0.007)	-0.012 (0.008)	0.014 (0.011)	-0.026 (0.014)	0.026 (0.014)	-0.012 (0.005)	0.001 (0.006)	-0.013 (0.008)	-0.008 (0.007)	-0.003 (0.008)
2 years pre-event	-0.013 (0.011)	-0.011 (0.019)	0.029 (0.015)	-0.039 (0.032)	0.026 (0.019)	-0.005 (0.007)	0.005 (0.006)	-0.010 (0.010)	0.004 (0.010)	0.002 (0.008)
1 year pre-event	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)
Year of event	0.009 (0.010)	-0.001 (0.014)	-0.006 (0.010)	0.005 (0.018)	0.028 (0.020)	0.005 (0.006)	-0.006 (0.008)	0.011 (0.012)	0.007 (0.009)	0.000 (0.011)
1 year post-event	0.001 (0.008)	0.009 (0.010)	0.006 (0.009)	0.003 (0.017)	0.005 (0.011)	-0.002 (0.006)	-0.006 (0.005)	0.003 (0.008)	-0.012 (0.009)	0.007 (0.007)
2 years post-event	-0.006 (0.012)	0.002 (0.015)	0.011 (0.016)	-0.009 (0.025)	-0.014 (0.019)	0.003 (0.008)	0.010 (0.010)	-0.006 (0.014)	0.003 (0.011)	-0.003 (0.010)
3 years post-event	-0.003 (0.009)	0.004 (0.013)	-0.002 (0.013)	0.006 (0.022)	-0.002 (0.013)	-0.008 (0.007)	-0.004 (0.006)	-0.004 (0.008)	0.008 (0.009)	-0.009 (0.008)
4 years post-event	0.003 (0.010)	-0.012 (0.011)	0.005 (0.010)	-0.018 (0.014)	0.029 (0.025)	0.005 (0.009)	-0.006 (0.011)	0.011 (0.014)	0.049 (0.017)	-0.008 (0.010)
Observations	72118	57439	57439	57439	57439	72049	72049	72049	72049	73639
Policies	233	191	191	191	191	233	233	233	233	241
Events	135	52	52	52	52	82	82	82	82	148
Events during years 1990 to 2020										
4 years pre-event	0.000 (0.007)	-0.002 (0.009)	-0.002 (0.008)	-0.001 (0.014)	0.005 (0.010)	0.003 (0.005)	0.002 (0.005)	0.002 (0.007)	0.005 (0.006)	0.019 (0.008)
3 years pre-event	-0.013 (0.007)	-0.006 (0.008)	0.004 (0.008)	-0.010 (0.014)	0.017 (0.007)	-0.005 (0.007)	0.009 (0.005)	-0.014 (0.007)	-0.001 (0.005)	0.013 (0.007)
2 years pre-event	-0.007 (0.008)	-0.006 (0.008)	0.019 (0.008)	-0.024 (0.012)	0.026 (0.008)	0.003 (0.006)	-0.007 (0.006)	0.010 (0.010)	-0.009 (0.007)	0.017 (0.008)
1 year pre-event	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)	– (–)
Year of event	0.022 (0.007)	0.012 (0.008)	0.003 (0.006)	0.008 (0.010)	0.015 (0.011)	0.028 (0.008)	-0.005 (0.006)	0.033 (0.010)	0.001 (0.008)	0.001 (0.010)
1 year post-event	0.022 (0.009)	0.011 (0.009)	-0.003 (0.006)	0.014 (0.011)	0.012 (0.009)	0.020 (0.009)	-0.011 (0.006)	0.031 (0.012)	-0.006 (0.005)	0.016 (0.009)
2 years post-event	0.026 (0.009)	0.008 (0.010)	-0.006 (0.008)	0.014 (0.014)	-0.001 (0.015)	0.026 (0.010)	-0.009 (0.007)	0.035 (0.013)	-0.007 (0.007)	0.009 (0.010)
3 years post-event	0.018 (0.007)	0.007 (0.010)	-0.010 (0.007)	0.017 (0.013)	0.013 (0.014)	0.014 (0.007)	-0.007 (0.006)	0.021 (0.010)	0.002 (0.009)	0.005 (0.008)
4 years post-event	0.028 (0.015)	-0.014 (0.008)	-0.010 (0.010)	-0.005 (0.014)	0.005 (0.011)	0.041 (0.022)	-0.023 (0.009)	0.064 (0.027)	0.013 (0.009)	-0.002 (0.009)
Observations	124482	112437	112437	112437	112437	123679	123679	123679	123679	122133
Policies	427	417	417	417	417	426	426	426	426	426
Events	115	49	49	49	49	64	64	64	64	99

Appendix D

Appendix for Bottlenecks for evidence adoption

D.1 Example of BIT trial report

THE
BEHAVIORAL
INSIGHTS TEAM

Reducing errors in business license renewal applications

A Trial Report from the Behavioral Insights Team

What was the context and goal?

The City of [redacted] worked with the Behavioral Insights Team to see if we could reduce the error rate for business license renewal applications. There are approximately 7,200 businesses and individuals that hold one or more [redacted] business licenses. Licenses are renewed annually.

The requirements for renewing a business license are complex, and the existing renewal notice does not effectively help applicants navigate that complexity. Approximately half of license renewal applications have one or more errors. When an error is spotted, city staff call the applicant to resolve the issue (which often include re-submission of paperwork) or even have to mail back the application, which re-starts the process. By reducing the error rate, the city will save time and resources, as will business licensees.

What did we test?

We designed a new license renewal notice aimed at better supporting applicants in navigating the renewal process. There were three primary changes:

1. We developed and included a comprehensive guide to what supporting documentation each licensee needed to provide. Based on their specific licenses up for renewal, the end of each notice include a picture of each required document, guidance on how to avoid common errors, and contact information for the relevant authority should the applicant have questions.
2. We re-organized the information in the notice so that the requirements for all licenses were grouped together. For example, we listed the required documents for all licenses to be renewed in one consolidated list. This makes the task appear simpler and reduces potential duplication of work effort. Similarly, we consolidated all payment requirements into one step.
3. We prompted applicants to set aside a specific date and time for completion of the renewal application.

We tested the re-designed notice to determine whether it reduces the error rate in license renewal applications.

Why did we think it might work?

Our new renewal notice was designed to respond to several key issues that we thought might play a significant role in the overall error rate. These issues were identified by program staff who manage the renewal process and were

supplemented by our prior experience with similar complex administrative tasks that governments require businesses to undertake.

1. Frontline staff indicated that the greatest source of error were issues with supporting documentation (e.g. certificates, affidavits or other licenses that licensees were required to submit to get their business license renewed). We believed that by providing licensees with a guide at the end of the notice, customized to their requirements, it would be easier for them to understand what supporting documentation they needed to provide.
2. To get their licenses renewed, businesses need to complete their renewal applications, provide the required documentation and make a payment of the correct amount. This is a complex administrative task that business owners or responsible employees may seek to avoid or put off. We thought that prompting licensees to set a specific date and time for completing the task would help avoid procrastination. Similar approaches to help people set "implementation intentions" have been effective in other contexts.
3. We thought that the structure of the existing notice, which listed the requirements for each notice separately, made the task seem more complex and burdensome than it was in reality. For example, if three licenses required a copy of the business owners drivers' license, the owner might think they need to provide three photocopies of that license, but one would be enough. By consolidating the requirements of all licenses to be renewed, we sought to reduce the perceived complexity of the task and reduce procrastination.

How did we test it?

We designed a two-armed randomized controlled trial (RCT), sending either the original or re-designed renewal notice to all business license holders who were scheduled to renew business licenses between January 2018 and December 2019. Randomization was clustered on the first letter of the business owner's name.

By randomly assigning some businesses to receive the new notice and others to receive the old notice, we could be confident that differences in error rate between these two groups would be the result of the new notice itself, rather than any other factors. For this reason, RCTs are often considered the "gold-standard" in evaluating the impact of new approaches.

What did we find?

The re-designed notices **reduced the error rate in license renewal applications from about 48% to about 44%, a 7% relative decrease.** The re-designed notices also required staff to mail back about 19% fewer applications (it was 9.3% for the old notice and 7.5% for the new notice). This suggests that in addition to reducing the

proportion of renewal applications that had any error, the new notice reduced the severity of the errors.

In further exploratory analysis, we found that the re-designed notice reduced documentation errors and renewal information errors, but had no effect on payment errors.

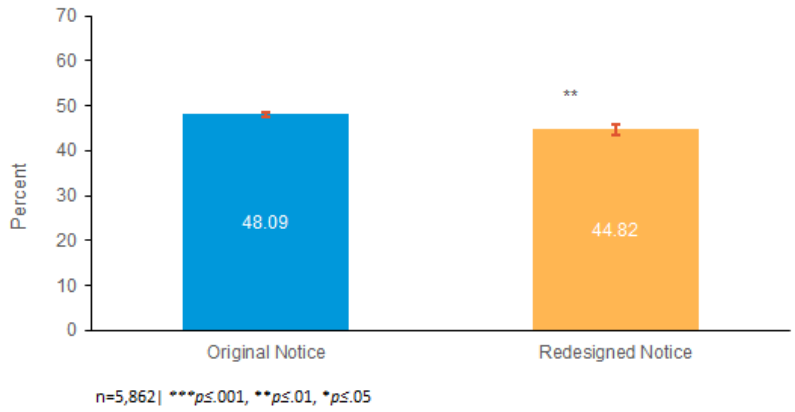
Recommendations

- We recommend that the city switch all licensees to the new format as soon as is convenient to reduce error and mail back rates.
- Even with the new notice, the error rate (44%) and mail back rate (7.5%) are still quite high. We believe that further changes to the renewal notice are unlikely to substantially bring this rate down. As a result, we recommend that the city consider changes to:
    - Policy, with a focus on simplifying renewal requirements;
    - Process, potentially reducing the frequency of renewals (which are currently annual); and
    - Systems, as moving to online renewals could potentially allow the city to pre-populate renewal applications, automatically validate payments or documentation, or take other actions to limit the likelihood and impact of errors.

Results

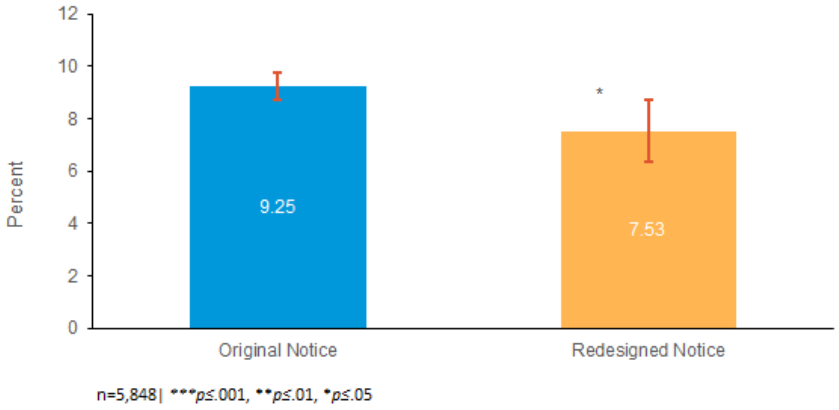
The re-designed notice decreased the likelihood that a renewal application would have a documentation, renewal information, or payment error by 3.27 percentage points (standard error = 1.19) for a relative decrease of 7 percent. This difference is statistically significant at a p-value of less than 0.01, meaning it is very unlikely to have occurred by chance.

Figure 1: Error rates in license renewal applications



The re-designed notice also decreased the likelihood that a renewal application would be mailed back to the applicant by 1.72 percentage points (standard error = 0.78) for a relative decrease of 19 percent. This difference is statistically significant at a p-value of less than 0.05, meaning it is very unlikely to have occurred by chance.

Figure 2: Mailback rates in license renewal applications



D.2 Online forms and communications on city websites

From the BIT-trial sample alone, we do not know whether the city departments that self-select into experimenting are more prone to communicate with their citizens and update their communications in general. Furthermore, though we find that city departments are more likely to adopt evidence on pre-existing communications, we cannot preclude that the BIT city departments that were already sending pre-existing communications have some unobserved factor, such as institutional priority or a diligent employee, that makes them more communicative as well as more willing to adopt evidence. Nor can we know how often pre-existing communications are updated in the absence of evidence generated from a trial.

We partly address these issues of self-selection and endogeneity by collecting communications outside the BIT-trial sample from both BIT and comparable non-BIT city departments. For each city in the BIT sample, we find the 10 closest cities in terms of population size in 2020 and in the same geographic census region that are not in the BIT sample.

For each BIT city that ran a trial and the non-BIT cities matched to it, a team of research assistants used the Wayback Machine and to find public forms or communications on those cities' websites that are related to the BIT trial departmentally or topically. We look for forms that were posted and downloadable at least once during 2015-19, when the BIT trials were conducted in our sample.

Specifically, the communication must either be (a) in the same *department* that ran the BIT trial (e.g., Utilities, Business Licensing, or Police), or (b) for a purpose related to the *trial communication*. This can include communications that are posted by a different department than the one that ran the trial. For example, suppose the Environmental Services Department conducted the BIT trial, which sent a postcard encouraging residents to use a large-item collection service, but the Environmental Services Department website has no public forms or communications. Suppose the Collection Services Department website, however, has a form for household appliance pickup, which is related to the trial communication even though it is not on the Environmental Services Department website. As departments are not standardized across cities and departmental organization sometimes changes even within cities, this criterion provides enough flexibility to find forms and communications matched at the BIT-trial level.

When a city ran multiple BIT trials in different departments, we find communications in the corresponding departments of the non-BIT cities matched to that BIT city. For example, suppose City A ran two BIT trials, Trial 1 in the Code Enforcement Department and Trial 2 in the Finance Department. If City B is a non-BIT city matched to City A, we try to find forms or communications from City B's Code Enforcement Department as well as their Finance Department.

For every city department that ran a BIT trial, we attempt to find an online form or communication. For the 10 non-BIT city departments matched to each BIT trial, we stop once we find a form or communication for 2 of them. For instance, if we find 2 communications from the first 7 matched non-BIT city departments, we do not attempt to find any from the remaining 3. There are cases when we tried all 10 of the matched non-BIT city departments without finding 2 with an adequate form or communication.

For the rate of finding forms, Figure A.2a shows that 26 out of 52 (50%) of the BIT city departments that ran a trial with a new communication have an online form or com-

munication, as do 13 out of 21 (62%) of the BIT city departments that ran a trial with a pre-existing communication. For the non-BIT city departments, 99 out of the 280 attempts for departments matched to a BIT trial with a new communication and 44 out of the 93 attempts for departments matched to a BIT trial with a pre-existing communication were successful. However, these raw numbers are likely downward biased as there would be more attempts for the city departments for which it is difficult to find forms. We thus report the weighted rate for the matched non-BIT cities in Figure A.2a, which inversely weights by the number of attempts for non-BIT city departments matched to the same BIT trial. For example, if there were 8 attempts for non-BIT city departments matched to the same BIT trial, each of the attempts is given a weight of 1/8.

The BIT and non-BIT city departments are similar in two ways in Figure A.2a. First, they have a similar rate of downloadable forms or communications on their websites. On this measure, the BIT cities seem no different from matched non-BIT cities in how communicative they are with their citizens. Second, they both have a higher rate in departments matched to BIT trials that tested a pre-existing communication. That is, the types of departments that sent pre-existing communications in the BIT trials seem to be more communicative in general, and not just in the BIT cities.

Conditional on having an online form or communication, how much do cities update from year to year? Based on cosine similarity, a measure explained in Appendix D.3, Figure A.2b reports the average similarity in the forms across years. For example, if we downloaded versions of the form in 2015, 2018, and 2022, we compute the cosine similarity between the 2015 version and the 2018 version and between the 2018 version and the 2022 version.

The forms change very little on average and not at all in most years as shown in Figure A.2b, regardless of whether the city department is in the BIT sample or whether the (matched) BIT city department ran a trial with a new or a pre-existing communication. The overall average cosine similarity is 0.96 (SE=0.005), where 1 represents perfect similarity. For reference, the average cosine similarity between the control and treatment communications in the BIT trial sample is 0.88 (SE=0.01), and the average between the control and the adopted communication (in cases of adoption) is 0.82 (SE=0.02).

From the out-of-sample forms and communications, we discover that city departments with pre-existing communications in the BIT trials, such as Utilities, Treasury, and Parking, are more communicative not just in BIT cities but in general. One possibility is that communicating with citizens is more a core function of these departments; after all, city governments need to send citizens their utility bills and parking fine citations. It still could be that unobserved factors in departments with pre-existing communications, such as institutional priority or an assiduous city employee, make them both more communicative and responsive to evidence. If so, we might also expect that the same factor would make those departments update their other communications more regularly, but that is not the case with the forms. We also learn that the changes to the communications from the BIT trials seem to be far beyond what occurs in other pre-existing communications, at least the ones posted online, in the absence of evidence.

D.3 Cosine similarity in BIT trial communications and out-of-sample online forms

We use the average of the textual and graphical cosine similarity to measure how alike two versions of a communication are in the BIT trials and in the online forms described in Appendix D.2. In general, the cosine similarity between two vectors a and b of length n is the dot product divided by the Frobenius norms of the two vectors, or $\sum_i^n a_i b_i / \sqrt{\sum_i^n a_i^2 \sum_i^n b_i^2}$. Values range from -1 to 1, where 1 indicates complete similarity, -1 complete dissimilarity, and 0 orthogonality. In our case, calculating cosine similarity between two images of a communication or form then amounts to representing the text and graphics in the images as vectors.

For a vector representation of the text in the images, we perform optical character recognition using Google's Tesseract-OCR Engine on the two versions of the communication. This gives us a vector of words scraped from the images. We keep only the set of words that appear at least twice across the two versions (i.e., at least once in each or at least twice in one). This normalization reduces noise by eliminating one-off words, such as placeholder names, addresses, dates, or payment amounts used in the templates. In each of the two versions, we then count the number of times each word appears, which is our vector representation of the text. For example, suppose the first version has the text "Remember to submit your application! Continue application here – Officer Lee," and the second version has the text "Remember to submit here – Officer Wayne." After dropping the words that appear only once, the dictionary for these two versions is (remember, to, submit, application, here, officer), and the vector representations are (1,1,1,2,1,1) and (1,1,1,0,1,1) respectively, which result in a cosine similarity of $5/\sqrt{9 * 5} \approx 0.75$. As demonstrated in the example, this measure computes the similarity in the frequency but not the order of the words.

For a vector representation of the graphics in the images, we use the ResNet-50 model, a convolutional neural network that is commonly used for evaluating image similarity. We read the image pixels into an array with three dimensions for the red, blue, and green channels and pre-process the array, which involves scaling pixel values and normalizing each channel. The ResNet-50 model runs the image through the convolutional layers and predicts the image embedding, a vector that represents the graphical features of the image. We use this vector in the calculation of the cosine similarity.

We evaluate both our human coding of (non-)adoption and the machine measure of cosine similarity using the 11 BIT trials for which we have the templates for the control, treatment, and current communications. This requires that the control group in the trial received some communication and that the city is still sending some version of the communication. We coded 4 of the 11 trials as cases of non-adoption. In line with our judgment, the cosine similarity in both text and graphics is higher between the control and the current versions than between the treatment and the current versions in all 4 cases. In the 7 cases that we coded as adoption, the cosine similarity is higher between the treatment and the current versions than between the control and the current versions in either text or graphics for 3 of them and in both for the remaining 4.

D.4 Marginal cases of non-adoption

As mentioned in the text, we defined a city as adopting a trial if the city has used one of the communications in the nudge treatment arms again following the RCT. This includes cases when the city had used the communication after the trial but was not currently doing so, for example, because it was not an election year. When cities have made further changes to the communication since the trial, we counted adoption as incorporating at least 50% of the nudge features as pre-specified in the internal trial protocol or report. For example, suppose a trial tested a utility bill by (i) simplifying the payment request, (ii) adding a peer comparison, and (iii) personalizing the message. If the current utility bill incorporates the simplification and the peer comparison but not the personalization, we count it as adoption, but if it only includes personalization, we do not.

Most cases of (non-)adoption were clear according to this rule, but there were 4 cases of non-adoption for which the post-trial communication seemed to include some nudge features, but did not meet our criteria upon close inspection. We describe each of these marginal non-adoption cases below.

1. This city sent new postcards encouraging local business owners to renew their license online. The control arm used a slogan on convenience, whereas the treatment arm used one with normative language. Both postcards were equally effective. The city no longer sends the postcards, but uses the same exact slogan from the control arm in different letters sent to businesses about their licenses.

2. This city police department sent recruitment postcards to local neighborhoods. The version with a message emphasizing the benefits and salary had the strongest effect. Now on its website, the police department has adopted this type of messaging for recruitment.

3. This city police department used online ads to recruit applicants from Historically Black Colleges and Universities (HBCUs). The control ads, based on a prior pilot, highlighted the relatable background of current police officers, and the treatment ads also offered a "personal concierge" service to guide applicants through the process. The control ads were significantly more effective. The police department still uses online ads to target applicants from HBCUs, but does not use the treatment messaging.

4. In both the control and treatment arms of this trial, the city added a new checkbox to the water utility bill for easy enrollment into a local charity program. The utility bill in the treatment arm also included colorful ASCII art and a message requesting recipients to sign up for the program. There was not a significant difference between the control and treatment arms. The city continues to send the utility bill with only the checkbox from the control arm.

Furthermore, in 11 out of the 20 cases of adoption, the city contacts verbally provided a description of the communication they had used after the trial that matched the treatment arm, but they could not send us a template of the exact communication for us to independently verify due to bureaucratic or technical issues (e.g., they were no longer using the same email system from which the newsletter had been sent before).

As a robustness check in Table A.6, we drop the marginal non-adoption cases and/or the verbal-only adoption cases. We replicate the main specification (Column 4) from Table 3. The key finding, namely on the importance of pre-existence, remains large and significant.

D.5 Broad cases of adoption

As introduced in Section 2.3, we also code cases of broad adoption in which the city contacts stated that either the trial with BIT-NA inspired another communication in a separate context or the city adopted the process of experimentation for their own internal trials. A brief description of each case is listed below with the number of related trials counting as broad adoption shown in parentheses.

- In this trial, a police department sent postcards encouraging applications from minority groups. From the trial, the city identified successful language that they added to subsequent phone and email recruitment scripts. (1 trial)
- A city police department used implementation intentions in an email trial targeting inactive applicants. The department did not continue the prompts for implementation intentions, but incorporated emailing inactive applicants in their long-term recruitment process. (1 trial)
- In this trial, the city sent text reminders for show-cause hearings. The department no longer sends these text reminders, but now sends similarly worded texts for citations, a step prior to the show-cause hearings. (1 trial)
- This city ran an email trial to recruit police applicants. After the trial, the city conducted three internal RCTs in other contexts. (1 trial)
- A city used a nascent text messaging system in two trials to remind citizens under a Medicaid waiver program to use their free health check-up. Motivated by these trials, the city began to use text reminders for a variety of purposes. (2 trials)
- This city ran three trials with BIT-NA for charitable giving, police recruitment, and paperless utility billing. These collaborations inspired the city to create its own internal team to experiment with nudge interventions in city communications. (3 trials)
- In this trial, the city sent postcards to encourage applications to the police force. The Tax and License Division adopted nudges and the process of experimentation in their communications. (1 trial)
- This city conducted two trials with BIT-NA for donations to local charities and voter registration. After the trials, the city established an internal nudge team that has run at least four trials, for example, on library fine payment and water conservation. (2 trials)

- This city police department sent postcards for police recruitment in a trial. The postcards were discontinued, but the findings informed other recruiting materials such as bus advertisements and language on the website. (1 trial)
- This city police department ran three trials with BIT-NA to improve recruiting practices by implementing social media advertisements as well as email and text reminders. These trials led to expanded communication efforts for police recruitment through these mediums. (3 trials)

D.6 Forecasting survey

This section details the 10-minute forecasting administered through the Social Sciences Prediction Platform[73]. In total, 118 forecasters submitted their predictions on the platform over 25 days. The survey first summarized the setting and main result of DellaVigna and Linos (2022), and then introduced the focus for the current paper on the adoption rate of the nudge interventions *after* the RCT collaborations with the cities and on the determinants of adoption. The survey described the sample of trials and highlighted that each trial was co-designed by BIT-NA and the partnering city and that the results were shared with the city in a report after the trial. Next, the survey showed two randomly selected examples of communications used in trials with a brief description of the policy area and targeted outcome.

The forecasters then made their first prediction on the baseline adoption rate. Specifically, we asked, "*What percent of the 73 trials do you think have been adopted by the cities?*" The forecasters provided their answer in percentages (from 0 to 100). We defined adoption as: "*We count a city as "adopting" a trial if one of the nudge treatment arms has been used in city communications after the trial with BIT.*" We gave an example of an adopted trial, showing the nudge communication used in the trial next to the comparable current communication in use by the city. For reference, we provided two statistics: 78% of the trials had at least one nudge intervention arm that led to an improvement relative to the control group, and 45% of the trials found a nudge that led to a significant improvement with $p < 0.05$. On the same page, we asked forecasters to write a short list or a couple sentences in an open-ended text box on which determinants of adoption they expect to matter most.

We then introduced the determinants of adoption that we consider: statistical significance, effect size, retention of the original city staff collaborator, state capacity (proxied by city population), What Works Cities certification, pre-existence of the communication in the trial, and behavioral mechanism used in the nudge intervention. (At this point, forecasters could not return to the previous page to change their baseline prediction nor their open-ended responses.) Next, the survey asked for the predicted adoption rate for each of these determinants separately page-by-page. The survey randomized the order of the determinants between two different orderings.

For each determinant, the sample of trials was separated into relevant bins with the number of trials in each bin shown, and forecasters predicted the adopted rate within each

[73] `https://socialscienceprediction.org/`

bin. For example, for statistical significance, we asked what percent the forecasters think have been adopted for trials that found: (i) a statistically significant improvement (i.e., $t \geq 1.96$, covers 45% of all trials, $n = 33$), (ii) a statistically insignificant improvement (i.e., $0 < t < 1.96$, covers 33% of all trials, $n = 24$), and (iii) a zero or negative effect (covers 22% of all trials, $n = 16$).

On every page, we reminded forecasters of their predicted baseline adoption rate from the very first question. For comparison, we displayed the weighted average adoption rate implied by their forecasts for the determinant on the page as a soft "nudge" to help them give forecasts that were consistent with their initial predicted baseline rate. For example, if they predicted that the adoption rates were 50% for statistically significant trials, 30% for statistically *in*significant trials, and 10% for zero or negative trials, then the weighted average we calculated for them would be $(50\% \times 0.45) + (30\% \times 0.33) + (10\% \times 0.22) = 34.6\%$.

Lastly, we asked forecasters to compare our sample of RCTs in U.S. municipal cities with similar representative samples of trials conducted by large multinational firms and by governments of low-income countries. We asked forecasters to rank these three samples by the overall adoption rate and by the responsiveness to evidence in adoption.

D.7 Categorizing trials combining new and pre-existing communications

In 6 trials, a new insert or letter was sent in addition to a pre-existing mailer. For example, a new postcard insert was added in the same envelope for a pre-existing routine utility bill. In these cases, we focus on the adoption of the new insert or letter and count them as new communications, since the pre-existing component (i.e., the utility bill in the example) remained unchanged and the nudge was entirely contained in the new inserts. In 2 trials, pre-existing email and letter notices were modified to include behavioral mechanisms and new text reminders were also used. We count these two cases as pre-existing communications, as the nudge intervention involved changes to the pre-existing email and letter. In one of these trials, the city adopted both the modified nudge email and the new text reminder, and in the other trial, the city adopted neither the nudge version of the letter nor the text reminder.

D.8 Appendix figures and tables

Figure D.1: Example of adoption of BIT-NA trial

(a) Status-quo control arm communication

[redacted]

MUNICIPALITY OF [redacted]

June 5, 2017

This is A Test
123 Test Street
[redacted]

RE: Case # 3AN-77-77777MO Balance Due $ 96.67

Our records indicate you still owe on the court case listed above. This is a courtesy reminder for payment, which must be made within **10 days** from the date of this letter to avoid further collection action. Your case may have been referred to our 3rd party collection agency and additional collection fees assessed.

Your outstanding balance due is reflected on our public website at **www.**[redacted](Link: Your Government > Delinquent Criminal & Civil Fines > Search the DCF Database). Your payment options are listed below:

- Pay online using a credit card, debit card or electronic check through Municipal Services Bureau at www.[redacted].
- Call [redacted] and pay with a credit card, debit card or electronic check through Municipal Services Bureau.
- Mail a check or money order to: Municipal Services Bureau, PO Box [redacted]

Note: a convenience fee is assessed by Municipal Services Bureau for electronic payment services.

IMPORTANT: To ensure the accurate application of your payment, please reference the case number(s).

Failure to resolve this matter within 10 days from the date of this letter may result in the exercising of our rights under the Court's judgment, including one or more of the following actions: garnishment of your [redacted], wages, and/or bank account(s), or Municipal referral of your account to an outside collection agency.

If you have any questions, please feel free to contact us at [redacted].

This is an attempt to collect a debt and any information obtained will be used for that purpose.

(b) Nudge treatment arm communication

[redacted]

PAYMENT DUE: August 31, 2017

MUNICIPALITY OF [redacted]
TREASURY DIVISION

August 4, 2017

This is A Test Case
123 Test St
[redacted]

[redacted]

PAY YOUR COURT-ORDERED CRIMINAL FINES/FEES NOW

DELINQUENT AMOUNT DUE: $ 96.67*

HOW TO PAY

Pay Online:	**www.**[redacted]
Pay by Phone:	**1** [redacted]
Pay by Mail:	**Check or Money Order Payable To:** [redacted] **P.O. Box** [redacted]

Note: Reference your [redacted] Case Number on check or money order for prompt processing. You may also use the enclosed postage-paid envelope to mail your payment. A convenience fee is assessed by MSB for electronic payment services. Credit/debit card transactions may appear as charges from Gila Corporation on your bank or credit card statement.

INFORMATION YOU NEED TO MAKE A PAYMENT

[redacted] **Case Number:** 3AN-77-77777MO [redacted] **Ticket Number: A123456** **Offense Date: 01/01/2000**

Charge: AMC9.22.040(C): Stop-Decree Speed Without Notice To Rear Driver

Important: Delinquent cases, including the amount still due, can be viewed by anyone (examples: employers, landlords, and insurance companies) by visiting the public records website at: www.[redacted]

IF YOU DON'T PAY NOW, WE CAN GARNISH THE FOLLOWING:

Your PFD, wages, and/or bank account(s).

TROUBLE PAYING?

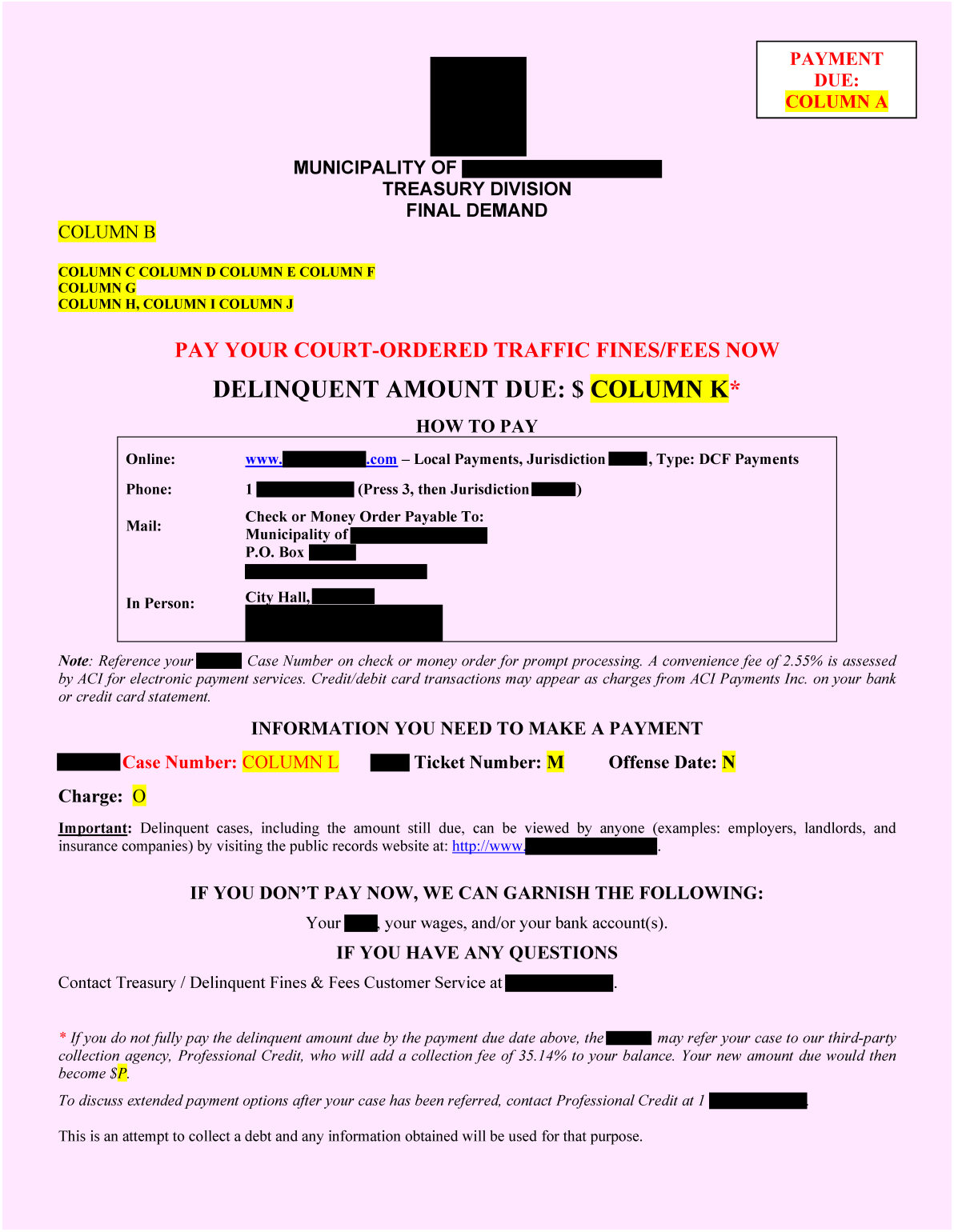
Contact Treasury / Delinquent Fines & Fees Customer Service at [redacted]

* *Your case has been referred to our third-party collection agency, MSB, so the delinquent amount due may not include MSB's full collection fee. Contact MSB for exact balance due. To discuss extended payment options, contact MSB at 1* [redacted]

This is an attempt to collect a debt and any information obtained will be used for that purpose.

Figure D.1: Example of adoption of BIT-NA trial

(c) Current communication



PAYMENT DUE: COLUMN A

[redacted]

MUNICIPALITY OF [redacted]
TREASURY DIVISION
FINAL DEMAND

COLUMN B

COLUMN C COLUMN D COLUMN E COLUMN F
COLUMN G
COLUMN H, COLUMN I COLUMN J

PAY YOUR COURT-ORDERED TRAFFIC FINES/FEES NOW

DELINQUENT AMOUNT DUE: $ COLUMN K*

HOW TO PAY

Online:	www.[redacted].com – Local Payments, Jurisdiction [redacted], Type: DCF Payments
Phone:	1 [redacted] (Press 3, then Jurisdiction [redacted])
Mail:	Check or Money Order Payable To: Municipality of [redacted] P.O. Box [redacted] [redacted]
In Person:	City Hall, [redacted] [redacted]

***Note**: Reference your [redacted] Case Number on check or money order for prompt processing. A convenience fee of 2.55% is assessed by ACI for electronic payment services. Credit/debit card transactions may appear as charges from ACI Payments Inc. on your bank or credit card statement.*

INFORMATION YOU NEED TO MAKE A PAYMENT

[redacted] **Case Number:** COLUMN L [redacted] **Ticket Number: M** **Offense Date: N**

Charge: O

Important: Delinquent cases, including the amount still due, can be viewed by anyone (examples: employers, landlords, and insurance companies) by visiting the public records website at: http://www.[redacted].

IF YOU DON'T PAY NOW, WE CAN GARNISH THE FOLLOWING:

Your [redacted], your wages, and/or your bank account(s).

IF YOU HAVE ANY QUESTIONS

Contact Treasury / Delinquent Fines & Fees Customer Service at [redacted].

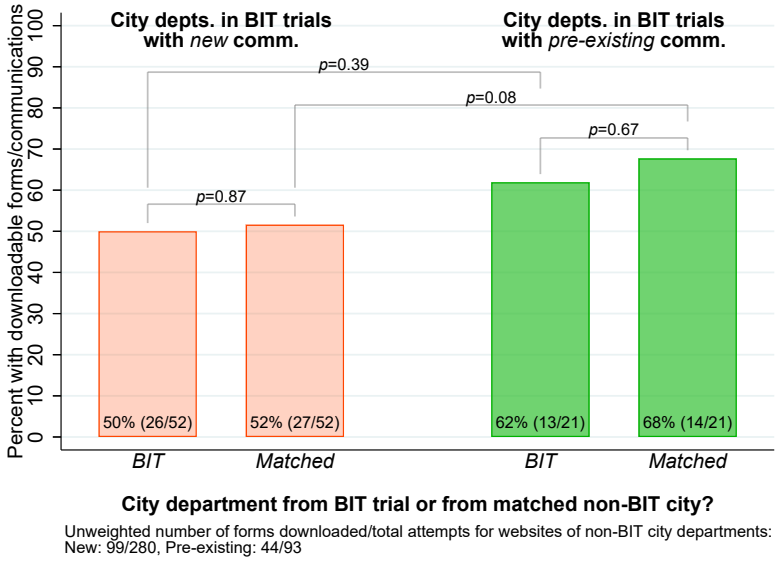
* *If you do not fully pay the delinquent amount due by the payment due date above, the [redacted] may refer your case to our third-party collection agency, Professional Credit, who will add a collection fee of 35.14% to your balance. Your new amount due would then become $P.*

To discuss extended payment options after your case has been referred, contact Professional Credit at 1 [redacted].

This is an attempt to collect a debt and any information obtained will be used for that purpose.

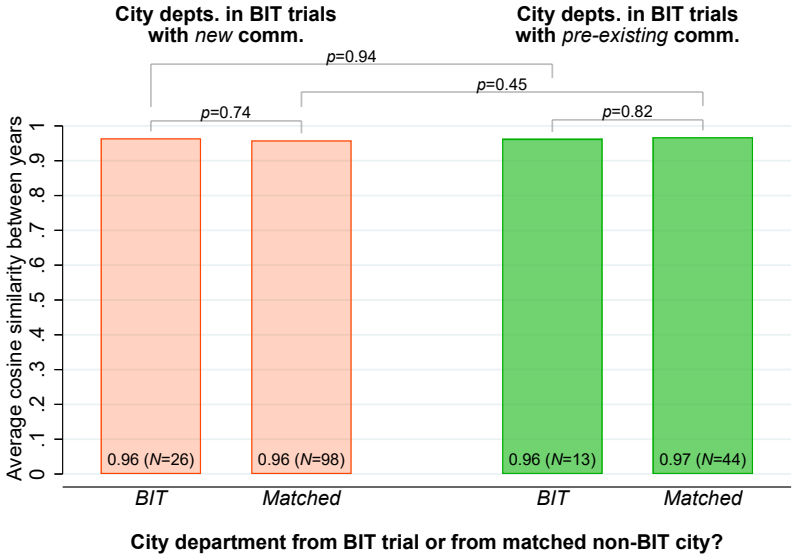
Figure D.2: Availability and updating of public forms/communications on city websites

(a) Percent with downloadable forms/communications available



Unweighted number of forms downloaded/total attempts for websites of non-BIT city departments: New: 99/280, Pre-existing: 44/93

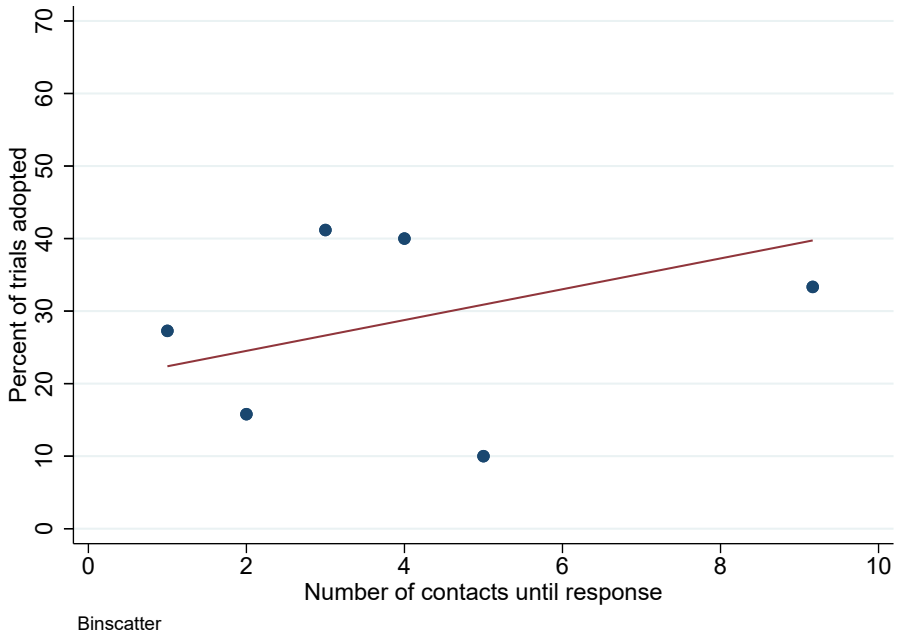
(b) Cosine similarity in forms/communications across years



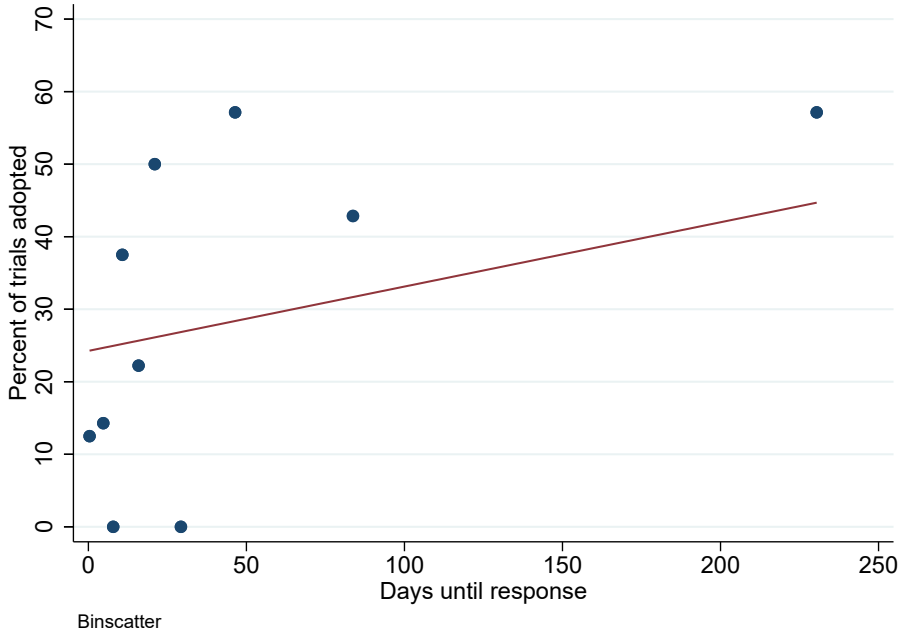
These figures summarize the availability and rate of change in the sample of citizen-facing online forms/communications collected from city department websites. See Online Appendix D.2 for more details. Figure D.2a shows the percent of city departments that had a citizen-facing form/communication on their websites downloadable at least once during 2015-19 and also at some point after 2019. Figure D.2b reports the average graphical and textual cosine similarity in the forms between each sequence of years for which the forms were downloaded. Standard errors are clustered by the (matched) BIT city. Non-BIT city departments are inversely weighted by the number of non-BIT city departments matched to the same BIT trial.

Figure D.3: Adoption by response times (bin scatters)

(a) Number of times contacted until final response

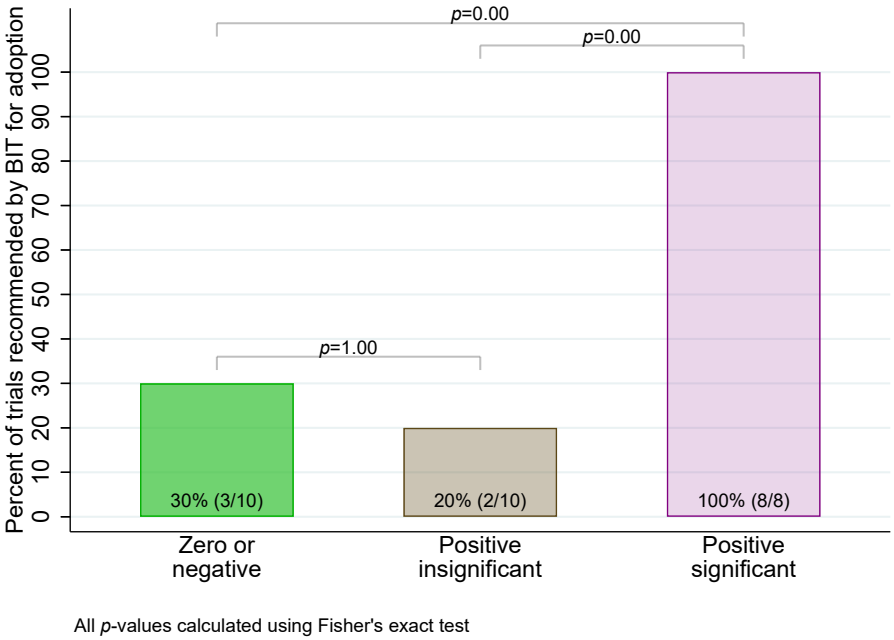


(b) Number of days from first request to final response



Figures D.3a and D.3b show binscatters relating the adoption rate to (a) the number of exchanges with the city contact (e.g., emails and phone calls) and (b) the days from first contact to final response, until all the information on the trial was collected.

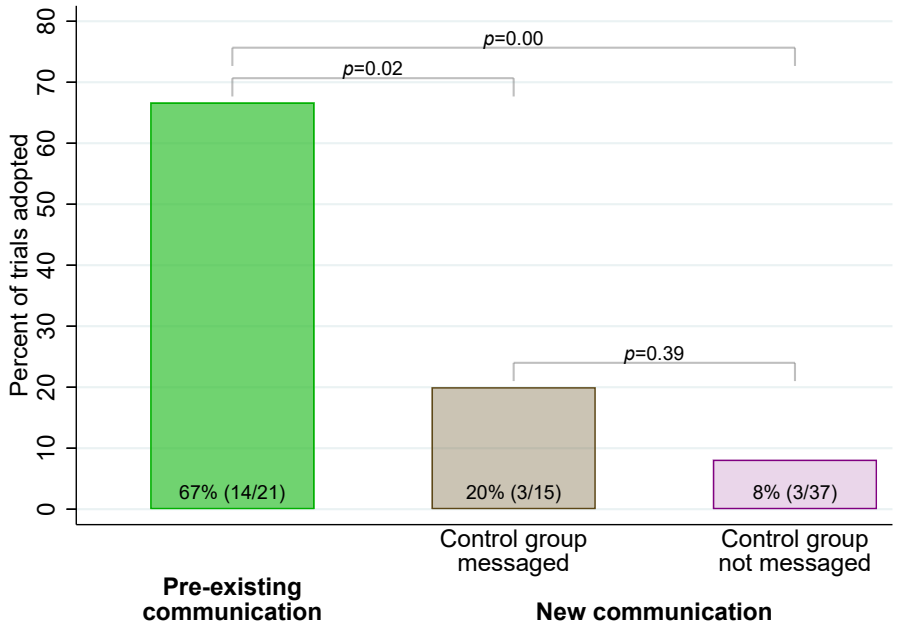
Figure D.4: BIT recommendations for adoption by statistical significance



This figure shows the percent of trials that BIT recommended for adoption within three groups separated by the sign and significance of the best-performing nudge treatment effect. BIT began documenting recommendations in their trial reports in mid-2017. 28 trials in the sample have these recommendations.

Figure D.5: Adoption of nudges by pre-existence: Additional results

(a) By control group communication



(b) By effect size (three bins)

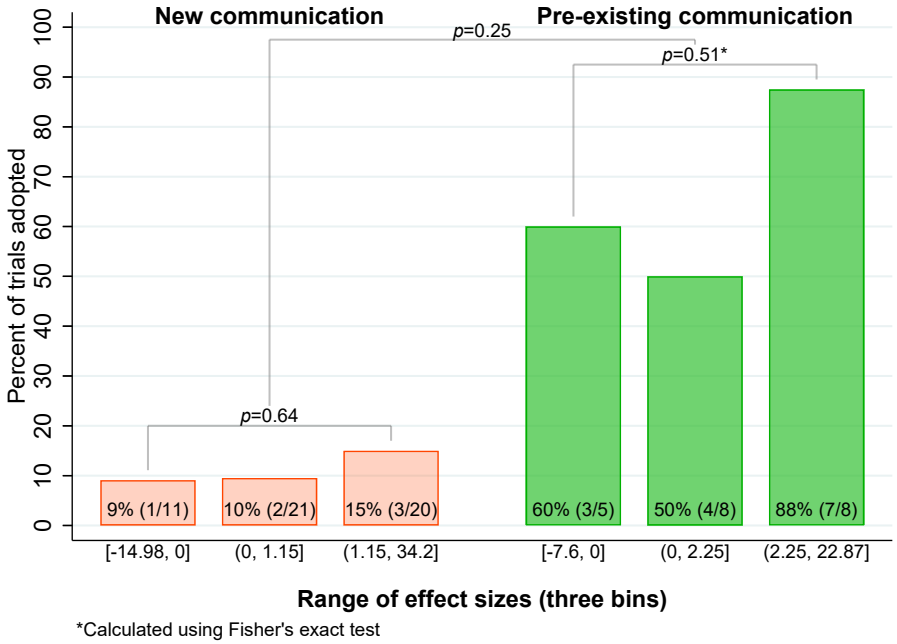
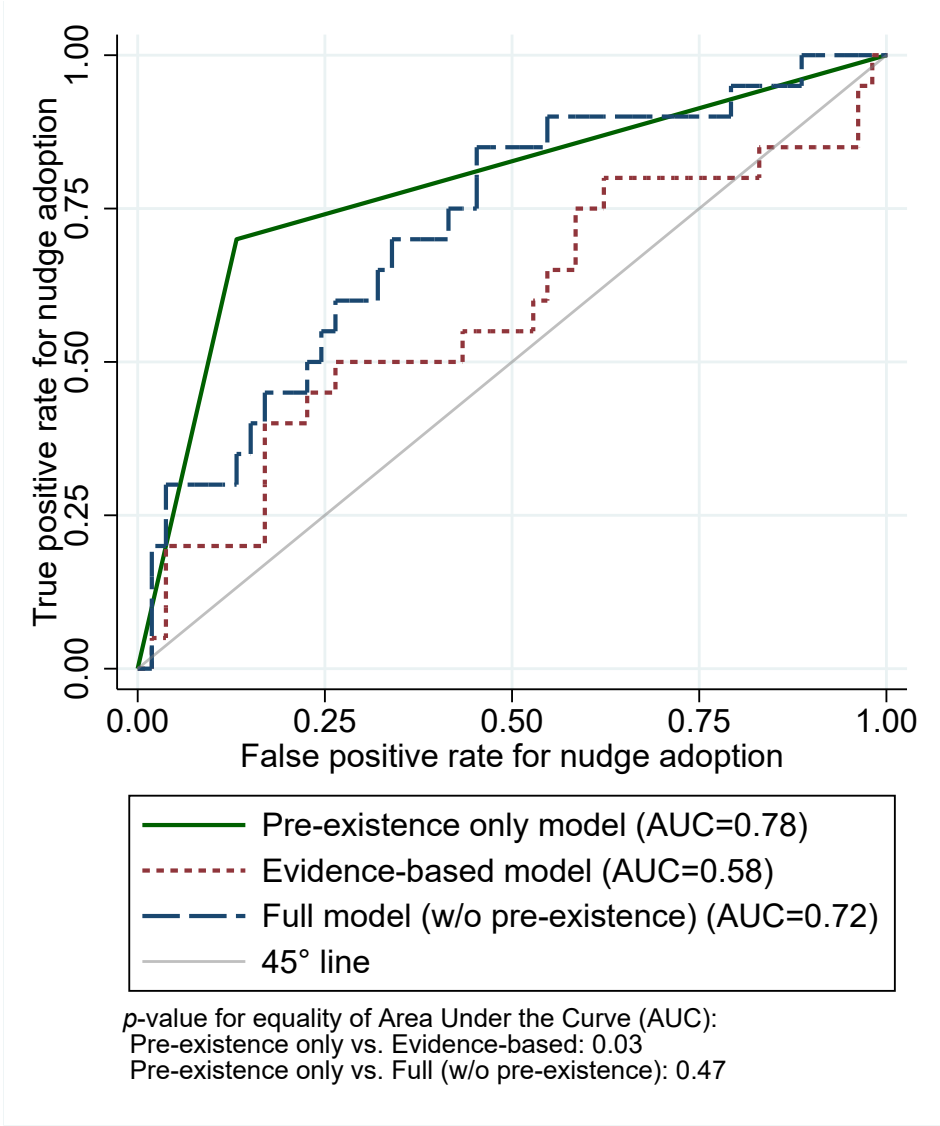


Figure D.5a shows the adoption rate for trials on new communications conditional on whether the control group received any communication in the trial. Figure D.5b bins together trials that found a negative or zero effect, and those that are below or above the median among the trials with a positive effect. It compares the adoption rates within each bin for new and pre-existing communication trials separately.

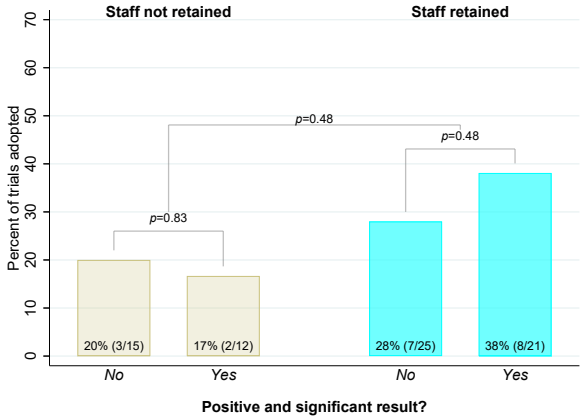
Figure D.6: Receiver operating characteristic (ROC) curves



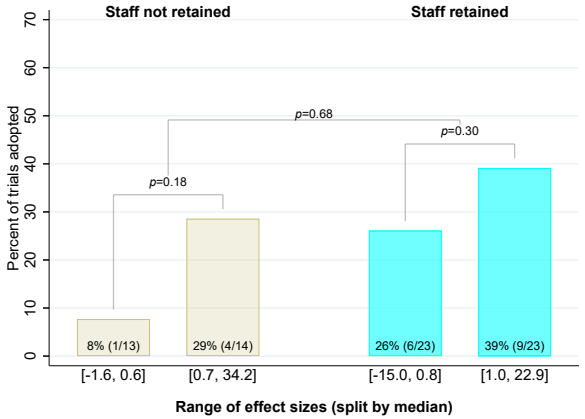
This figure shows the receiver operating characteristic (ROC) curves and the computes the area under the curve (AUC) for three models estimated by logistic regression. Each model includes a constant and different sets of explanatory variables.
The *Pre-existence only* model includes an indicator for whether the communication in the trial was pre-existing.
The *Evidence-based* model includes an indicator for the whether the most effective treatment arm in the trial was positive and significant (i.e. max $t \geq 1.96$) as well as the percentage-point treatment effect.
The *Full without pre-existence* model includes all the controls in Column 4 of Table 5.3 except for the pre-existing communication indicator.

Figure D.7: Interaction effects: Staff retention

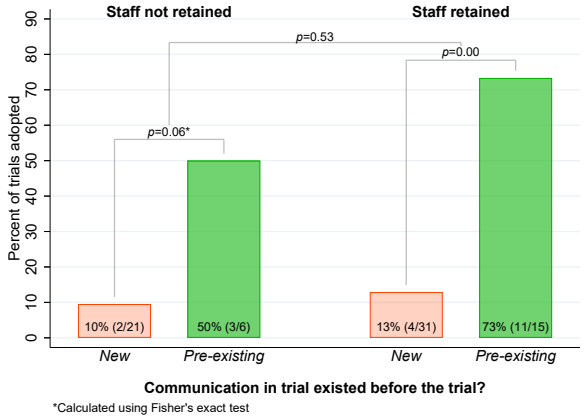
(a) By staff retention and statistical significance



(b) By staff retention and effect size



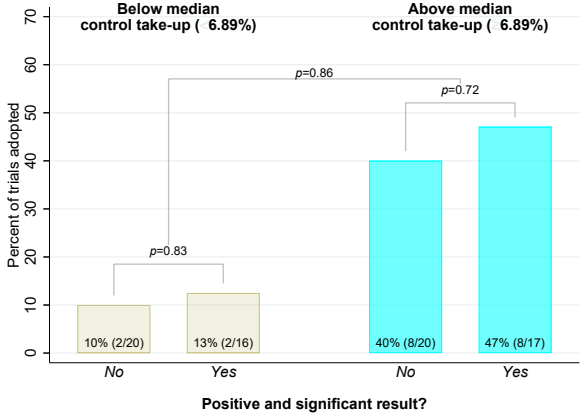
(c) By staff retention and pre-existence



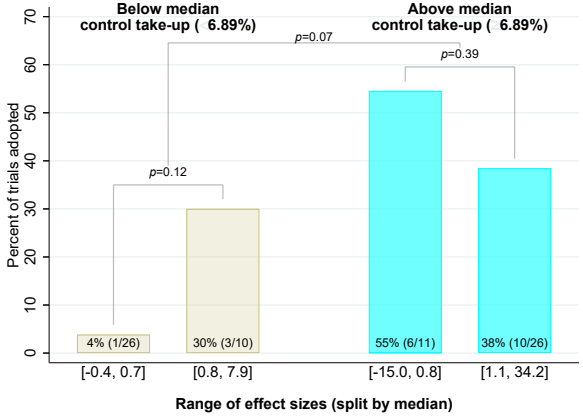
Figures D.7a and D.7b show the adoption rates conditional on finding an effect that is (a) positive and significant and (b) above or below the median effect size, separately for trials where the original city collaborator has left or has been retained by the city department. Figure D.7c compares the adoption rate of interventions in new (orange) versus pre-existing (green) communications separately for cases when the original city collaborator has left or has been retained.

Figure D.7: Interaction effects: Control take-up

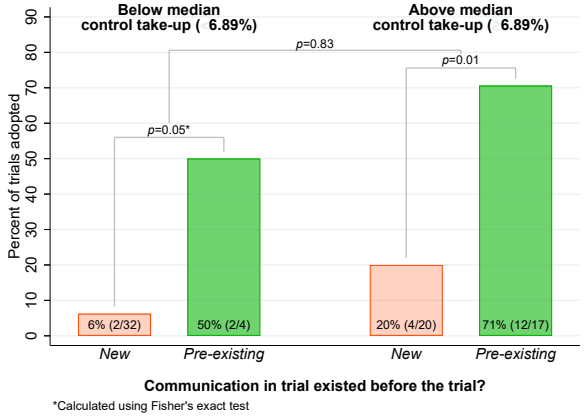
(d) By control take-up and statistical significance



(e) By control take-up and effect size

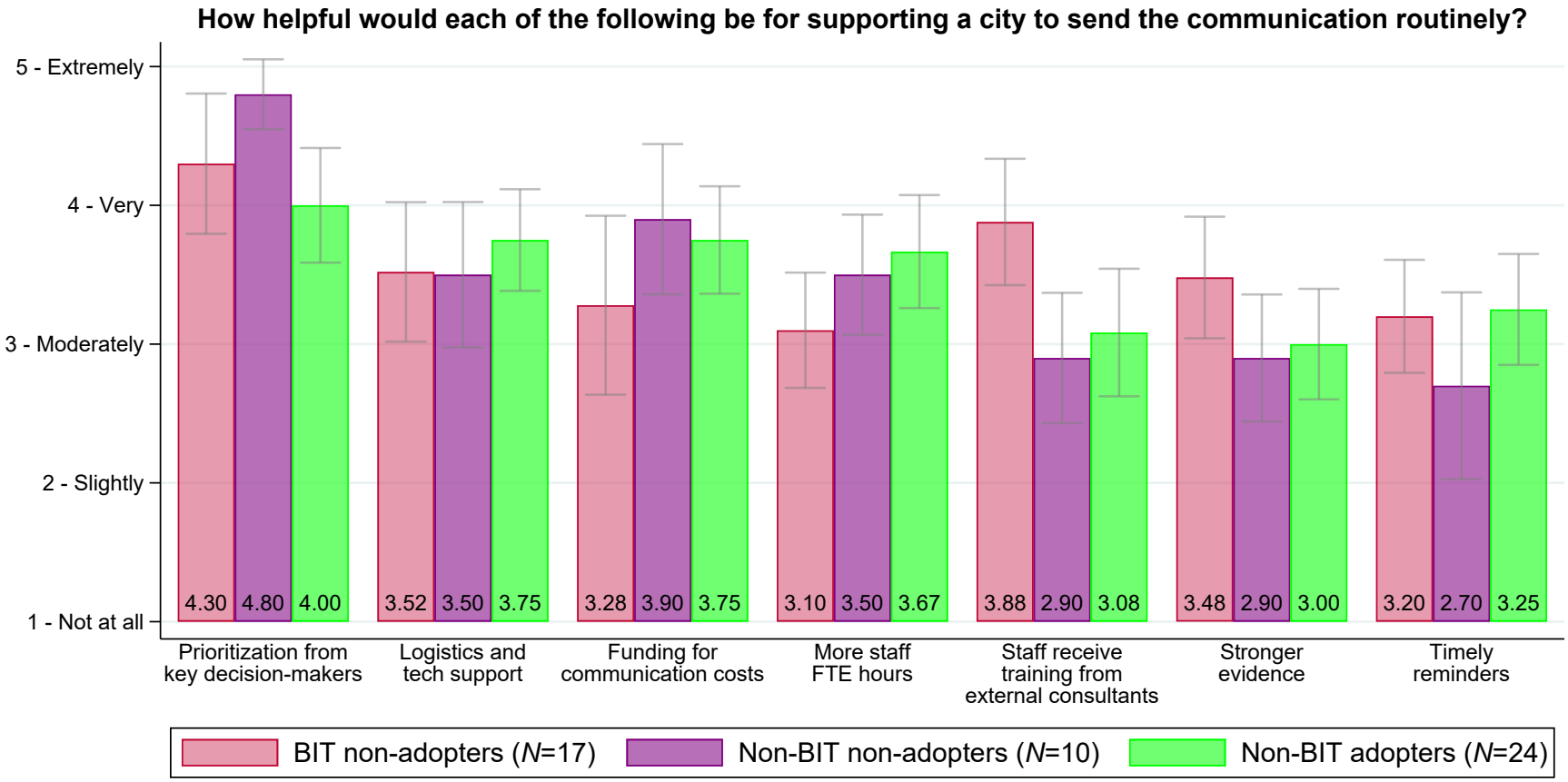


(f) By control take-up and pre-existence



Figures D.7d and D.7e show the adoption rates conditional on finding an effect that is (a) positive and significant and (b) above or below the median effect size, separately for trials above and below the median in the control group take-up rate. Figure D.7f compares the adoption rate of interventions in new (orange) versus pre-existing (green) communications separately for trials above and below the median in the control group take-up rate.

Figure D.8a: Survey evidence on organizational inertia: Non-adopters vs. adopters



This figure splits the sample of respondents in Figure 5.10 into (i) staff from BIT-trial cities where a nudge was not adopted despite having an effect size that was positively significant or at least 1 pp. in the trial ("BIT non-adopters"), (ii) staff from U.S. cities outside the BIT sample that received evidence on an effective communication but did not implement it ("non-BIT non-adopters"), and (iii) staff from U.S. cities outside the BIT sample that implemented a communication based on evidence of its effectiveness ("non-BIT adopters"). 11 respondents from Figure 5.10 who did not report whether their city had or had not adopted evidence are dropped. 95% confidence intervals are shown with bootstrap standard errors.

Figure D.8b: Survey evidence on organizational inertia: Word cloud



This word cloud is based on the responses in the survey represented in Figure 5.10 to the open-ended question "In general, what do you think are the main obstacles that could prevent cities from putting evidence into practice?" The size of the words is proportional to their frequency in the responses.

Figure D.9: Hjort et al. (2021) policy adoption experiment: Policy brief for tax payment reminder letter (reproduced)





HOW TO INCREASE COMPLIANCE WITH LOCAL TAXES

A Policy Brief Based on Scientific Research

INTRODUCTION

Raising tax revenue locally is an important task for municipal governments in Brazil. Local taxes increase the municipal budget, but also provide untied funds which the municipality can spend in line with its own priorities. But municipalities in Brazil face a serious challenge when it comes to collecting local taxes: many businesses and individuals who owe tax payments do not comply with the tax laws by paying the full amounts on time.

Governments throughout the world, including Brazil, have tried many innovative methods to solve this problem. But what works, and what does not? This policy brief provides simple results from scientific research on how governments can increase compliance with taxes.

A LOW-COST AND EFFECTIVE WAY TO INCREASE TAX COMPLIANCE: REMINDER LETTERS

Research conducted in Latin America has revealed one very simple and inexpensive action that has proven to be effective in increasing compliance: **sending taxpayers reminder letters before the due date of the taxes.**[1] For example, an academic researcher worked with two municipal governments in Peru, and found that property tax compliance increased by 10% simply by sending a letter to taxpayers which reminded them of the tax payment deadline![2] Similar results have been found in other studies, including in the United States, Austria and the United Kingdom.[3]

1 Taxpayers are those legally responsible to pay taxes. For instance, taxpayers of the urban property tax (IPTU) are the owners of the property (or the tenants if is explicitly stated in the lease agreement). Taxpayers of the services of any nature tax (ISSQN), are the professionals or businesses that provide the service.
2 Del Carpio (2013)
3 Coleman (1996), Hallsworth et al. (2014), Fellner et al. (2013)

Research can also guide how to make the reminder letters even more effective. An important policy lesson is that **the letter should emphasize that most people pay their taxes on time.** The same study in Peru found that tax compliance increased by 20% if the reminder letter also included a sentence like "The vast majority of your neighbors pay their taxes on time!" or "75% of your neighbors pay their taxes on time!" Such a message highlights that paying taxes on time is a "social norm", and those who don't pay are deviating from the desirable social norm.

There is one final lesson from research on how to increase the effectiveness of tax reminder letters: **highlight the threat of audits or penalties due to not paying taxes on time.** For example, a study in Argentina found that sending out a letter to property owners (who are supposed to pay property taxes) emphasizing the possible fines and audits due to evading taxes increased tax compliance by 12%.[4]

An important point to keep in mind is that reminder letters are inexpensive to send. All that is needed is for the municipal tax authorities to know the addresses of potential taxpayers. In many cases, letters are already being sent to such taxpayers.

Simply by choosing the correct content of the letter, to remind taxpayers of the payment deadline, to emphasize social norms, and to highlight the threat of audits or penalties, governments have been able to increase tax compliance and revenues, and reduce tax evasion. This can be a very cost-effective policy, and is moreover easy to implement compared to most other strategies to increase tax revenues.[5]

POLICY LESSONS

To summarize, this brief provides a total of **three policy lessons:**

- Send letters to taxpayers reminding them of the deadline to pay taxes.
- Emphasize in the letter that most people pay their taxes on time.
- Highlight the potential bad consequences of avoiding taxes: fines and audits.

An example letter is provided on Page 3 of this policy brief. Contact the Project team at **contato@pesquisadoresdeharvardcnm.com** to receive an electronic copy of the letter.

4 Castro and Scartascini (2013)
5 A cost-effective action is one that produces good results with a small cost.

Figure D.9: Hjort et al. (2021) policy adoption experiment: Template for tax payment reminder letter (reproduced)

EXAMPLE
REMINDER
LETTER FOR
TAX PAYMENT

A Policy Brief
Based on
Scientific
Research

FEATURING PAYMENT DEADLINE, SOCIAL NORMS, AND THREAT OF PENALTIES

Dear Sir/Madam,

Your municipal tax payments are due by **01 November 2016.**

Our statistics show that the **vast majority of your neighbors will pay their taxes on time.** We greatly appreciate your doing the same.

Don't forget to report your taxes accurately and in a timely manner to avoid the **risk of an audit,** which is a time-consuming and costly process that may lead to substantial financial and other penalties if your tax reporting is found to be wrong.

It is easy to pay your taxes. Please follow the enclosed instructions for more information.

If you have already paid your taxes, thank you very much! If not, please act now.

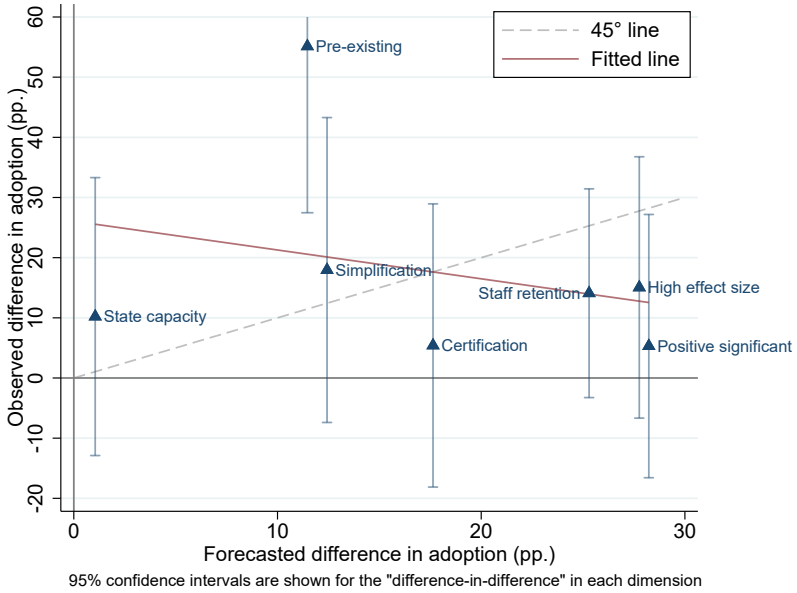
Yours faithfully,
Name of Tax Authority



How to Increase Compliance with Local Taxes • 3

Figure D.10: Comparisons of forecasts and observed adoption

(a) By dimension



(b) Ranking of adoption compared to firms and governments of low-income countries

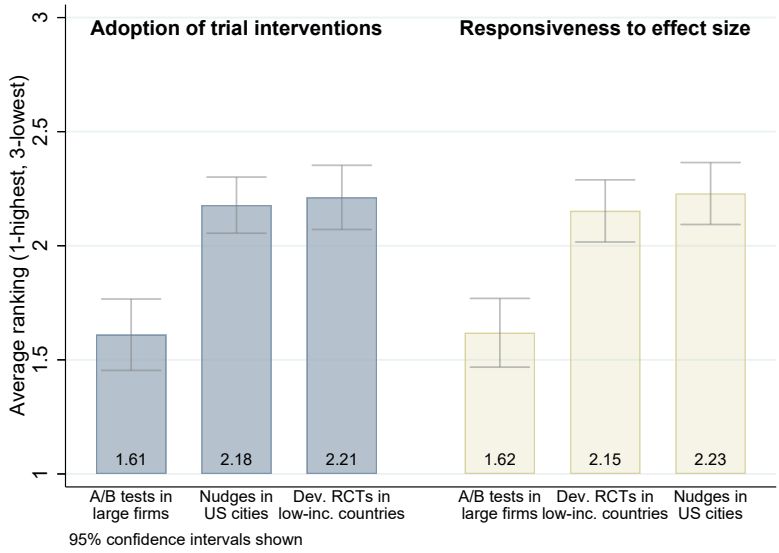


Figure D.10a evaluates the accuracy of the expert forecasts across determinants and plots the predicted impact of each determinant (horizontal axis) against the observed impact (vertical axis). Figure D.10b shows the average of the forecasters' ranking of adoption rates for three samples: (1) nudges in U.S. cities (this paper), and similar hypothetical representative samples of trials conducted by (2) large multinational firms and (3) governments of low-income countries. The left side shows the average rankings for the overall adoption rate (where 1 corresponds to the highest adoption and 3 to the lowest), and the right side shows the ranking for each sample's responsiveness to evidence in adoption.

Table D.1: City features of completed vs. abandoned trials

	Completed trials	Abandoned trials	Difference (SE)	Diff. p-value
Log(pop)	13.00	12.79	0.22 (0.24)	0.37
Median age of residents	36.08	33.72	2.36 (0.85)	0.01
Median household income	63348.16	67981.22	-4633.06 (6007.24)	0.45
Median property value	329809.59	391200.00	-61390.41 (84489.91)	0.47
Employed percent	49.82	51.01	-1.19 (1.33)	0.38
White percent	48.43	50.23	-1.80 (4.61)	0.70
Tax revenue per capita	1562.32	2185.21	-622.89 (449.16)	0.17
What Works Cities certification	0.60	0.63	-0.03 (0.13)	0.83
Trial pre-existed	28.77	40.00	-11.23 (15.92)	0.49
	73 trials (30 cities)	27 trials (19 cities)		

Standard errors are clustered by city. There are 7 cities in the Abandoned trials sample that are not in the Completed trials sample. The remaining cities in the Abandoned trials sample are also represented in the Completed trials sample.

Table D.2: Average nudge treatment effects

	Nudge Units* (1)	Updated BIT-NA (2)
Average treatment effect	1.390	1.906
	(0.304)	(0.587)
Nudges	241	116
Trials	126	73
Observations	23,556,095	1,800,382
Average control group take-up (%)	17.33	15.07
Distribution of treatment effects		
25th percentile	0.06	0.01
50th percentile	0.50	0.40
75th percentile	1.40	1.72

This table shows the average treatment effect of nudges. Standard errors clustered by trial are shown in parentheses. pp. refers to percentage point. *Column 1 replicates Column 2 of Table III in DellaVigna and Linos (2022).

Table D.3: Sample characteristics: City features

	Overall	Effect size>median		City staff retained		Comm. pre-existed	
Frequency in category (%)	(1)	(2) No	(3) Yes	(4) No	(5) Yes	(6) No	(7) Yes
City characteristics							
Log(pop)	13.00	12.91	13.10	12.70	13.18	12.97	13.09
Median age of residents	36.08	36.63	35.51*	35.58	36.37	36.46	35.13
Median household income	63348.16	62022.38	64710.78	59824.63	65416.33	64774.40	59816.52
Median property value	329809.59	334472.97	325016.67	277892.59	360282.61	352273.08	274185.71
Employed percent	49.82	49.38	50.28	49.44	50.04	49.81	49.85
White non-Hispanic percent	48.43	48.15	48.72	49.66	47.72	48.20	49.02
Tax revenue per capita	1562.32	1717.54	1402.80	1770.97	1439.86	1635.43	1381.30
	73	37	36	27	46	52	21

This table reports the average city-level features for trials in the sample. Column 1 shows the sample averages. Columns 2 and 3 partition the sample along the median of the maximum effect size in each trial. Columns 4 and 5 consider separately trials for which all the city collaborators from the trial have departed versus trials that have at least one original staff member still working in the same city department. Columns 6 and 7 distinguish between trials that tested nudges in a new communication and those that added nudges to a pre-existing communication that the city had been sending before the trial.

*Asterisk indicates that the p-value of the difference<0.05. Standard errors are clustered by city.

Table D.4: Forecasts summary

	Observed	Forecasts (Mean % [SD])			
Category	(%)	(1) Overall ($N = 118$)	(2) Nudge unit staff ($N = 19$)	(3) Reseachers ($N = 67$)	(4) Government workers ($N = 14$)
Baseline adoption rate	27.40	32.47 [19.06]	37.16 [20.64]	31.91 [19.16]	32.00 [20.40]
By sign and significance:					
Positive & significant	30.30	46.58 [23.66]	48.00 [23.91]	46.49 [23.13]	44.29 [28.01]
Positive & insignificant	25.00	23.22 [20.27]	34.84 [23.51]	21.31 [18.59]	23.29 [22.61]
Zero or negative	25.00	11.06 [16.21]	17.47 [19.25]	10.43 [16.35]	12.93 [17.40]
By effect size:					
High third	37.50	49.31 [24.81]	54.26 [24.80]	48.85 [24.15]	45.43 [30.89]
Middle third	28.00	31.61 [19.60]	40.32 [23.51]	29.57 [16.95]	32.29 [25.08]
Low third	16.67	12.94 [15.42]	18.16 [18.39]	12.60 [15.80]	11.57 [11.88]
By staff retention:					
With original staff	32.61	43.75 [22.64]	48.26 [26.59]	42.13 [20.74]	44.07 [28.02]
Without original staff	18.52	18.45 [16.31]	24.32 [15.14]	17.51 [16.33]	19.71 [18.59]
By state capacity (proxied by 2020 city population size):					
Above median	32.43	33.26 [19.62]	40.16 [22.93]	32.07 [17.76]	31.79 [24.12]
Below median	22.22	32.21 [18.67]	35.84 [19.25]	32.15 [18.89]	29.50 [20.26]
By What Works Cities certification:					
Certified	29.55	41.69 [21.23]	45.26 [22.86]	40.06 [20.37]	42.93 [24.94]
Not certified	24.14	24.06 [17.90]	32.42 [20.61]	22.58 [17.21]	22.14 [17.16]
By pre-existing or new communication:					
New	11.54	29.79 [20.42]	36.32 [22.87]	29.48 [19.48]	25.50 [19.00]
Pre-existing	66.67	41.25 [25.43]	45.63 [26.68]	37.78 [23.11]	47.71 [28.97]
By behavioral mechanism:					
Simplification	33.33	42.42 [22.06]	51.16 [24.34]	40.21 [20.56]	41.00 [25.45]
Personal motivaton	19.05	30.14 [19.30]	35.84 [20.55]	28.37 [18.35]	26.79 [19.42]
Social cues	24.39	29.83 [20.97]	34.74 [23.05]	28.81 [19.58]	30.36 [22.01]

Table D.5: BIT recommendations for trial adoption

Dep. Var.:	BIT recommended adopt	Trial adopted
(OLS)	(1)	(2)
Max $t \geq 1.96$	0.71	
	(0.12)	
Max treatment effect (10pp.)	0.05	
	(0.06)	
BIT recommended for adoption		0.25
		(0.17)
BIT did not recommend for adoption		-0.16
		(0.11)
Communication pre-existed		0.45
		(0.14)
Constant	0.25	0.13
	(0.10)	(0.05)
Average recommendation rate	0.46	0.18
Trials with recommendations only	✓	
Number of trials	28	73
Number of cities	16	30
R^2	0.46	0.38

Standard errors clustered by city are shown in parentheses. BIT has included recommendations for or against adoption in their trial reports since mid-2017. In Column 2, the omitted group are the earlier trials without BIT recommendations in the trial reports.

Table D.6: Determinants of nudge adoption (robustness)

	Baseline	Robust SEs	Robustness to marginal cases		
Dep. Var.: Nudge adopted (0/1, OLS)	(1)	(2)	(3)	(4)	(5)
Max $t \geq 1.96$	-0.03	-0.03	-0.06	0.06	0.03
	(0.08)	(0.10)	(0.09)	(0.10)	(0.10)
Max treatment effect (10pp.)	0.10	0.10	0.11	0.02	0.02
	(0.07)	(0.07)	(0.07)	(0.05)	(0.05)
City staff retained	0.07	0.07	0.04	0.02	-0.02
	(0.08)	(0.09)	(0.09)	(0.07)	(0.07)
Above-median city population	0.06	0.06	0.05	0.03	0.02
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)
What Works Cities certified	0.14	0.14	0.18	0.06	0.11
	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)
Communication pre-existed	0.52	0.52	0.56	0.40	0.45
	(0.13)	(0.12)	(0.13)	(0.18)	(0.18)
Mechanism					
Simplification & information	0.03	0.03	0.04	0.10	0.11
	(0.10)	(0.10)	(0.09)	(0.06)	(0.07)
Personal motivation	-0.12	-0.12	-0.15	-0.11	-0.14
	(0.12)	(0.10)	(0.11)	(0.07)	(0.07)
Social cues	-0.08	-0.08	-0.09	-0.10	-0.13
	(0.08)	(0.10)	(0.09)	(0.10)	(0.10)
Constant	0.05	0.05	0.10	0.03	0.07
	(0.15)	(0.14)	(0.16)	(0.12)	(0.12)
Average adoption rate	0.27	0.27	0.29	0.15	0.16
Dropping marginal non-adopts			✓		✓
Dropping verbal-only adopts				✓	✓
Number of trials	73	73	69	62	58
Number of cities	30	30	29	29	27
R^2	0.38	0.38	0.41	0.37	0.42

Standard errors (shown in parentheses) are clustered at the city level except in Column 2, which provides heteroskedastic-robust standard errors. Baseline replicates Column 4 of Table 2. See Appendix D for details on marginal non-adoption and verbal-only adoption cases.

Table D.7: Sample characteristics: Survey of non-adopters

	Overall	Responded	
Frequency in category (%)	(1)	(2) No	(3) Yes
Nudge effectiveness			
Max $t \geq 1.96$	74.19	83.33	72.00
Max treatment effect ≥ 1 pp.	70.97	83.33	68.00
Organizational features			
City certified by What Works Cities	64.52	16.67	76.00*
City staff member from trial retained	58.06	33.33	64.00
Partner city dept. in charge of implementing	83.87	83.33	84.00
Senior city staff on trial (Director/Chief)	51.61	16.67	60.00
Experimental design			
Communication pre-existed before trial	12.90	16.67	12.00
Nudge communication uses Simplification	48.39	66.67	44.00
Nudge communication uses Personal Motivation	70.97	83.33	68.00
Nudge communication uses Social Cues	54.84	50.00	56.00
Policy area			
Revenue collection & debt repayment	25.81	50.00	20.00
Registration & regulation compliance	19.35	33.33	16.00
Workforce & education	19.35	16.67	20.00
Take-up of benefits and programs	12.90	0.00	16.00
Community engagement	9.68	0.00	12.00
Health	9.68	0.00	12.00
Environment	3.23	0.00	4.00
Medium			
Physical letter	32.26	50.00	28.00
Email	29.03	16.67	32.00
Postcard	32.26	33.33	32.00
Text message	16.13	0.00	20.00
Website	0.00	0.00	0.00
Number of trials	31	6	25

This table shows the frequencies of trials for each category listed in the leftmost column. Column 1 shows the frequences for all trials that had a treatment effect size ≥ 1 pp. or t-stat > 1.96 but were not adopted by the city. Columns 2 and 3 show the frequencies separately for cases when the did or did not city respond to the survey for Figure 10.

*Asterisk indicates that the p-value of the difference<0.05. Standard errors are clustered by city. Except when there are fewer than 5 trials in one of the 2×2 cells, p-values are calculated using the two-sided Fisher's exact test instead.

Table D.8a: Hjort et al. (2021) policy adoption experiment: Control and treatment rates by adoption definition

Adoption definition	All 3 mechanisms		$\geq$ 2 of 3 mechanisms		Social cues	
% (N)	Control	Treatment	Control	Treatment	Control	Treatment
No Letter + No Nudge	80.43% (945)	77.83% (853)	73.36% (862)	69.80% (765)	79.40% (933)	76.28% (836)
Letter + No Nudge	10.38% (122)	10.22% (112)	2.55% (30)	1.92% (21)	9.28% (109)	8.21% (90)
No Letter + Nudge	4.00% (47)	3.65% (40)	11.06% (130)	11.68% (128)	5.02% (59)	5.20% (57)
Letter + Nudge	5.19% (61)	8.30% (91)	13.02% (153)	16.61% (182)	6.30% (74)	10.31% (113)

This table shows the frequency and number of mayors and city staff stating whether their city sends a tax payment reminder communication and what language it contains. The 3 mechanisms mentioned in the template for the tax reminder letter are the due date, the threat of audits or fines, and social norm language.

Table D.8b: Hjort et al. (2021) policy adoption experiment: Balance in letter adoption (control group)

	Avg. in No Letter Group	Δ in Letter Group	p-value
Mayor's Characteristics			
Male	90.52	-4.47	0.41
Age	46.96	0.67	0.67
College or more	57.76	-8.92	0.26
2nd Term	14.66	-5.35	0.25
Electoral Margin Victory	16.73	-0.27	0.94
Leftist Political Party	33.45	-12.52	0.06
Municipalities' Characteristics			
Population	20.99	0.84	0.75
College Population	5.54	-0.28	0.56
Public Adm College	33.51	-1.68	0.40
Poverty	23.56	-0.85	0.75
Gini	49.66	-0.68	0.48
Big South	58.79	-0.65	0.93
Per Capita Income	488.38	-12.33	0.71
Local Taxes Revenue (2010-15)	6.40	0.51	0.54
Placebo Adoption			
Use of E-Procurement	0.52	-0.04	0.71
P(Use of E-Procurement$\|X$)	0.52	0.01	0.55
Joint F-test			0.31
N	580	43	

This table compares observable charactertistics between the municipalities in the Control Group from Hjort et al. (2021) that were not sending a taxpayer reminder letter with those that were. Municipalities in which the mayor and the city staff gave conflicting responses to letter adoption are excluded. Δ in Letter Group is the difference in the means from the No Letter Group to the Letter Group. The p-value is computed from robust standard errors. The Use of E-Procurement is considered as a placebo adoption outcome in Hjort et al. (2021). P(Use of E-Procurement$|X$) is the predicted probability of adopting E-Procurement based on a logit model including all listed Mayor and Municipalities' characteristics. The Joint F-test includes all variables in the table.