# Stanford Economic Review

*Stanford Undergraduate Economics Journal*

*Winter 2023*

*Volume 11*

## Editors and Staff

**Note from the Editor**

---

On behalf of the *Stanford Economic Review* Editorial Board, I am pleased to present the eleventh volume, winter issue, of Stanford University's undergraduate economics journal.

Building on our momentum from last year, our publication has continued expanding its global reach over the course of the 2022-2023 academic year. As our readership climbs to new heights, we have remained steadfast in our commitment to publishing both exceptional empirical research and incisive analyses of modern economic issues.

This journal issue spotlights undergraduate work on a wide variety of topics ranging from electric vehicle adoption in California to asymmetric matching markets on 7 Cups, a social-emotional support site. In addition to the six original research papers in this volume, commentary pieces written for our publication over the last few months have evaluated popular domestic policy proposals like raising the minimum wage and implementing larger-scale basic income programs and have analyzed important political and economic developments including China's recent housing crisis and rising homelessness in Los Angeles.

As always, we are incredibly grateful to the authors whose writing is featured in this journal edition and on the commentaries section of our website. Lastly, we would like to thank the Stanford Economics Association (SEA) and the Stanford Economics Department for their continued support.

Karthick Arunachalam
*2022-23 Editor-in-Chief*

# Contents

# Do Public Chargers Accelerate Mass EV Adoption? Evidence from California

Robert Huang

*University of Southern California*

*Abstract*— **In the United States, the transportation sector contributes to 30% of the total emissions, 58% of which are produced by private passenger vehicles. One barrier of mass electric vehicle adoption is the lack of public chargers. Using a panel dataset on over 1800 Californian ZIP codes from 2010 to 2021, I employ a shift-share instrumental variable to estimate the EV demand elasticity with respect to chargers and the heterogenous treatment effects of public charger deployments. I document that a 1% increase in charger counts leads to a 0.7% to 1.1% increase in EV sales on average, with a larger increase in upper-middle income suburbs. I also use the difference-in-differences strategy to estimate the differential treatment effect of fast versus regular chargers. While PHEVs are incompatible with fast chargers, early deployments of fast chargers significantly boost BEV sales.**

## I. Introduction

The transportation sector is a major contributor to climate change, accounting for 20% of the global carbon emissions and 30% of emissions in the United States. In 2019, transportation generated 8.5 gigatons of CO2 worldwide, 41% of which came from private passenger cars[1]. In the US, approximately 58% of transportation emissions were from private light-duty vehicles in 2020[2]. Given the significant share of GHG emissions from the transportation sector, especially private transport, governments have rolled out subsidies to accelerate the adoption of electric vehicles (EVs). The government intervention is justified by the positive environmental externality from EV adoption. From 2015 to 2020, the sales share of EVs jumped from 0.9% to 5.8% in China, 0.8% to 2% in the US, and 1.2% to 10% in Europe[3]. Nevertheless, to limit global warming below 2°C, transportation electrification needs to take place more quickly.

The market share of EVs in the US pales when compared to that in other major economies such as the EU and China. Many early adopters value EVs' environmental features over their technical features. They voluntarily pay the price premium of EVs for environmental goods (Kotchen and Moore, 2007; Langbroek et al., 2016). However, high price and low quality are two major barriers to mass EV adoption (Egbue and Long, 2012; De Rubens, 2019). On average, EVs are 20% more expensive than gas vehicles (GVs). Most EVs have a limited driving range and underperform GVs on uneven terrains such as mountainous areas. Teslas are high-quality EVs, but they are at least 20% more expensive than other EVs. Given this tradeoff, it is essential to build public chargers as a complement to EVs. As documented in Neubauer and Wood (2014), access to public chargers can ease middle-income consumers' range anxieties and incentivize them to substitute GVs with affordable EVs despite the limited driving ranges.

To increase the price competitiveness of EVs, states have been providing direct rebates to EV purchases. These rebates feature diminishing marginal returns (Holland et al., 2019). Previous works have found that rebates on EV purchasing reduce little emissions at high costs and may cause an increase in net emissions in regions with dirty electric grids (Holtsmark and Skonhoft, 2014; Holland et al., 2016). This suggests welfare gains in reallocating spendings on EV purchase subsidies to other incentives programs including charger installation rebates.

Previous research has documented indirect network effects in the EV market (Li et al., 2017). The network effects kick in through the installation of new chargers following the rising EV sales. As it becomes easier to find chargers, the time cost associated with driving EVs declines. The increasing charger availability would thus in turn accelerate EV adoption. More recent research has studied the effect of charger deployment over time in different stages of EV adoption. Van Dijk et al. (2022) document that the provision of charging network boosts EV sales in early adoption. Springel (2021) has found medium to long-run positive effects of charger deployment on EV adoption, and the marginal returns of public charger investments decline slower than that of other incentives including purchasing rebates. Most of these past papers use data at the state or metropolitan area level. My paper contributes to this literature by studying the local dynamics of EV adoption using ZIP-code level data from California. Following the approach in Li et al. (2017), I estimate the elasticity of EV sales with respect to public chargers using a shift-share instrumental variable (Bartik, 1991). I document a significantly positive effect of public charger deployment on EV adoption, and the sales elasticity is higher in (upper) middle-income ZIP codes with more single-family homes.

A recent development in the EV industry is the invention of DC fast chargers that increase the charging speed by more than three times. Past literature has not thoroughly studied the differential effect of regular and fast chargers.

Acknowledgments: I thank Professor Matthew E. Kahn for his advice on this paper.

[1] https://www.iea.org/reports/tracking-transport-2021

[2] https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions

[3] https://www.iea.org/reports/electric-vehicles

In this paper, I employ a difference-in-differences (DID) design to document a positive effect of fast chargers on BEV adoption. I use PHEVs as the control group because most PHEVs are not compatible with fast chargers. Their sales should not respond to the deployment of fast chargers. My paper contributes to a growing literature investigating whether quality improvements in EVs and related products incentivize non-environmentalists to adopt EVs (Egbue and Long, 2012; Delmas et al., 2014; De Rubens, 2019). This hypothesis of "accidental environmentalists" implies there are higher returns to the installation of fast chargers than regular chargers.

Environmental economics literature has studied the heterogenous treatment effects of non-binding "green nudges." For example, liberal households with high electricity consumption are more likely to be nudged by electricity usage reports to cut their energy use (Allcott 2011; Costa and Kahn, 2013). Water conservation messages are more effective when targeting wealthier households (Ferraro and Miranda, 2014). Public chargers are a form of "green nudges" because they do not incur any financial burden on potential EV purchasers but could ease their concerns about EVs' limited driving ranges. I study whether the public charger installation exhibit heterogenous effects as do energy conservation nudges. Related EV literature has found that low-income population are more responsive to financial incentives such as EV purchase subsidies (Xing et al., 2021).

This paper is organized as follows: In Section 2, I present a microeconomic framework of how charger deployment incentivizes EV adoption. In Sections 3 and 4, I introduce the dataset and study the cross-sectional distribution of EVs and chargers. In Sections 5 and 6, I estimate the effects of public chargers on EV adoption and the differential impacts of regular versus fast chargers. Then I conclude and point to future areas of study.

## II. Conceptual Framework

In this subsection, I introduce a basic framework of how public charger investments affect EV demand. In this simplified framework, I consider a duopoly market, where one manufacturer exclusively produces EVs and the other exclusively produces GVs. Both manufacturers seek to maximize their profits. The government maximizes social welfare and subsidizes both EV purchases and public charger installations. I follow the approach in Shao et al. (2017) and Kumar et al. (2021) to set up my model. This model is an application of Mussa and Rosen's (1978) model, where consumers with different preferences for specific product qualities coexist, to the vehicle market featuring environmentalists and non-environmentalists.

Consumers can choose to purchase an EV, purchase a GV, or stay inactive. The utility from each of these scenarios is respectively given as:

$$U_e = (1+\delta)\theta - (p_e - S) + \kappa I \tag{1a}$$
$$U_g = \theta - p_g \tag{1b}$$
$$U_n = 0 \tag{1c}$$

In the above equations, $U_e$, $U_g$, and $U_n$ are the utilities from purchasing EVs, GVs, and staying inactive respectively. is consumers' valuation of the car, depending on the individual tastes on size, color, etc. is a measure of consumers' environmental awareness. For simplicity, I assume follows a uniform distribution on $[0, 1]$, and is in the interval $[0, 1]$. The interaction term between $\theta$ and $\delta$ is positive for environmentalists, showing that they derive more utility from driving EVs than non-environmentalists do. $p_e$ and $p_g$ are the prices of EVs and GVs, and $S$ is the purchase subsidy per EV. $I$ is the total investment in public chargers, and $\kappa$ is a measure of consumers' utility change with respect to the charger investment. $\kappa$ is larger for high-income consumers whose marginal time cost of driving EVs is high and consumers without access to residential charging. I assume it can take any value from $[0, 1]$.

I solve the indifference points by equating $U_e$ to $U_g$ and $U_g$ to $U_n$ respectively. Consumers are indifferent between EVs and GVs when $\theta = \frac{p_e - p_g - S - \kappa I}{\delta}$, denoted as $\theta_1$, and they are indifferent between GVs and staying inactive when $\theta = p_g$, denoted as $\theta_2$. The demand for EVs is $1 - \theta_1$, and the demand for GVs is $\theta_1 - \theta_2$. I substitute in $\theta_1$ and $\theta_2$ and rearrange to get the inverse demand functions:

$$p_e = 1 + \delta + S + \kappa I - (1+\delta)q_e - q_g \tag{2a}$$
$$p_g = 1 - q_e - q_g \tag{2b}$$

Without loss of generality, I assume the production cost per GV is zero and that per EV is C $(C > 0)$ because EV batteries are expensive. The profit functions of the EV and GV manufacturers can be expressed as $\pi_e = (p_e - C)q_e$ and $\pi_g = p_g q_g$. These profit functions are concave because $\frac{\delta^2 \pi_e}{\delta q_e^2} = -2\delta - 2 < 0$ and $\frac{\delta^2 \pi_e}{\delta q_e^2} = -2 < 0$. The manufacturers supply the quantity that maximizes their profits. To derive the optimal sales of EVs and GVs, I differentiate $\pi_e$ and $\pi_g$ with respect to $q_e$ and $q_g$ and set the derivatives to 0:

$$1 + \delta + S + \kappa I - (2+2\delta)q_e - q_g - C = 0 \tag{3a}$$
$$1 - q_e - 2q_g = 0 \tag{3b}$$

When (3a) and (3b) are solved, the EV sales and GV sales are given as:

$$q_e = \frac{1 + 2\delta + 2S + 2\kappa I - 2C}{4\delta + 3} \tag{4a}$$
$$q_g = \frac{1 + \delta + C - S - \kappa I}{4\delta + 3} \tag{4b}$$

Equations (4a) show the mechanism of EV adoption. EV sales are higher in areas with higher environmental awareness ($\delta$), more purchase subsidies (S), and larger investments in chargers (I). The sales are also a monotonically increasing function of $\kappa$, which is higher for potential purchasers who benefit more from public charger deployment. The EV demand of this subset of consumers is more elastic, so charger deployment could yield larger returns when targeted to them. is the key variable of interest in this paper because it measures the responsiveness of EV sales to charger investments. I will study the heterogenous treatment effects of public chargers (i.e. variations of $\kappa$ across space).

Equation (4b) shows that EVs and GVs are substitutes. More people are incentivized to substitute away from GVs when EV purchase rebates are higher (S) and public chargers become more accessible (I). However, the higher cost of producing EVs (C) causes higher EV prices. This induces negative income effects and thus hinders EV adoption.

## III. Data

To analyze the effects of charger deployment on EV adoption, I compile a comprehensive dataset at ZIP/EV type/year level with 1800 Californian ZIP codes from 2010 to 2021. In most of my regression analysis, I focus on the 1200 ZIP codes that have installed at least one public charger by 2021. The EV sales data is from the California Energy Commission.[4]. This dataset provides the sales of BEV and PHEV respectively in each ZIP code each year. The charger data is from the Alternative Fuels Data Center (AFDC)[5]. This dataset offers the information on every public EV charging station in the US, including its location, number of regular and fast chargers, and open year. Based on this information, I calculate the number of regular and fast chargers by ZIP/year and merge it into the EV sales data.

I calculate the net vehicle price by EV type (BEV or PHEV) for every year from 2010 to 2021. For each EV brand/type, I calculate the average manufacturer's suggested retail price (MSRP) of all its available vehicle models in each given year. I then average the brand prices by EV type/year[6]. I merge this dataset by EV type/year into the main dataset. To calculate the net price at ZIP/EV type/year level, I use the EV rebate dataset from the California Clean Vehicle Rebate Project[7]. This dataset includes data on the location, EV brand, type, rebate, and time of each EV purchase that has applied for rebates since 2010. This enables me to calculate the purchase rebates per BEV and per PHEV respectively in each ZIP/year. I merge it into the sales and chargers dataset, and the net price by ZIP/type/year equals the average price minus the rebates.

Another essential part of my EV dataset consists of demographic data. The household income, education, population density, and single-family home data are from the American Census Survey (ACS)[8]. The original data are available at census tract level. I map each census tract to a ZIP area and take the population-weighted average within each ZIP code. The income data is available each year, while the other three variables are cross-sectional from 2020. I merge in the income data to the main dataset by ZIP/year and the cross-sectional data by ZIP. In the following sections, I use this ZIP/EV type level panel dataset to test multiple hypotheses.

## IV. EV Adoption and Charger Deployment in California

### A. General time trends

In the past decade, California has launched various state initiatives to incentivize the substitution of GVs with EVs[9]. In 2021, an average Californian BEV purchaser receives \$2500 point-of-sale rebates from the California Air Resources Board (CARB), and a PHEV purchaser receives \$1600. Aside from state incentives, electric utility companies such as Southern California Edison offer multiple EV rebate projects to expedite the EV adoption[10].

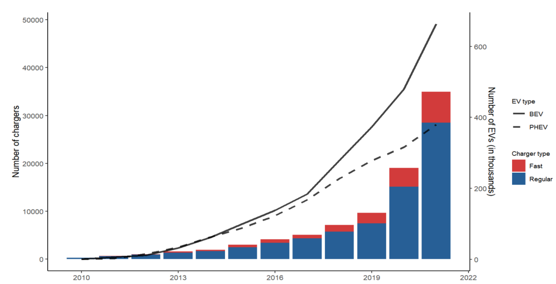



Fig. 1

Figure 1 shows the number of private light-duty EVs in California over time. In 2010, there were only 759 EVs (593 BEVs and 166 PHEVs), and this number rose to 187,769 in 2015 (99,883 BEVs and 87,886 PHEVs), and to 1,042,138 in 2021 (663,013 BEVs and 379,125 PHEVs). The derivative of EV counts, especially BEVs, with respect to time is increasing, showing that EV sales are rising. As of 2021, there were almost twice as many BEVs as PHEVs. The adoption trend of BEVs and PHEVs was similar before 2017 but began to diverge sharply afterward. Although PHEVs have lower upfront costs, they are not fully electric and have smaller batteries supporting shorter driving ranges. The rapid increase in BEV counts in recent years indicates that vehicle quality is an important consideration for new EV adopters.

Along with the rising EV sales is the deployment of public chargers, which are essential to ease EV drivers' range anxiety. The charger rebate policy varies across counties. A regular charger is eligible to a rebate from \$3500 to \$6000 and a fast charger from \$45,000 to \$70,000. Counties in Southern California and Bay Area generally offer higher rebates. Charger installers can also apply for rebates from the electric utility companies.

Figure 1 shows that the number of public chargers has grown exponentially in the past few years. From 2010 to

[4]https://www.energy.ca.gov/files/zev-and-infrastructure-stats-data

[5]https://afdc.energy.gov/

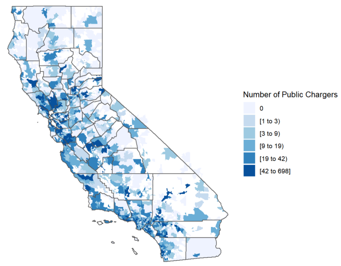
[6]For example, Tesla produced three models in 2018: Model S, Model X, and Model 3. All are BEVs. The average price per Tesla (BEV) in 2018 is calculated by taking the average MSRP of these three models in 2018. To calculate the average price per BEV in 2018, I take the average BEV prices across all brands including Tesla, Nissan, etc.

[7]https://cleanvehiclerebate.org/en/rebate-map

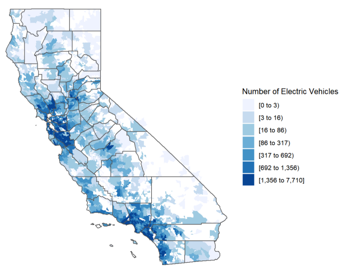
[8]https://data.census.gov/cedsci/

[9]https://afdc.energy.gov/laws/all?state=CA

[10]https://www.sce.com/residential/ev-rates-rebates




(a) EVs




(b) Chargers

Fig. 2

2015, the total count of regular chargers grew from 305 to 2473, and that of fast chargers increased from 3 to 551. The count of both public regular and fast chargers rose by more than ten times from 2015 to 2021. In 2021, there were 28,495 regular chargers and 6473 fast chargers in California. Most public chargers are regular chargers, but the share of fast chargers rose over time. Figure 1 suggests that BEVs and fast chargers are complements because the jump in the BEV count coincides with the jump of the fast charger count since 2017. This is because most existing PHEV models are not compatible with fast chargers.

### B. Mapping the spatial variations

Figures 2 shows the spatial distribution of EVs and chargers across ZIP codes in 2021. EVs and public chargers mostly concentrate in Southern California (55%) and the Bay Area (36%). The distribution of new EV sales is similar. Californians purchased 248,470 new EVs in 2021, 59% were in Southern California and 31% in the Bay Area. Over one-fourth of EVs and new EV sales in California are in Los Angeles County. EV ownership is the lowest in Superior California counties (counties to the north of the Bay Area). Most counties in the region still have fewer than 100 EVs today. Regional geographical features could play a role in people's decision to purchase EVs. Superior California is mountainous and sparsely populated, so drivers may prefer gas-powered vehicles because of their longer driving ranges and stronger car engines. These counties are also generally more conservative.

The distribution of public chargers aligns with that of EVs. In 2021, 47% of public chargers are in Southern California and 36% in the Bay Area. Los Angeles County has almost 9000 public chargers, about 23% of the total count in California. The demand for charging is high given the large EV count in the county. This incentivizes the expansion of the public charging network. Access to public chargers reduces the marginal time cost of driving EVs and thus in turn accelerates the EV adoption. This demand-side economies of scale imply the presence of network effects in the market (Katz and Shapiro, 1994). The positive network effects explain the collocation of EVs and chargers. Also because of this positive feedback loop, areas already with more EVs and chargers tend to adopt EVs and deploy chargers at faster rates, leading to over-dispersed distributions of EVs and chargers. Both EVs and chargers are highly right-skewed, with a skewness of 3.68 and 6.87 respectively.

### C. Regression results on spatial variations

Local demographic attributes are important in determining the pace of EV adoption and charger deployment. EVs are more expensive, and the emergence of high-quality EVs has incentivized high-income people to substitute GVs with EVs (Delmas et al, 2014). The voluntary constraint hypothesis posits that environmentalists would cut their carbon footprints in the absence of a Pigouvian tax (Kotchen and Moore, 2007). Educated people with higher environmental awareness are thus more likely to purchase EVs (Kahn, 2007; Okada et al., 2019). Given the higher demand, businesses would install more chargers in these areas. However, in ZIP codes that have many single-family homes with residential charging, businesses install fewer public chargers even if the EV count is high. The expected profits are lower because residential charging is a substitute to public chargers. In this section, I use a negative binomial model to study how EVs and public chargers distribute across California. The model is chosen based on the facts that the distribution of EVs and chargers are over-dispersed and most ZIP codes have at least one EV and one public charger. I use cross-sectional data from 2021, and the unit of analysis is a ZIP code. The density function can be written as:

$$f(y_i|x_i) = \frac{\tau\theta + y_i}{\tau(\theta)\tau(1+y_i)}(\frac{\mu_i}{\theta+\mu_i})^{y_{it}}(\frac{\theta}{\theta+\mu_i})^{\theta}, \quad (5a)$$

where y is the count of EVs or public chargers, and $\theta$ is the dispersion parameter and denotes the gamma function. $\mu_i$ is the conditional mean defined as:

$$\mu_i = exp(\beta' x_i) \quad (5b)$$

with x a vector of local attributes. I include county-fixed effects and cluster standard error by county. The results are shown in Table 1. In columns (1) to (4), the dispersion parameter is significantly positive at 5% level, confirming the appropriateness of the negative binomial model.

**Table 1. Cross-sectional Variations in EV and Public Charger Counts**

| | (1) BEV | (2) PHEV | (3) L2 charger | (4) Fast charger | (5) Fast dummy |
|---|---|---|---|---|---|
| log(Income) | 1.805*** | 1.672*** | 0.466* | 0.652 | 1.108** |
| | (0.380) | (0.317) | (0.273) | (0.503) | (0.507) |
| Education | 0.818 | 0.187 | 0.592 | -0.445 | -1.072 |
| | (0.526) | (0.527) | (0.471) | (0.967) | (0.937) |
| log(Density) | 0.325*** | 0.317*** | 0.0203 | 0.0224 | 0.0314 |
| | (0.0526) | (0.0516) | (0.0408) | (0.0508) | (0.0642) |
| Single family | 0.234 | 0.425* | -2.713*** | -1.500*** | -1.870*** |
| | (0.209) | (0.247) | (0.421) | (0.564) | (0.617) |
| Constant | -17.80*** | -16.67*** | -0.779 | -4.698 | -11.36** |
| | (4.475) | (3.653) | (3.099) | (5.302) | (5.227) |
| County fixed effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 1,077 | 1,077 | 1,077 | 1,077 | 1,051 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Columns (1) and (2) show that there are more BEVs and PHEVs in richer ZIP codes. The expected count of BEVs and PHEVs increase by a factor of 6.08 and 5.32 respectively when income rises by a factor of e (approximately 2.718). BEV sales are more sensitive to income, confirming the "accidental environmentalists" as BEVs have higher quality. The significantly positive coefficient of population density suggests that urban areas have adopted more EVs. Education is insignificant, and single-family home percentage is only weakly significant in column (2).

In columns (3) and (4), single-family home percentage is significantly negative, while most other variables do not significantly affect public charger counts. A conversion into incidence-rate ratio shows that the expected regular and fast charger count declines by 24% and 14% respectively following a 10% rise of the single-family home percentage. The difference in coefficients indicates that residential charging crowds out more investments in regular chargers than in fast chargers. By 2021, 70% of the ZIP codes still have not installed any fast chargers. In column (5), I run a logit regression on the fast charger dummy. A 1% increase in median income raises the odds of installing fast chargers by 1.11%. High-income people have higher marginal costs of time, so they value fast charging more. Consistent with column (4), the single-family home percentage is significantly negative in column (5). Education and population density are insignificant.

## V. Effects of Charger Accessibility on EV Adoption

### A. EV sales model

In the previous section, I have found a positive spatial correlation between EV adoption and charger accessibility. Previous research has used MSA or city-level data to estimate the elasticity of EV demand with respect to public chargers (Li et al., 2017; Springel, 2021; Van Dijk et al., 2022). While policy differences usually explain a large part of variations across MSAs or cities, Figure 2 shows that EVs and chargers could differ significantly even in adjacent ZIP codes. In this section, I estimate the effect of charger installation on EV sales at the ZIP level and study the heterogenous treatment effects. I estimate the following basic regression for EV type e (BEV or PHEV) in ZIP i (located in county k) in year t:

$$\begin{aligned} log(Sales_{iet}) = \beta_0 + \beta_1 log(Chargers_{it}) \\ + \beta_2' X_{iet} + \gamma_t + \sigma_{ke} + \epsilon_{iet} \end{aligned} \tag{6}$$

In equation (6), X is a vector of covariates including the net vehicle price, income, etc. I include year-fixed effects ($\gamma_t$) and county/type effects ($\sigma_{ke}$). Standard errors are clustered by county.

Because of the indirect network effects, the charger count is endogenous. I employ the instrumental variable (IV) strategy. Similar to the approach in Li et al. (2017), my chosen IV is the interaction of two ZIP-code level variables: the percentage of local population living within half a mile from supermarkets or grocery stores[11] and the one-year lagged number of outside the county where the ZIP code is in. The formula of the IV for ZIP code i (in county k) in year t is given by:

$$Z_{it} = Grocery\%_i \times \sum_{j \neq k, j \in J} Chargers_{j,t-1} \tag{7}$$

This is a shift-share instrument (Bartik, 1991; Goldsmith-Pinkham et al., 2020). The lagged charger count captures the state-level trend in charger deployment independent of potential county-level shocks (the "shift" component). The food access variable is an estimation of the percentage of potential buyers who have access to new chargers (the "share" component). In the US, grocery stores like Kroger are major owners of charging ports. They install chargers to attract consumers who drive EVs. In California, stores are also incentivized to invest in chargers to gain green tax credits. Using this interaction term as an IV requires that a shock on charger supply has larger impacts on ZIP areas where people live closer to the chargers. This is a valid assumption because the California Vehicle Survey shows that people are more likely to consider purchasing EVs if they see more chargers near their homes or workplaces[12].

[11]The food access is provided by the US Department of Agriculture (USDA): https://www.ers.usda.gov/data-products/food-access-research-atlas/download-the-data/. I use the data from 2019. The original data is at census tract level. I convert it to ZIP-code level by taking the average percentage across all census tracts within a ZIP-code area.

[12]https://www.energy.ca.gov/data-reports/surveys/california-vehicle-survey

**Table 3. Validity of the Instrumental Variable**

| | (1) | (2) | (3) |
|---|---|---|---|
| | log(Chargers) | log(GV count) | |
| IV | 0.0663** | | 0.0111 |
| | (0.0268) | | (0.0447) |
| log(Chargers) | | 0.861*** | |
| | | (0.0348) | |
| log(Income) | 0.730*** | | |
| | (0.195) | | |
| log(Price) | -0.0941*** | | |
| | (0.0318) | | |
| BEV x Trend | -0.000786 | | |
| | (0.00103) | | |
| Education | -0.217 | | |
| | (0.444) | | |
| log(Density) | -0.0664*** | | |
| | (0.0230) | | |
| Single family | -1.600*** | | |
| | (0.192) | | |
| Constant | -5.770*** | 8.530*** | 8.971*** |
| | (2.077) | (0.0291) | (0.153) |
| Year fixed effects | Yes | Yes | Yes |
| County fixed effects | No | Yes | Yes |
| County/model fixed effects | Yes | No | No |
| Observations | 20,810 | 25,467 | 22,769 |
| R-squared | 0.467 | 0.224 | 0.082 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

I argue that this IV passes the exogeneity test. Both the state-level trend in charger deployments and supermarket access are positively related to the local count of public chargers. Column (1) of Table 3 shows the IV is positively significant in the first-stage regression, satisfying the monotonicity assumption from the local average treatment effect (LATE) framework (Angrist et al., 1996). The F-statistic from the first stage is 357, much higher than the weak instrument threshold of F=10. The IV satisfies the independence assumption because the local EV sales and public charger counts affect neither people's decision to open grocery stores nor the lagged charger count. The exclusion restriction assumption should also be satisfied. I have added in county/EV type fixed effects to control for the possible common unobservable variables across counties. Conditional on them, grocery store access and the lagged out-of-county charger count are unlikely to directly affect EV sales. Columns (2) and (3) from Table 3 suggest the IV is uncorrelated with the error term. GV counts and EV chargers are both correlated with unobservable shocks in the transit sector (e.g. failing public transit that incentivizes more people to drive, either GVs or EVs), so column (2) reports the spurious result that EV chargers cause an increase in GV counts. Yet, under the same specification, the IV is insignificant in column (3) with a p-value of 0.805. The highly insignificant result implies the validity of the IV.

## B. *Main estimation results*

**Table 2. Determinants of EV Sales**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | OLS | IV | OLS | IV | IV | IV |
| | | | log(Sales) | | | |
| log(Chargers) | 0.435*** | 1.091*** | 0.125*** | 0.725*** | 1.133*** | |
| | (0.0285) | (0.189) | (0.0210) | (0.241) | (0.216) | |
| log(Accessible chargers) | | | | | | 1.008*** |
| | | | | | | (0.253) |
| log(Income) | 1.167*** | 1.014*** | 1.363*** | 0.986*** | 1.045*** | 1.189*** |
| | (0.0542) | (0.0741) | (0.234) | (0.301) | (0.0693) | (0.0976) |
| log(Price) | -2.931*** | -2.423*** | -2.038*** | -2.019*** | -2.425*** | -2.155*** |
| | (0.151) | (0.268) | (0.128) | (0.133) | (0.270) | (0.127) |
| BEV x Trend | 0.0692*** | 0.0773*** | 0.103*** | 0.103*** | 0.0759*** | 0.100*** |
| | (0.00784) | (0.00635) | (0.00594) | (0.00607) | (0.00684) | (0.00597) |
| Education | | | 0.663 | 0.747* | | |
| | | | (0.458) | (0.433) | | |
| log(Density) | | | 0.224*** | 0.239*** | | |
| | | | (0.0246) | (0.0231) | | |
| Single family | | | 0.294 | 1.301*** | | |
| | | | (0.202) | (0.486) | | |
| Constant | 17.52*** | 14.72*** | 4.001 | 7.995** | 14.46*** | 10.22*** |
| | (1.886) | (2.681) | (3.536) | (3.723) | (2.798) | (1.194) |
| County/model fixed effects | Yes | Yes | Yes | Yes | No | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes | No | Yes |
| County/year fixed effects | No | No | No | No | Yes | No |
| Model fixed effects | No | No | No | No | Yes | No |
| Observations | 29,248 | 28,750 | 21,074 | 20,810 | 28,750 | 21,077 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2 presents the estimated results of different specifications of equation (6). Columns (1) and (2) show the basic models, and local demographic attributes are added in columns (3) and (4). In these columns, logged chargers have a significantly positive coefficient. Given the log-on-log specification, the coefficient can be interpreted as elasticity. This corresponds to in equation (1a). From the OLS estimation, a 1% increase in accessible chargers leads to a 0.44% and 0.13% (see columns (1) and (3)) increase in EV sales, and from the IV estimation, 1.09% and 0.73% respectively (columns (2) and (4)). These values are slightly higher than the estimates from Li et al. (2017), which is based on MSA-level data across the US. California is more liberal than the rest of the nation, so more people are at the margin of substituting their fuel-powered cars with EVs. The downward bias of OLS implies that the unobserved shocks to EV sales are negatively correlated with the charger count. An example is electric utilities' incentive programs for residential charging. The residential charging programs will increase EV sales but slow down the deployment of public chargers.

The coefficient of the net price of EVs is significantly negative in all columns. The estimated price elasticity is -2.02 to -2.93. Berry et al. (1995) have found that the price elasticities of automobiles range from -3 to -10, with an average of -7.2. The EV price elasticity is lower because many people purchase them out of environmental concerns and are thus less price sensitive (Kahn, 2007; Langbroek et al., 2016). As more environmentalists have made the substitution, the price elasticity may start to rise. A limitation of my specifications is that they do not account for the differential quality between EV brands but only EV types (BEV and PHEV). This could cause price elasticities to bias downward (Berry et al., 1995).

The estimated income elasticity is roughly 1 to 1.2. It is significantly greater than 0 at the 1% level. This is consistent with the spatial distribution of EVs because EV counts are higher in wealthier neighborhoods. In column (4), I find that more educated ZIP codes with higher population density and more single-family homes adopt more EVs.

If local EV sales are affected by chargers in other ZIP codes, the Stable Unit Treatment Value Assumption (SUTVA) of the potential outcome framework would be violated. My baseline specifications cluster standard errors by county to address this spatial correlation within each county. In columns (5) and (6), I further demonstrate my results are robust to spatial correlations. In column (5), I include county/year fixed effects. This captures the county-level shock in a given year. For example, if a ZIP code in downtown installs many chargers, this could raise EV sales in all other ZIP codes because people work in downtown and can charge during work. The coefficient of logged chargers is still significantly positive and numerically similar as in column (2).

In column (6), I address the concern that EV sales may be affected by public chargers in neighboring ZIP codes. Instead of using the charger count within a ZIP code, I calculate the accessible charger count using inverse-distance weighting:

$$Accessiblechargers_i = Chargers_i + \sum_{j \in J} \frac{Chargers_j}{d_{ij}^2} \quad (8)$$

where J is a set of ZIP codes within 5 miles from ZIP code i, and d is the distance between two ZIP codes in miles. When accessible chargers is used, the estimated elasticity is still significantly positive. It is roughly 10% smaller than the value in column (2), suggesting that 10% of the EV sales increase is due to charger installation in nearby ZIP codes. This numerically small value indicates that most people would not substitute to EVs unless chargers are available very close to where they live or work.

### C. *Heterogenous treatment effects across ZIP codes*

**Table 4. Heterogenous Treatment Effects of Public Chargers**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | | log(Sales) | | |
| log(Chargers) | 1.180*** | 0.990*** | 1.748*** | 0.597* |
| | (0.240) | (0.227) | (0.485) | (0.362) |
| log(Income) | 1.112*** | 1.098*** | 1.022*** | 1.012*** |
| | (0.0832) | (0.140) | (0.344) | (0.382) |
| log(Price) | -3.072*** | -3.008*** | -2.422*** | -2.436*** |
| | (0.234) | (0.224) | (0.140) | (0.133) |
| BEV x Trend | 0.0623*** | 0.0816*** | 0.105*** | 0.105*** |
| | (0.0149) | (0.00710) | (0.00661) | (0.00656) |
| Education | | | 0.729 | 0.774 |
| | | | (0.538) | (0.522) |
| log(Density) | | | 0.381*** | 0.266*** |
| | | | (0.0770) | (0.0278) |
| Single family | | | 1.653*** | 0.832 |
| | | | (0.483) | (0.818) |
| BEV x log(Chargers) | 0.0859 | | | |
| | (0.0925) | | | |
| High-income x log(Chargers) | | 0.241 | | |
| | | (0.176) | | |
| Mid-income x log(Chargers) | | 0.330*** | | |
| | | (0.0956) | | |
| log(Density) x log(Chargers) | | | -0.100 | |
| | | | (0.0677) | |
| Single family x log(Chargers) | | | | 0.626* |
| | | | | (0.361) |
| Constant | 19.99*** | 19.47*** | 10.06** | 11.69*** |
| | (2.576) | (2.674) | (4.528) | (4.107) |
| County/model fixed effects | Yes | Yes | Yes | Yes |
| Year fixed effects | Yes | Yes | Yes | Yes |
| Observations | 28,115 | 27,053 | 20,587 | 20,587 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Previous literature has documented lower-income households are more responsive to EV purchase subsidies (Xing et al., 2021). The effectiveness of charger deployment is also likely to vary across population and geography. I hypothesize the elasticity of EV sales with respect to chargers is higher in higher-income and more urban ZIP codes. The price premium of EVs is less of a concern for richer people who care more about the quality of EVs and the convenience of driving EVs. Public chargers make it more convenient to charge and would incentivize them to substitute to EVs. Urban areas are more compact and less mountainous than rural areas, so they are more suitable for EVs given the existing technical limitations. I test these and some other hypotheses by including the interaction terms between demographic variables and the charger count in equation (6). The demographic variables are assumed to be exogenous, so I interact the demographic variable with the shift-share IV specified in equation (7) to construct a new IV for each interaction term. The estimation results are reported in Table 4.

In column (1), I test whether the elasticity differs across BEV and PHEV buyers and find no significant result. Although fast chargers are only compatible with BEVs, more than 80% of public chargers are regular chargers. This explains why the treatment effect of an average charger does not differ by EV type. The result is consistent with previous findings that EV demand elasticity with respect to purchase subsidies do not vary across EV type (Clinton and Steinberg, 2019).

In column (2), I study the heterogeneities across income groups. The high-income dummy equals 1 for ZIP codes above the 75th percentile of median income, and the mid-income dummy equals 1 for ZIP codes between the 25th and the 75th income percentile. I find that the elasticity in mid-income areas is 33% higher than that in low-income ZIP codes, whereas the elasticity in high-income and low-income areas do not differ significantly. This suggests that the returns to charger investment is a concave function of income. The high price of EVs is the biggest barrier to adoption for low-income households. In rich ZIP codes, households concern more about vehicle quality (Egbue and Long, 2012). With a fixed budget, they would prefer GVs with better engines such as Cadillac over Tesla. Access to public charging offers them smaller incentives for substitution.

In columns (3) and (4), I test whether the elasticity differs by neighborhood features. I find no evidence that people in high-density urban areas are more responsive to the deployment of public chargers, but the elasticity is higher in neighborhoods with more single-family housing. The pattern is significant when the median income is controlled, so it cannot solely be explained by spatial sorting (i.e. richer people drive more and live in the suburb with more single-family homes). A 10% increase in the percentage of single-family homes causes the elasticity to rise by 10.4%. This is significant at the 10% level. While drivers can install home charging in single-family houses, home charging features a high fixed cost, whereas public charging has zero upfront cost. The per-kWh saving from home charging compared to public charging is low, so the average cost of home charging drops below that of using public charging only in the long run. Single-family house owners are thus cautious in investing in residential charging.

## VI. Early Deployment of Fast Chargers

### A. Comparing trends of BEVs versus PHEVs

Figure 1 shows that the count of BEVs and PHEVs began to diverge in 2017, which coincides with the early deployment of fast chargers. This motivates the use of DID to estimate the treatment effect of early fast charger deployment on annual EV sales. The feasibility of this approach is due to the fact that PHEVs are generally not compatible with fast chargers, so they constitute an ideal control group for BEVs. This strategy is also boosted by results from Table 4 implying the elasticities of BEV sales and PHEV sales with respect to public chargers (mostly regular chargers as of now) do not differ significantly. This rules out the regular charger deployment as a confounding factor.

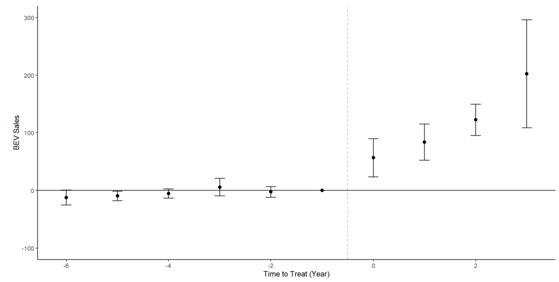
I focus on the 68 ZIP codes that initially deployed fast chargers in 2018 or 2019 (32 from 2018 and 36 from 2019) and have built at least 11 fast chargers (the 90th percentile among all ZIP codes with at least one fast charger) by 2021. These ZIP codes are located in 23 different counties and are at least five miles from each other, so the spatial spillovers are likely limited. I estimate the following event-study specification for ZIP code i, EV type e in year t:

$$\begin{aligned} Sales_{iet} = \sum_{\tau=-6}^{3} \beta_\tau Deployment_{i\tau} * BEV_e + BEV_e \\ + \gamma_t + \varphi_k + \epsilon_{iet} \end{aligned} \tag{9}$$

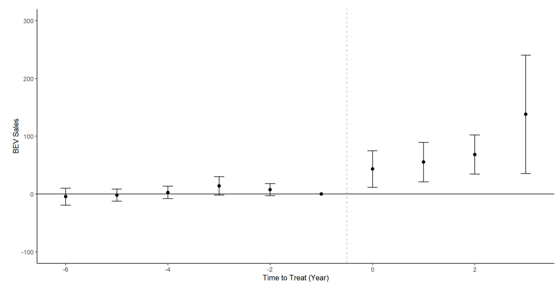
Deployment is a dummy that equals 1 if it is $\tau$ years from the deployment of the first fast charger in ZIP code i.$\tau$ is set to -6 for periods more than 6 years before the treatment. The omitted time category is $\tau = -1$. BEV is equivalent to the treatment group dummy. $\beta$ is a vector of DID estimates. I include year-fixed effects ($\gamma_t$) and county-fixed effects ($\varphi_k$). The standard error is clustered by county. The 95% confidence intervals of are plotted in panel (a) of Figure 3.

The DID coefficient is insignificant at the 5% level except when $\tau = -5$. It is significant at the 10% level when $\tau = -6$ and insignificant in other pre-treated periods. However, these could be explained by the limited BEV supply in the early years of the 2010s. For example, Tesla's first popular BEV model, Model S, was released in late 2012. Other major BEV brands such as Nissan released their first BEV models even later. Since various BEV models became available, BEV and PHEV sales have followed a parallel trend until the treatment period (see the insignificant DID coefficients since $\tau = -4$). The coefficients are also numerically small, none of which is larger than 10 in magnitude.

Following the opening of the first fast charging stations, BEV sales significantly increased, with highly significant positive coefficients in all post-treatment periods. In the year of treatment, BEV sales rise by 57 per ZIP code, and the magnitude of increase continues to rise to 202 per ZIP code




(a) DID




(b) DDD
Fig. 3

three years later. Fast charger installation is not a one-time event. It is likely that ZIP codes add in more fast chargers after the initial deployment, leading to large divergences in sales between BEVs and PHEVs.

In recent years, breakthroughs have been made in improving the quality of BEVs such as the invention of batteries that support longer driving ranges. Such improvements may coincide with the deployment of fast chargers and would cause BEV and PHEV sales to diverge regardless of fast chargers deployments. To rule out these potential explanatory factors, I employ a triple-difference (DDD) model to test the robustness of the previous result. The DDD estimator is the difference between two DID estimators: the DID estimator based on the 68 ZIP codes used to estimate equation (9) and the DID estimator based on another 484 untreated ZIP codes that satisfy (1) no fast charger prior to 2019; (2) fewer than 5 fast chargers by 2021; and (3) at least three miles away from any treated ZIP code. The difference in the DID estimators should take out the cross-ZIP codes unobservable trends that have caused BEV and PHEV sales to diverge. I estimate the following equation for ZIP code i, EV type e in year t:

$$\begin{aligned} Sales_{iet} = \alpha_1 BEV_e + \alpha_2 Fast_i + \alpha_3 BEV_e \cdot Fast_i \\ + \omega_1' BEV_e \cdot Year_t + \omega_2' Fast_i \cdot Year_t \\ + \sum_{\tau=-6}^{3} \beta_\tau Deployment_{i\tau} \cdot Fast_i \cdot BEV_e + \gamma_t \\ + \phi_k + \epsilon_{iet} \end{aligned} \tag{10}$$

In equation (10), fast is a dummy equal to 1 if the ZIP code is one of the 68 ZIP codes that initially installed fast chargers in 2018 or 2019. Year is a vector of year dummies. Their interaction terms with the BEV dummy capture factors that accelerate the BEV adoption aside from fast chargers (e.g. quality improvements or rising purchase rebates for BEVs). Other variables are the same as in equation (9). $\omega_1'$ is a vector of DID estimates by year for the control ZIP codes where fewer than 5 fast chargers have been built as of 2021. This captures the factors affecting the BEV adoption trend aside from the initial deployment of fast chargers. $\beta$ is a vector of DDD estimates equal to the difference between the DID estimate from the treatment ZIP codes and the control ones. This should equal to the average treatment effect of fast chargers installation. Standard errors are clustered by county. The 95% confidence intervals of $\beta$ are shown in panel (b) of Figure 3.

Before the installation of fast chargers, none of the DDD coefficients is statistically significant at the 5% level, not even at the 10% except when $\tau = -3$. The BEV sales surge significantly after the installation. The annual sales increase by 43 per ZIP code in the treatment year and by 143 three years after the initial deployment. All coefficients are significant at the 1% level, but they are numerically smaller than the DID estimates. This indicates that exogenous factors such as BEV quality improvements cause the DID estimates to bias upward. Their effects are mitigated, if not cancelled out, when the DDD estimator is used.

**Table 5. Effects of Fast Charger Deployment on BEV Sales**

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | DID (treated ZIPs) | | | DDD (full sample) | | |
| | Sales | | | | | |
| BEV x post | 105.9*** | 103.1*** | 104.6*** | | | |
| | (19.72) | (20.61) | (20.16) | | | |
| BEV x post x fast | | | | 63.22*** | 58.41*** | 58.77*** |
| | | | | (20.41) | (19.73) | (19.29) |
| Covariates | No | No | Yes | No | No | Yes |
| County fixed effects | Yes | No | No | Yes | No | No |
| Year fixed effects | Yes | No | No | Yes | No | No |
| County/year fixed effects | No | Yes | Yes | No | Yes | Yes |
| Observations | 1,632 | 1,632 | 1,632 | 11,616 | 11,616 | 11,616 |
| R-squared | 0.473 | 0.540 | 0.622 | 0.439 | 0.495 | 0.585 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To estimate the average treatment effect, I simplify equations (9) and (10) to include a single post-treatment dummy instead of a separate deployment dummy for each period. The simplified equations are shown below.

$$\begin{aligned} Sales_{iet} =& \beta BEV_e \cdot Post_{it} \\ =& + \omega' X_{it} + BEV_e + fixedeffects + \epsilon_{iet} \end{aligned} \tag{11a}$$

$$\begin{aligned} Sales_{iet} =& \alpha_1 BEV_e + \alpha_2 Fast_i + \alpha_3 BEV_e \cdot Fast_i \\ & + \omega_1' BEV_e \cdot Year_t + \omega_2' Fast_i \cdot Year_t \\ & + \beta Post_{it} \cdot Fast_i \cdot BEV_e \\ & + \omega_3' X_{it} + fixedeffects + \epsilon_{iet} \end{aligned} \tag{11b}$$

The post dummy equals 1 if ZIP code i has installed fast chargers in year t. X is a vector of demographic variables as appear before. In equation (11b), I include the interaction terms between year dummies and the BEV dummy and the fast charger deployment dummy (i.e. the treatment dummy) receptively. The former controls for shocks particular to BEV sales in year t such as the introduction of a new Tesla model. The latter controls for shocks to EV sales particular to the treated ZIP codes such as the roll out of new EV purchase subsidies in these areas. Failing to account for such positive shocks at the BEV/year or treatment/year level would bias (the estimated treatment effect) upward. I include different fixed-effects such as county-fixed effects, year-fixed effects, and county/year-fixed effects. The estimation results are reported in Table 5.

From columns (1) to (3), the DID estimator is significantly positive and has a numerical value around 105. This represents a large increase as the average BEV sales in the treated ZIP codes is 87 one year prior to the treatment. With county/year fixed effects, I control for county-level factors such as the opening of fast charging stations in downtown, which could affect the EV adoption in all parts of the county. The DID coefficients are smaller in columns (2) and (3), but the difference is small in magnitude. In columns (4) to (6), the DDD estimator is around 60 and highly significant. The BEV sales increase is smaller but still sizable under this specification. Again, the coefficient is only slightly smaller in magnitude when county/year fixed effects are added. Assuming that the fast charger deployment in downtown is the primary factor captured by county/year fixed effects, results from Table 5 indicate that EV demand is more responsive chargers near where people live. Although people may drive to downtown regularly (e.g. for work), the driving distance is relatively short, so they can make a round trip without charging. Fast chargers in downtown are thus less effective in incentivizing BEV purchases when people expect not to use them a lot. Future research can study whether this holds for all cities, especially multicentric cities such as Los Angeles.

### B. Placebo tests

One key assumption of the DID strategy is the "no anticipation" effect (Borusyak et al., 2022). If people expect fast chargers to be built soon and purchase BEVs before they are actually built, this could bias the estimators. I conduct a temporal placebo test by moving the installation time one year earlier and replot Figure 3. These results are shown in panel (a) of Figure 4.

Results show that the mean placebo difference remains insignificant and numerically small in magnitude in period 0, the placebo treatment period. Starting from period 1, the true treatment period, all estimates are highly significant and numerically close to the estimates from the corresponding estimates in the original specifications. These placebo results provide further confidence to the positive effect of fast chargers that I have found.

Recent literature has noted the potential drawbacks of the canonical two-way fixed effects DID framework when

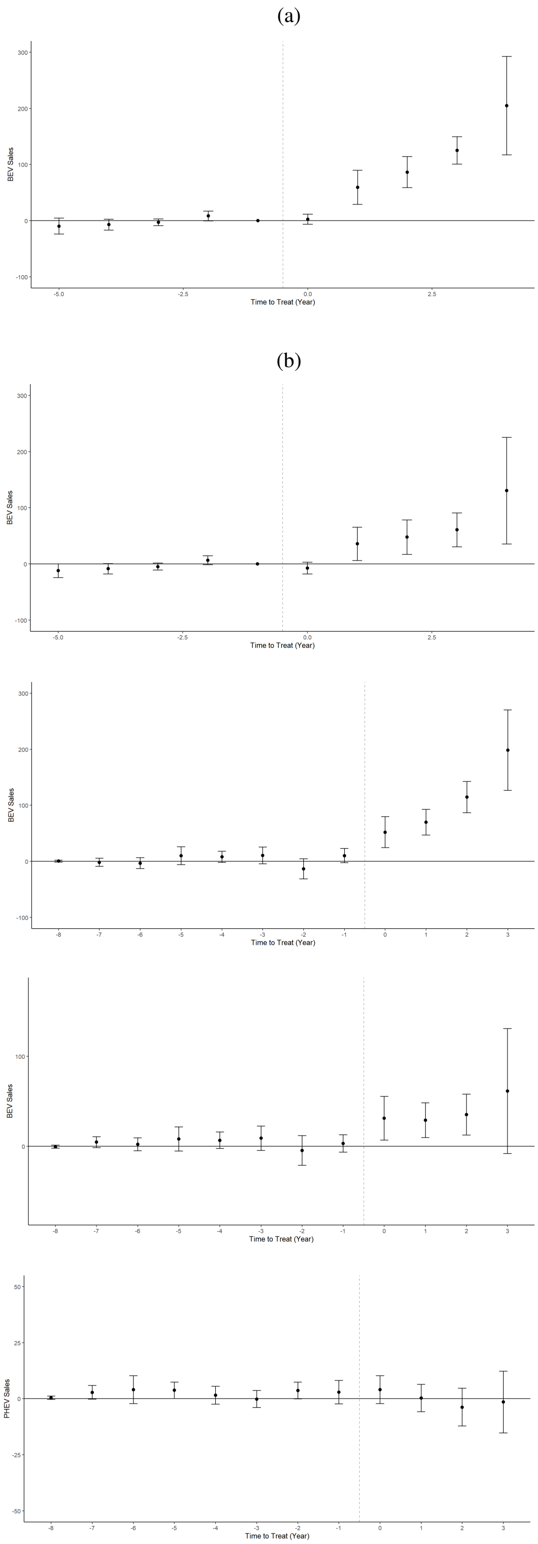



Fig. 4

applied to scenarios with staggered treatment and dynamic treatment effects (De Chaisemartin and D'Haultfœuille, 2020; Athey and Imbens 2022; Roth et al., 2022). In this case, the DID estimator (see from previous equations) represents a weighted average of the dynamic effects, where the weights could be negative. This is problematic because the estimated treatment effect may be negative, while the effect of participation is always non-negative. Given these concerns, as a second placebo test, I employ the doubly-robust DID estimator proposed by Callaway and Sant'Anna (2021). Their estimator allows for multiple time periods, variations in treatment time, and conditional parallel trends. The event-study plot using this estimator is shown in panel (b) of Figure 4. The estimates treatment effect similar to the conventional DID estimates in both the significance level and the numerical values.

In an alternative setup, the 68 ZIP codes that initially deployed fast chargers in 2018 or 2019 constitute the treatment group, and the never-treated ZIP codes (i.e. no fast chargers as of 2021) are used as controls. This setup would not pass a conventional test for parallel trend assumption due to the structural differences between treatment and control units (e.g. the never-treated ZIP codes tend to be poorer and more rural). However, the parallel trend assumption is satisfied when conditional on demographic variables including education, income, and population density. The event-study plots are shown in panel (c) of Figure 4. The outcome variable is BEV sales in the left plot and PHEV sales in the right plot as a robustness check.

In the left plot, the DID estimator is insignificant and numerically close to 0 in all pre-treatment periods. Starting period 0, the DID estimator becomes significantly positive at the 5% level until period 3. Under this specification, the annual BEV sales rise by 39.15 on average following the initial deployment of fast chargers. The estimated sales increase is smaller than the estimate from Figure 3, which could be attributed to the different chosen control groups. When I use PHEV sales as the control in Figure 3, the estimated treatment effect could bias upward due to the alternative explanatory factors such as the larger investments in improving BEV quality.

**Table 6. Heterogenous Effects of Fast Chargers Deployment**

| | (1)<br>High | (2)<br>Mid/Low | (3)<br>Urban<br>Sales | (4)<br>Rural | (5)<br>Suburb |
|---|---|---|---|---|---|
| BEV x post x fast | 97.52***<br>(28.94) | 12.70*<br>(7.444) | 46.90***<br>(14.87) | 19.41*<br>(11.25) | 96.39**<br>(35.59) |
| County/year fixed effects | Yes | Yes | Yes | Yes | Yes |
| Observations | 3,552 | 7,920 | 4,440 | 3,096 | 4,344 |
| R-squared | 0.638 | 0.503 | 0.526 | 0.471 | 0.591 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Because I do not control for the count of regular chargers, this could be a confounding factor. Following the initial fast charger installations, the count of regular chargers may also surge as developers expect more EV purchases and a rising demand for charging. Were this the case, it would be hard to tell whether the surge in BEV sales should be attributed to regular chargers or the initial fast charger deployment. To investigate this, I conduct a placebo test using the PHEV sales instead of the BEV sales as the outcome variable. Since PHEVs can only use regular chargers, a surge in regular chargers would likely lead to a sharp increase in PHEV sales as well. Based on the insignificant treatment effect shown on the right, I can rule out regular chargers as an alternative explanation. This again confirms that the fast charger deployment contributes to rapid BEV adoption. Meanwhile, the insignificant coefficients in the PHEV plot show that fast chargers do not reduce PHEV sales. The new BEV purchasers do not originally plan to buy PHEVs. Instead, they are likely the "accidental environmentalists" who are incentivized to purchase BEVs due to an increase in quality, namely shorter charging time in this case (Delmas et al., 2014). This is consistent with the results from Table 6 because these accidental environmentalists tend to have high income and live in the suburb.

## VII. Conclusion

Using a panel dataset at ZIP code level from 2010 to 2021, I have found that high-income and urban areas have adopted more EVs, and EV sales are increasing over time in California. There is a strong spatial correlation between EV adoption and public charger deployment due to the indirect network effects. By employing a Bartik-style IV, I have documented that a 1% increase in public chargers leads to a 0.72% to 1.1% increase in annual EV sales. The elasticity of EV sales with respect to public charger counts is higher in middle-income ZIP codes with more single-family homes. Due to the current high fixed cost of residential charging, the average total cost of home charging falls below that of public charging only in the long run. This incentivizes middle-income homeowners to defer installing home charging, so their EV demand is more responsive to the public charger deployment. While I study the heterogenous effects using interaction terms, more accurate results could be derived from more flexible models such as random coefficient models and causal forests (Swamy, 1970; Wager and Athey, 2018). Future research should also study how this elasticity varies over time. It is an open question whether this elasticity would decline as the driving range of EVs increases in the near term.

I have shown that the EV sales elasticity with respect to regular chargers are different from that to fast chargers. Fast chargers significantly boost BEV sales, especially in high-income suburbs, yet do not reduce PHEV sales. I interpret this as accidental environmentalists incentivized to substitute GVs with high-quality BEVs. A related emerging trend in the EV industry is the development of vehicle-to-grid (V2G). V2G chargers support reverse charging, which is to discharge the electricity from EVs to the power grids amid electricity shortages, and EV owners would be paid for this. An emerging literature studies the effect of V2G on EV adoption (Noel et al., 2019; Huang et al., 2021). Future research can use data from California's pilot V2G programs to examine the heterogenous effects of regular, fast, and V2G chargers on EV sales.

The Biden administration has signed multiple packages on expanding the EV charging network. In California, current rebates per regular charger are $3500 to $6000, and rebates per fast charger are $45000 to $70000. Standard microeconomics argues that the optimal amount of rebate should equal the positive environmental externalities associated with each public charger. In this paper, I do not study whether the large investments in public charging are cost-effective. Holland et al. (2016) have documented that EV purchases are over-subsidized in regions with dirty electric grids. In these regions, charger subsidies may well be higher than optimal. Xing et al. (2021) have found that EV purchase subsidies are regressive and inefficient when given mostly to the high-income population. Similar results may hold for EV charger subsidies when chargers are built in high-income versus low-income neighborhoods. The cost-effectiveness and distributional effects of public charger subsidies are important areas for future research.

## References


[1] Allcott, H. (2011). Social Norms and energy conservation. Journal of Public Economics, 95(9-10), 1082–1095. https://doi.org/10.1016/j.jpubeco.2011.03.003

[2] Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

[3] Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. Journal of Econometrics, 226(1), 62–79. https://doi.org/10.1016/j.jeconom.2020.10.012

[4] Bartik, T. J. (1991). Who benefits from state and local economic development policies? W.E. Upjohn Institute for Employment Research.

[5] Borusyak, K., Jaravel, H., & Spiess, J. (2022). Revisiting event study designs: Robust and efficient estimation. https://doi.org/10.47004/wp.cem.2022.1122

[6] Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica, 63(4), 841. https://doi.org/10.2307/2171802

[7] Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

[8] Clinton, B. C., & Steinberg, D. C. (2019). Providing the spark: Impact of financial incentives on Battery Electric Vehicle Adoption. Journal of Environmental Economics and Management, 98, 102255. https://doi.org/10.1016/j.jeem.2019.102255

[9] Costa, D. L., & Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. Journal of the European Economic Association, 11(3), 680–702. https://doi.org/10.1111/jeea.12011

[10] de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review, 110(9), 2964–2996. https://doi.org/10.1257/aer.20181169

[11] Delmas, M., Kahn, M., & Locke, S. (2014). Accidental environmentalists? Californian demand for teslas and solar panels. https://doi.org/10.3386/w20754

[12] Egbue, O., & Long, S. (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. Energy Policy, 48, 717–729. https://doi.org/10.1016/j.enpol.2012.06.009

[13] Ferraro, P. J., & Miranda, J. J. (2014). The performance of non-experimental designs in the evaluation of Environmental Programs: A design-replication study using a large-scale randomized experiment as a benchmark. Journal of Economic Behavior & Organization, 107, 344–365. https://doi.org/10.1016/j.jebo.2014.03.008

[14] Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. American Economic Review, 110(8), 2586–2624. https://doi.org/10.1257/aer.20181047

[15] Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2016). Are there environmental benefits from driving electric vehicles? the importance of local factors. American Economic Review, 106(12), 3700–3729. https://doi.org/10.1257/aer.20150897

[16] Holland, S. P., Mansur, E. T., Muller, N. Z., & Yates, A. J. (2019). Distributional effects of air pollution from Electric Vehicle Adoption. Journal of the Association of Environmental and Resource Economists, 6(S1). https://doi.org/10.1086/701188

[17] Holtsmark, B., & Skonhoft, A. (2014). The Norwegian support and subsidy policy of electric cars. should it be adopted by other countries? Environmental Science & Policy, 42, 160–168. https://doi.org/10.1016/j.envsci.2014.06.006

[18] Huang, B., Meijssen, A. G., Annema, J. A., & Lukszo, Z. (2021). Are electric vehicle drivers willing to participate in vehicle-to-grid contracts? A context-dependent stated choice experiment. Energy Policy, 156, 112410. https://doi.org/10.1016/j.enpol.2021.112410

[19] Kahn, M. E. (2007). Do greens drive hummers or hybrids? environmental ideology as a determinant of consumer choice. Journal of Environmental Economics and Management, 54(2), 129–145. https://doi.org/10.1016/j.jeem.2007.05.001

[20] Katz, M. L., & Shapiro, C. (1994). Systems competition and network effects. Journal of Economic Perspectives, 8(2), 93–115. https://doi.org/10.1257/jep.8.2.93

[21] Kotchen, M. J., & Moore, M. R. (2007). Conservation: From voluntary restraint to a voluntary price premium. Environmental and Resource Economics, 40(2), 195–215. https://doi.org/10.1007/s10640-007-9148-x

[22] Kumar, R. R., Chakraborty, A., & Mandal, P. (2021). Promoting Electric Vehicle Adoption: Who should invest in charging infrastructure? Transportation Research Part E: Logistics and Transportation Review, 149, 102295. https://doi.org/10.1016/j.tre.2021.102295

[23] Langbroek, J. H. M., Franklin, J. P., & Susilo, Y. O. (2016). The effect of policy incentives on electric vehicle adoption. Energy Policy, 94, 94–103. https://doi.org/10.1016/j.enpol.2016.03.050

[24] Li, S., Tong, L., Xing, J., & Zhou, Y. (2017). The market for electric vehicles: Indirect Network Effects and policy design. Journal of the Association of Environmental and Resource Economists, 4(1), 89–133. https://doi.org/10.1086/689702

[25] Mussa, M., & Rosen, S. (1978). Monopoly and product quality. Journal of Economic Theory, 18(2), 301–317. https://doi.org/10.1016/0022-0531(78)90085-6

[26] Neubauer, J., & Wood, E. (2014). The impact of range anxiety and home, workplace, and public charging infrastructure on simulated battery electric vehicle lifetime utility. Journal of Power Sources, 257, 12–20. https://doi.org/10.1016/j.jpowsour.2014.01.075

[27] Noel, L., Papu Carrone, A., Jensen, A. F., Zarazua de Rubens, G., Kester, J., & Sovacool, B. K. (2019). Willingness to pay for electric vehicles and vehicle-to-grid applications: A nordic choice experiment. Energy Economics, 78, 525–534. https://doi.org/10.1016/j.eneco.2018.12.014

[28] Okada, T., Tamaki, T., & Managi, S. (2019). Effect of environmental awareness on purchase intention and satisfaction pertaining to electric vehicles in Japan. Transportation Research Part D: Transport and Environment, 67, 503–513. https://doi.org/10.1016/j.trd.2019.01.012

[29] Roth, J., Sant'Anna, P., Bilinski, A., & Poe, J. (2022). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. https://doi.org/10.48550/arXiv.2201.01194

[30] Shao, L., Yang, J., & Zhang, M. (2017). Subsidy scheme or Price Discount Scheme? mass adoption of electric vehicles under different market structures. European Journal of Operational Research, 262(3), 1181–1195. https://doi.org/10.1016/j.ejor.2017.04.030

[31] Springel, K. (2021). Network externality and subsidy structure in two-sided markets: Evidence from electric vehicle incentives. American Economic Journal: Economic Policy, 13(4), 393–432. https://doi.org/10.1257/pol.20190131

[32] Swamy, P. A. (1970). Efficient inference in a random coefficient regression model. Econometrica, 38(2), 311. https://doi.org/10.2307/1913012

[33] van Dijk, J., Delacrétaz, N., & Lanz, B. (2022). Technology adoption and early network infrastructure provision in the market for electric vehicles. Environmental and Resource Economics, 83(3), 631–679. https://doi.org/10.1007/s10640-022-00703-z

[34] Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

[35] Xing, J., Leard, B., & Li, S. (2021). What does an electric vehicle replace? Journal of Environmental Economics and Management, 107, 102432. https://doi.org/10.1016/j.jeem.2021.102432

[36] Zarazua de Rubens, G. (2019). Who will buy electric vehicles after early adopters? using machine learning to identify the electric vehicle mainstream market. Energy, 172, 243–254. https://doi.org/10.1016/j.energy.2019.01.114

# The Political Implications of Corporate Philanthropy: Evidence from Pivotal Politics and Legislator Behavior on Environmental Issues

JC Martinez
Advisor: Natalie Cox; Assistant Instructor: Jessica Min

*Princeton University*

*Abstract*— **In this paper, we explore (I) whether large corporations use charitable giving strategically to influence politics and (II) the extent to which it might be effective in changing legislator behavior. First, we investigate the extent to which a senator's importance predicts PAC and philanthropic contributions from Fortune 500 and S&P 500 corporations to nonprofits with which the senator is personally affiliated. We conduct this analysis by fashioning a novel measure of U.S. senator importance to firm profitability using the pivotal politics theoretical model of U.S. lawmaking, and exploiting U.S. Senate panel data from the 105th to 113th Congresses. Estimating different specifications of standard and fixed effects regression models, we find little evidence suggesting corporate increases in response to this political incentive; however, (1) the measure does not always predict PAC contributions well either and (2) we do find evidence suggestive of non-politically strategic CSR usage and corporate awareness of the political careers of receiving charity board members. Second, using panel data on the House of Representatives from the 106th to 113th Congresses, we explore the differing extents to which oil and gas companies' PAC contributions and philanthropic giving to nonprofits linked to House Representatives affect how they vote on environmental issues and talk about climate change during congressional floor speech. Using a two-way fixed effects regression, which accounts for temporal precedence by incorporating lead and lagged values, we find a more robust relationship between energy corporations' PAC contributions and anti-environmental legislator behavior than we do for philanthropic contributions. Further, we use two instruments to implement a two-stage-least-squares within estimator that also controls for two-way-fixed-effects, finding that—if instrumental variable assumptions hold—PAC contributions cause an anti-environmental shift in behavior while we find little evidence that philanthropic giving does. Finally, the analysis for the second question presents mixed evidence that energy corporations use philanthropy as a tool for political influence.**

## I. Introduction

Since Madison's warnings of the "mischiefs of faction" in *Federalist Paper* No. 10, there have been concerns that the political activity of organized interest groups has a corrosive effect on the quality of democratic representation in America. Many argue that large corporations, in particular, are effective in using their resources to shift public policy towards fulfilling business objectives at the expense of voter welfare (Ferguson 1995; Gilens and Page 2014; Hacker and Pierson 2010; Lindblom 1977; McConnell 1966; Schlozman et al. 2012). The concentration of economic power attained by giant modern corporations, whose revenues often surpass the GDPs of national governments, exacerbates this concern and heightens the risk of a positive feedback loop in which economic power and political power reinforce each other to the detriment of democracy (Zingales 2017).

However, a classic question in the study of political economy asks why so *little* corporate money is observed in politics (Tullock 1972), denoted as Tullock's Puzzle. Given the sheer dollar value of public expenditures at stake, the amount of corporate money observed in politics would imply an exceptionally high return on investment if firms were to succeed in tilting policy favorably to their ends (Ansolabehere et al. 2003). In response to this puzzle, Bertrand et al. (2020), henceforth BBFT, argue that the traditional, directly observable channels of corporate political influence—campaign finance and lobbying—may not be the only ways big firms try to exert influence, and that nominally benign corporate activities falling under the umbrella of corporate social responsibility (CSR) may function as opaque avenues of influence. Specifically, BBFT find systematic evidence that large companies donate more, in both value and frequency, to charities linked to politicians (either by residing in their congressional district or by direct personal affiliation—e.g., occupying a board position) who are particularly important to the company's financial gain. Legislators benefit politically from these donations as they provide an opportunity to claim credit and associate their brand with the charity's good deeds in the eyes of the electorate (p. 2067).

Their findings suggest that large firms use philanthropy as an instrument for political influence. Moreover, using a stylized political-economic model (p. 2096), they estimate the amount of annual politically motivated corporate charitable giving to be 2.5 to 6.3 times *greater* than annual PAC contributions. Politically motivated CSR giving, then, may involve very large sums of money which (1) potentially have nontrivial political influence, (2) bypass current regulatory frameworks and voter attention, and (3) amount to corporate influence subsidized by taxpayers (Charitable, 2021).

This paper builds on BBFT's seminal work to explore (I)

I foremost thank Professor Natalie Cox, my advisor, and Jessica Min, my assistant instructor, for helping me complete this project since its inception as well as advising and supporting me through its difficult periods. I thank Professor Nolan McCarty, Rafael Parente, Sebastian Roelsgaard, and Angela De Santis for discussions which contributed to my thinking on the paper's topic. I am especially grateful to Professor Michal Kolesár and Professor Mark Watson for allowing me to bother them countless times with JP questions after class. I thank Professor Jeremiah Bohr for generously sharing data on congressional rhetoric on climate change. Lastly, I thank Bobray Bordelon for data collection assistance and Oscar Torres Reyna for data cleaning assistance.

whether large corporations use charitable giving strategically to influence politics and (II) the extent to which it might be effective in changing legislator behavior. First, fashioning a measure of U.S. senator importance to firm profitability from the Pivotal Politics theoretical model of U.S. lawmaking and utilizing U.S. Senate panel data from four sources on the 105th to 113th Congresses, we investigate the extent to which a senator's importance predicts philanthropic contributions from Fortune 500 and S&P 500 corporations to nonprofits with which the senator is personally affiliated. We compare these trends to those corporations' PAC contributions, which are plainly political, to observe if the trends are similar. Estimating different specifications of standard and fixed effects regression models, we find little evidence suggesting corporate philanthropy increases in response to this political incentive but the measure does not always predict PAC contributions well either and we do find evidence potentially suggestive of non-politically strategic CSR usage and corporate awareness of the political careers of receiving charity board members.

Second, using panel data on the House of Representatives from the 106th to 113th Congresses, we explore the differing extents to which oil and gas companies' PAC contributions and philanthropic giving to nonprofits linked to Congress affect how representatives vote on environmental issues and talk about climate change during congressional floor speech. We quantify "how they vote" as the fraction of roll call votes on environmental issues during a given Congress in which the politician took the pro-environmental position as judged by the League of Conservation Voters (LCV). We quantify "how they talk" by using a set of quantitative text analysis variables generated by Guber, Bohr, and Dunlap (2021) which record the prevalence of 18 'topics' in Congressional floor speeches on climate change. As we show in Section IV, the prevalence of some of these topics (e.g., *anecdotal denial, cap-and-trade, climate change denial*) correlate strongly with anti-environmental voting while other topics go hand-in-hand with pro-environmental voting (e.g., *public health, extreme weather, economic opportunity*). We group the speech variables according to the sign of their correlation with voting on environmental issues to create tractable outcome variables. Using a two-way fixed effects regression, which accounts for temporal precedence by incorporating lead and lagged values, we find a more robust relationship between energy corporations' PAC contributions and anti-environmental legislator behavior than we do for philanthropic contributions. Further, we use two instruments to implement a two-stage-least-squares (2SLS) estimator that also controls for two-way-fixed-effects (2FWE), finding that—if instrumental variable assumptions hold—PAC contributions cause an anti-environmental shift in behavior while we find little evidence that philanthropic giving also does so. The analysis for the second question presents mixed evidence that energy corporations use philanthropy as a tool for political influence.

Importantly, utilizing our particular 'pivotal' measure of politician importance allows us to test BBFT's thesis in the US Senate, whereas much of BBFT's analysis was restricted to the House of Representatives. For instance, BBFT demonstrate that a House member's tenure drives corporate philanthropic giving (in a manner parallel to PAC contributions) by conducting an event study on House member exits from office, documenting an immediate withdrawal of charitable giving to charities located in the member's congressional district followed by a gradual build up once the politician secures tenure (p. 2087). But panel data on the Senate provides far less turnover and thus variation, hampering that methodology.

Our measure bypasses this problem. The pivotal politics theory, henceforth PPT, posits that legislative output is largely a function of the preferences of a few pivotal legislators whose unique statuses derive not from special parliamentary prerogatives, but rather from the position of their policy preferences relative to those of other legislators when ordered along an ideological spectrum. If their position on this continuum places them sufficiently close to a quantile made significant by the Senate's majority vote and supermajoritarian procedures—ending the filibuster and overturning an executive veto—we call them pivotal. This preference-based measure, therefore, has at least three advantages: first, it exploits variation in the preferences of individual politicians over time; second, even when using lifetime politician ideology scores, it generates relatively high variation insofar as the entry or exit of just one senator in a new Congress can shift rankings and thus also shift pivotality; third, it provides a measure of importance besides tenure, which is less important in the Senate, a chamber "far more egalitarian and individualistic than the hierarchical and institutionally driven House" (Volden and Wiseman 2018, p. 731).

The rest of this paper proceeds as follows. Section II provides a primer on the theoretical properties of the pivotal politics model that underpin our measure of senator importance. Section III briefly expounds the relevant aspects of the literatures straddled by the paper and connects them to the paper's unique and significant contributions. Section IV describes the data used in this paper. Section V explains the methodologies employed to explore the paper's two questions. Section VI presents and interprets results. Section VII summarizes and concludes.

## II. The Pivotal Politics Theory

Krehbiel (1998)'s pivotal politics theory (PPT) is devoted primarily to identifying the conditions under which congressional gridlock is broken. The PPT postulates that policy proposals can be ordered by their ideological content along a single line. This is the *policy space*, which is one-dimensional, continuous, and might conceptually be conceived of "as a continuum on which liberal policies are located on the left, moderate policies are located in the center, and conservative policies are located on the right" (p. 21). As will be discussed further in Section IV, we use different measures of ideology to span the unidimensional policy space. For now, however, we simply explain the model in terms of liberalism and conservatism.

The point representing the ideological content of the status quo policy is denoted $q$. The game consists of $n$ legislators and one president, and each player seeks to maximize their univariable, single-peaked, and symmetric utility function by trying to effectuate legislation closest to their ideal point of policy, which maps onto to the global maximum of their utility function. The legislators operate in a unicameral legislature, which we take to represent the U.S. Senate.

With the ideal points of all the legislators distributed along the policy space, the PPT then notates the median congressional voter's ideal point $m$, and at this point the model closely resembles traditional median voter games in which a simple majority vote will elect, among the choices offered, the proposal closest to m in the policy space. The theory then departs from the median voter literature by introducing "two *supermajoritarian procedures*: the executive *veto*, and the Senate's *filibuster* procedures" (p. 22). These two structures imply that any changes in legislation depend upon two crucial players: "the *filibuster pivot* with ideal point $f$ and the *veto pivot* with ideal point $v\ldots$ defined with reference to the president, whose ideal point is $p$" (p. 23). These pivots are equivalent to the American supermajoritarian procedures upon which they are based; it takes a $2/3^{\text{rds}}$ majority vote from Congress to nullify an executive veto and a $3/5^{\text{ths}}$ senatorial vote to invoke cloture of a filibuster which can otherwise delay the passage of a bill indefinitely. The pivots are defined in relation to the president in the sense that, if $p$ is to the left (right) of $m$, then the filibuster pivot is the legislator who no more than $2/5^{\text{th}}$ of the legislature is more conservative (liberal) than and the veto pivot is the legislator who no more than $1/3^{\text{rd}}$ of the legislature is more liberal (conservative) than. For clarification, see Krehbiel's diagram (p. 23), shown here, illustrating the case of a distinctly liberal president.

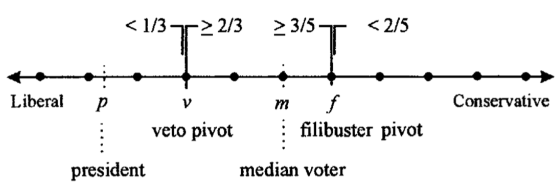



Figure 2.2
Pivotal legislators if the president is liberal

A simple majority vote can propose a bill mapping onto point $b$ in the policy space, which depends on whether $b$ or $q$ is closer to $m$. The bill, however, has a few more hurdles to clear before it becomes a law. The filibuster pivot decides if the bill can escape congressional purgatory by ending extended debate with the $3/5^{\text{ths}}$ vote (which, in turn, depends on whether $b$ or $q$ is closer to $f$). If so, the president can choose to veto the bill (if $q$ is closer to $p$ than $b$ is) or not. If they do, then the veto pivot decides if the legislature can muster the $2/3^{\text{rds}}$ majority vote necessary for an override (which occurs if $b$ is closer to $v$ than $q$ is).

As the proposed bill can be rejected at several different steps, the PPT provides an intuitive explanation for why gridlock is so common. More precisely, the PPT goes on to outline five intervals, informed by the foregoing ideal points of the pivotal players, in which the status quo policy's $q$ might lie. Each interval delineates the necessary ideological content a bill must have to be successfully passed, apart from the gridlock interval, which, if $q$ lies within it, ensures that no new bill will be passed, regardless of its ideological content.

As the preferences of the pivotal players determine where these intervals fall, the PPT therefore identifies the veto, median, and filibuster pivots as having an outsized importance in the legislative process. We hypothesize that corporations interested in influencing legislation via donations might disproportionately direct contributions towards senators whose ideological complexions identify them as pivotal.

Accordingly, we use both time-invariant (i.e., lifetime scores) and time-varying (i.e., scores for a given congressional cycle) measurements of three distinct ideologies—economic (e.g., taxation), social (e.g., abortion), and environmental—to create pivotal indicator variables for each of the three pivots every Congress. (We differentiate environmental from the other categories to specifically partition contributions from energy companies. We further elaborate on this distinction in Section III.A.)

We say a senator has *pivotality* if they fall within a $\pm 4$ range of the pivotal quantile. For example, in the case above, all 9 senators from the $56^{\text{th}}$ most liberal to the $64^{\text{th}}$ most liberal are deemed to have filibuster pivotality. We argue that including this range is important practically as well as theoretically. Practically, if the pivotality indicator were not 'turned on' for a range of senators around the pivot, but rather only identifies one individual per pivot, respectively, our results would likely be highly sensitive to any noise in the estimate of the 'true' ideological ranking (if such a concept exists). Theoretically, the reasons which make us think that corporations might target pivotal senators also imply that senators immediately next to the pivotal senators might also be targeted. To illustrate, consider the example where the $60^{\text{th}}$ most liberal senator (let this be the filibuster) is only marginally more liberal than the $61^{\text{st}}$ most liberal senator. If a corporation succeeds in shifting the preferences of the filibuster rightward, the preference of the new filibuster is only slightly less liberal than before. Lastly, the interval length of the pivotal quantile, equal to 9 senators, was chosen to balance the tradeoff between mitigating the effects of noisy ideology measures and identifying a differentiated set of pivotal politicians, comprising roughly 10% of the Senate body for each of the three kinds of pivotality.

## III. Literature Review

### A. Corporate Social Responsibility

CSR generally refers to the body of policies or doctrines of self-regulation by which corporations who state an intent to positively impact the world realize that intent. To clarify this somewhat ill-defined concept, Bénabou and Tirole (2010) classify three broad motivations for CSR: (1) *strategic CSR*, involving taking a socially responsible stance in order to secure a stronger market position and increase profits by attracting socially conscious consumers and employees; (2) *delegated philanthropy*, by which stakeholders give up money so that the firm can conduct socially positive behavior on their behalf; and (3) *insider initiated philanthropy*, in which willingness to spend money for a certain cause stems from the personal predilections of those in management, who take advantage of their company's flexible governance norms. If the motive of strategic CSR dominates, there can be equivocal social effects if it leads firms to contribute to charitable causes while, as BBFT (p. 2072) cite Baron (2001), acting as "a means of placating regulators and public opinion in order to avoid strict supervision in the future" (p. 10) and thereby imposing a social cost and benefit whose net effect is ambivalent.

Like BBFT's article, the questions of this paper seek to contribute to our understanding of the extent to which this is a problem. If evidence is found that business actors in the energy sector seek to use CSR to move policy in a direction beneficial to their interests but adverse to the environment, as they seek to do with other mechanisms of influence (Brulle, 2018; Downie, 2017; Kang, 2015; Tabuchi, 2021), it would serve as a powerful illustration of an example in which the net societal impact of strategic CSR is likely very negative.

### *B. Empirical Research Testing the Pivotal Politics Theory*

By investigating the relationship between our particular measure of senator pivotality and corporate PAC donations—especially in using environmental preferences and donations from large energy firms—we also contribute to the PPT literature an indirect test of the PPT which may be more immune to possible criticisms of previous empirical research.

First, note that the PPT's primary prediction that the length of the gridlock interval and legislative productivity are negatively related has not fared well in empirical studies. While Heitshusen and Young (2006) do find evidence of this negative relationship, the great majority of studies find no support (e.g., Chiou and Rothenberg 2003), weak support (e.g., Covington and Bargen 2004, Stiglitz and Weingast 2010, Clinton 2007), or mixed results (e.g., Woon and Cook 2015, Richman 2011, Krehbiel et al. 2005) for the negative relationship. These results might suggest that the PPT is too simple to reasonably capture the dynamics of the more complicated US lawmaking from which it abstracts, which may give pause to using it as a measure of senator importance to firm profitability.

However, Gray and Jenkins (2017) dispute this notion and identify a critical flaw in the methodologies of the preceding studies. The problem with these studies lay in their treatment of the PPT "as making legislative predictions from a single dimension of ideological policy preferences"—in practice this is primarily economic ideology—and failure to observe that "[s]ome policy areas draw on different dimensions of preferences" so that "if we narrow the testing to an empirically workable dimension of ideological preferences, we should only use outcomes that would rely on that dimension" (p. 126). They go on to provide evidence that when this issue is accounted for by examining legislative productivity only for the set of laws which would reasonably depend on the ideological preferences used to measure the gridlock interval, the PPT theory's core prediction holds.

By also including the relationship between environmental pivots and oil and gas company contributions specifically, we incorporate the Gray and Jenkins (2017) insight that empirical tests of the PPT should use outcome variables which are highly responsive to the ideological preferences that define the pivots. In fact, our indirect test may also provide evidence regarding the relevance of Gray and Jenkins (2017) argument by including economic and social ideology measures for comparison.

Finally, note that we are not the first to think of examining the relationship between senator pivotality and contributions. Mixon et al. (2005) find a positive, significant relationship between PAC contributions and pivotality in the US Senate from the 105$^{\text{th}}$ to the 107$^{\text{th}}$ Congresses. Nonetheless, our indirect test still offers a unique contribution to the empirical PPT literature for a few reasons; Mixon et al. (2005) provide estimates for data during only 3 Congresses (while we use data spanning 9 Congresses), run only cross-sectional regressions whereas we also control for entity fixed effects, and do not incorporate the Gray and Jenkins (2017) insight. Moreover, our test also gives information on how corporations might respond to pivotality with their philanthropic giving differently than they do with their PAC contributions, which can enrich our understanding of the PPT in practice.

### *C. The Influence of Corporate Contributions on Lawmaking in Congress*

The great majority of studies analyzing the extent to which contributions lead to influence in the legislative process investigate the impact of PAC contributions on Roll Call voting behavior and find mixed results (see Roscoe and Jenkins (2005), Ansolabehere et al. (2003) for meta-analyses). However, there are a few problems with this approach. First, influence from contributions in the legislative process is likely to manifest earlier and more subtly than Roll Call votes, when "less visible actions are taken to kill bills quietly or to negotiate the details of legislation that can matter so greatly to donors" (Powell, 2014, pp.75-6). A lot of influence simply comes in other ways such as impacting legislative involvement in committees (Hall and Wayman, 1990) and increased legislator access enjoyed by contributors (Kalla and Brookman, 2016). This drawback motivates including congressional rhetoric on climate change during floor speeches in addition to Roll Call voting as an outcome variable in our analysis. In fact, this is the first paper to my knowledge which examines the impact of contributions on legislator speech patterns. We also document strong correlations between climate speech behavior and environmental Roll Call voting behavior, which may lend credibility to using congressional speech as an outcome variable when studying the effect of money in corporate politics moving forward.

A second criticism of the wealth of studies analyzing PAC contributions' effect on roll call votes is that they do not account for simultaneous causality (Ansolabehere et al., 2003): in our case, do energy companies contribute to politicians who already act anti-environmentally to support their incumbency status (i.e., henceforth the *investment* hypothesis), or are corporate energy contributions influencing them to act anti-environmentally (i.e., henceforth the *influence* hypothesis)? Goldberg et al. (2020) argue that, for PAC contributions from energy corporations, the investment hypothesis fits the data better by conducting a so-called cross-lagged panel analysis which accounts for dynamics. We use this idea that temporal precedence can help distinguish the causal direction between donations and legislator behavior in analyzing speech and Roll Call voting response to CSR and PAC contribution. Moreover, we introduce two instruments which act as exogenous measures of firm generosity to help identify the causal effect of contributions on legislator behavior (and not vice versa).

## IV. Data

### *A. Ideology, Pivotality, and Corporate Contributions in the US Senate (105$^{th}$-113$^{th}$ Congresses)*

This section describes the panel dataset used to investigate the sensitivity of corporate CSR and PAC contributions to senator pivotality from the 105$^{th}$ to the 113$^{th}$ Congress, which were constructed from four sources. First, we obtain Poole and Rosenthal's popular DW-NOMINATE two-dimensional scores from voteview.com (Lewis et al., 2021), which use Roll Call votes to estimate the economic and social ideology in terms of liberalism-conservatism of all Congress members. There are the traditional DW-NOMINATE (DWN) scores, which are time-invariant, lifetime measures, and Nokken-Poole (NP) estimates, which allow for ideology to change each Congress (Boche et al. 2018). Both scores range from -1, the liberal extreme, to +1, the conservative extreme. Our chamber seniority and political party variables also come from Lewis et al. (2021). Second, to obtain a unidimensional measure of senators' preferences on environmental issues, we use scores from the League of Conservation Voters (LCV), an environmental advocacy organization that tracks voting records on environmental issues and assigns legislators a yearly score capturing how pro-environmental their voting record was that year. The score ranges from 0 to 100, increasing in pro-environmentalism. We average the yearly LCV scores to obtain a score for each 2-year Congress. With these 6 unidimensional measures of senatorial preferences in hand (3 ideological spectrums and lifetime versus time-varying scores), we create separate rankings for all senators in each Congress, and then create veto, median, and filibuster indicators equal to one if the senator falls within a $\pm$ 4 range of what the 'true' pivotal seat would be if our ideology measures were completely accurate. Each pivotal region, then, is 9 senators long (see Section II).

Third, data on CSR contributions come from BBFT, who compiled complete data on 324 corporations in the set of S&P 500 and Fortune 500 companies in 2014 which could be linked to an active corporate philanthropic foundation. All CSR contribution data measures giving only from this set of 324 corporations. These corporations donated to many nonprofits over the sample period, some of which are linked to legislators in two different ways. The first kind of link happens if a non-profit is located in the geographical congressional district which the legislator represents (more applicable to House members but can of course be aggregated to the state level for Senators), henceforth the *district* link. The second type of link is arguably stronger, occurring if the politician holds a position with the non-profit (e.g., a board membership). BBFT obtained this data from the legally required personal financial disclosure forms legislators must fill out, so this will henceforth be called the *PFD* link. In this section on the Senate, only the PFD link is used. We cleaned BBFTs raw data to create Senate panel data recording the total amount of CSR donations to Senators' PFD charities per Congress. We also compute the foregoing just from corporations of the 324 that are energy companies, of which there are 41 (e.g., ExxonMobil, Chevron, etc.). Fourth and finally, data on energy corporations' PAC contributions to senators come from Goldberg et al. (2020).

Table 1 presents summary statistics. Most observed ideologies range almost the length of their bounds, with most means falling near the middle of the measure except for lifetime social scores, with a liberal mean of -0.11. The average amount of a CSR contribution to a senators' PFD charities, \$807,000, is larger than average PAC contributions, \$630,000. This difference is more pronounced for energy companies (\$98,000 versus \$48,000). CSR donations are strictly nonnegative, whereas PAC contributions can be negative if returned or voided. Average seniority over all senators and Congresses is 11.8 years. Half of senators observed are Republicans.

**Table 1**

| Variable | N | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| ***Ideological Measures*** | | | | | |
| Econ. Score (lifetime) | 888 | .04 | .37 | -.76 | .9 |
| Econ. Score (time varying) | 888 | .04 | .39 | -.99 | .99 |
| Social Score (lifetime) | 888 | -.11 | .34 | -1 | .93 |
| Social Score (time varying) | 888 | .04 | .37 | -.76 | .9 |
| Env. Score (lifetime) | 887 | 48.3 | 34.8 | 1 | 98 |
| Env. Score (time varying) | 887 | 48.5 | 38.5 | 0 | 100 |
| ***Corporate Contributions*** | | | | | |
| All $CSR (PFD) | 379 | 806,609 | 2,261,627 | 0 | 21,292,685 |
| Energy $CSR (PFD) | 332 | 98,176 | 426,671 | 0 | 5,610,505 |
| All $PAC | 650 | 629,686 | 786,393 | -78,000 | 3,899,420 |
| Energy $PAC | 665 | 47,813 | 144,171 | -11,800 | 2,763,700 |
| ***Political Variables*** | | | | | |
| Chamber Seniority | 888 | 11.81 | 10.07 | 1 | 49 |
| Republican (0-1) | 888 | .5 | .5 | 0 | 1 |

Lastly, note differences in data availability. Ideally, for all variables we would have 100 senators and 9 Congresses, for a total of 900 observations. For ideology and political variables, we have 12-13 observations assumed to be missing at random. We see less than 900 PAC observations because Goldberg et.al (2020)'s data do not map perfectly onto ours, and again we assume data missing at random. Note that only a small subset of observations is recorded as having received CSR giving to PFD charities. This is because many politicians are not recorded as occupying a position in a charity.[2] Thus, reported statistics are only for senators who we know have charity positions. A concern might be that having a PFD charity is related to ideology. However, regressing an indicator for having a PFD charity on all non-contribution variables above and then conducting an F-test for joint significance yields a p-value of 0.46. Finally, as a sanity check we see if ideology measures relate to political party in the expected ways. Compared to Democrats and Independents, Republicans on average have a lifetime environmental score 62.2 percentage points lower (t = - 59.6) and a lifetime economic ideology 35 percentage points (t = 75.74) more conservative.

### *B. Energy Corporate Contributions, Environmental Roll Call Votes, and Rhetoric on Climate Change in the House of Representatives (106th-113th Congresses)*

This section describes the panel data used to explore the effect of energy corporations' PAC and CSR contributions on House Representative behavior on environmental issues in the 106th-113th Congresses.[3] The data come from three sources. First, we aggregate BBFT's company-charity-district-Congress specific data for each congressional district and Congress to obtain the amount donated by the 41 energy companies in our sample, one, to charities located in the district, two, to the PFD charities of the district's representative, and three, as PAC donations to the districts' representative. Second, from the LCV we obtain data on every House Roll Call vote pertaining to an environmental issue in our sample period and aggregate to compute, for each district and Congress, the fraction of such votes in a given Congress in which the district's representative made the pro-environmental decision as judged by the LCV.

Third, we employ Guber et al. (2021)'s quantitative text analysis on a corpus of Congressional floor speeches on climate change during our sample period. They use an unsupervised topic model to detect associations between words used over different speeches. They identify 18 such climate 'topics,' and for each speech assign 18 scores which can be thought of as representing the extent to which the speech is "about" each topic. We then aggregate to have 18 variables recording the average topic value for all of each politicians' speeches during each Congress. After cleaning these politician specific data to merge to our congressional-district data, we then group these 18 topics for purposes of tractability into pro-environmental, anti-environmental, and neutral categories using a rough heuristic described below.

Table 2 reports summary statistics. The average amount a district's charities receive from our sample of large energy companies is $410,00, far greater than the $12,500 average PAC contribution to the district's representative. For representatives recorded as having PFD charities, the average energy CSR donation amount to their PFD charities is $88,000. The maximum value, however, is far higher; one legislator (Rep. Kay Granger, 111th Congress) received $12.5 million energy company contributions in one Congress. The congressperson behavioral variables show that voting is almost perfectly split at 0.5, while neutral rhetoric is less common than either anti- or pro-environmental speech.

**Table 2**

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| ***House Representative behavior on Environmental issues*** | | | | | |
| Pro-env. Vote proportion | 3,387 | .5 | .41 | 0 | 1 |
| Neutral speech proportion | 437 | .26 | .17 | .002 | .892 |
| Pro-env. speech proportion | 437 | .39 | .25 | .004 | .979 |
| Anti-Env. Speech proportion | 437 | .35 | .23 | .011 | .994 |
| ***$ Energy Corporation Contributions*** | | | | | |
| $CSR (cong. district) | 3,387 | 409,944 | 1,378,350 | 0 | 32,543,856 |
| $CSR (PFD) | 1,348 | 88,045 | 518,171 | 0 | 12,488,598 |
| $PAC | 3,387 | 12,491 | 16,346 | 0 | 167,000 |
| ***Instrumental Variables*** | | | | | |
| PAC generosity proxy | 3,387 | 204,403 | 65,992 | 124,444 | 331,139 |
| CSR generosity proxy | 3,387 | 6,624,010 | 2,829,274 | 3,659,509 | 11,807,240 |

The instrumental variables, which will be explained more in Section V, are, for a given district, the average PAC and CSR (district) generosity to *other* districts of the set of companies donating to the given district.

How do we categorize the 18 speech topics? Even though topic prevalence variables do not directly map onto agents' preferences (Guber et al., 2021, pp 541-2; Lauderdale and Herzog, 2016), many go hand-in-hand with Roll Call voting behavior on environmental issues, and therefore in the aggregate arguably provide a good measure of preferences. We estimate $V_{it} = \beta_0 + \beta_1 * Topic^j{}_{it} + error_{it}$, for j = $1 \ldots 18$, in which $V_{it}$ is the proportion of relevant Roll Call votes in Congress t on which district *i*'s representative made the pro-environmental choice, and $Topic^j{}_{it}$ is one of 18 topic prevalence variables for the floor speeches given by district *i*'s representative in Congress *t*. Heuristically, we place the topic in the neutral category if p $>$ 0.1, and assign to the other categories depending on sign. These correlations comport with expectations, as shown by Table 3.

**Table 3: Regressing Pro-Env. Voting on Topic Prevalence**

| Pro-environmental | | Neutral | | Anti-environmental | |
|---|---|---|---|---|---|
| ***Topic*** | ***Reg. coef.*** | ***Topic*** | ***Reg. coef.*** | ***Topic*** | ***Reg. coef.*** |
| School Programs | .44 (t = 2.43) | Budget Issues | -.25 (t = -1.15) | Legislative Text | -.34 (t = -1.67) |
| Economic Opportunity | 1.52 (t = 12.35) | Climate Change Impacts | -.02 (t = -0.14) | Cap-And-Trade | -1.54 (t = -12.74) |
| Extreme Weather | .79 (t = 8.06) | Congressional Procedures | -.27 (t = -1.26) | Anecdotal Denial | -1.17 (t = -6.09) |
| Public Health | .84 (t = 6.16) | International Accords | -.01 (t = -0.05) | Climate Change Denial | -.75 (t = -5.44) |
| Energy Efficiency | .85 (t = 5.42) | Emissions | .09 (t = 0.42) | Oil Industry | -.88 (t = -3.74) |
| Foreign Policy | .72 (t = 3.31) | | | National Security | -.52 (t = -2.14) |
| | | | | Non-Fossil Energy | -1.07 (t = -3.88) |
| **Total** | 1.02 (t = 18.17) | **Total** | -.16 (t = -1.39) | **Total** | -1.14 (t = -19.12) |

Figure 1 displays these results graphically, while Figure 2 provides time series evidence that these three broad categories correlate with environmental Roll Call voting over time.

[2] I do not code all politicians who do not have charity positions as having $0 PFD charitable contributions because, one, the manner of data collection implemented by BBFT was not exhaustive and, two, it would likely dilute the significance of estimates insofar as corporations are less likely to try to use CSR to influence politicians who do not have a PFD charity in the first place; doing so, then, would weaken the observed relationship.

[3] When filtering BBFT's data for energy company contributions, the 105th Congress drops out, and hence I restrict my analysis to the 8 Congresses for which I energy CSR and PAC data. I do not impute 0's for reasons similar to those articulated in footnote no. 2.

Figure 1: Environmental Voting vs Rhetoric on Climate Change in the House, 106 - 113th Congress

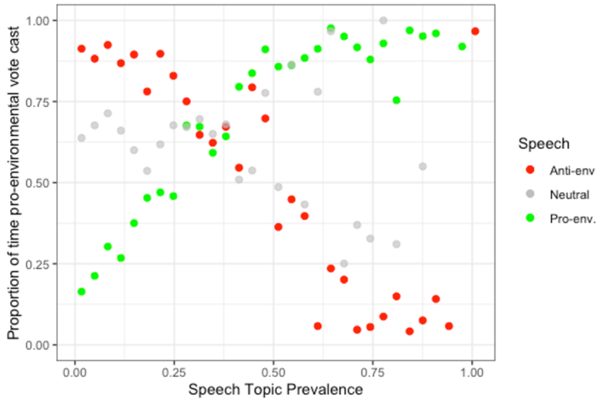



Figure 2: House Rep. Avg. Environmental Behavior over time

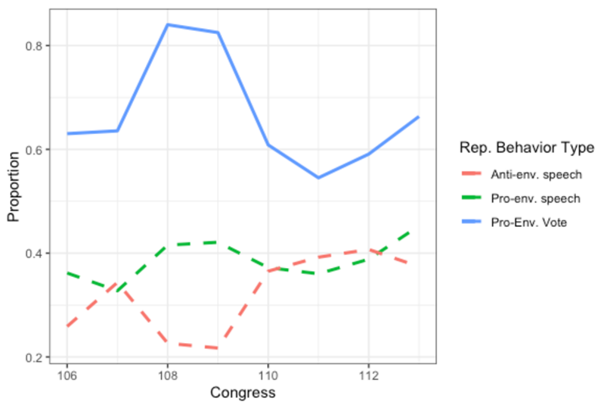
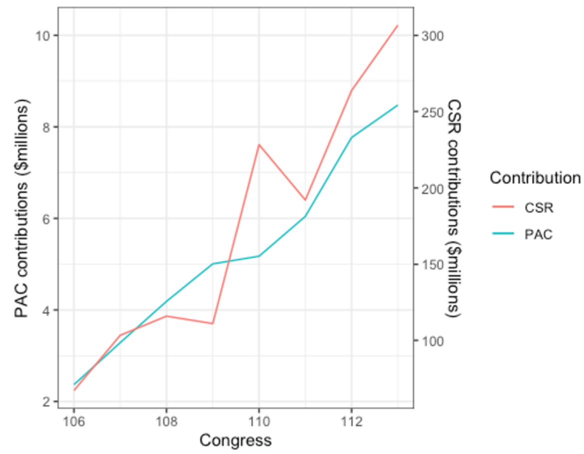



Figure 2 also demonstrates that average environmental behavior in the House changes significantly over the sample period. So too do energy companies' PAC and district CSR giving, as shown below in Figure 3. Note that the Congress in which energy CSR jumped the most—the 110th Congress—is also when the biggest anti-environmental shift in all three measures of behavior occurs. Figure IV.A in the appendix also documents this phenomenon for energy PFD giving and the fraction of total PFD giving done by energy companies. These graphs taken together speak to the prudence of controlling for time fixed effects when conducting inference.

Figure 3: PAC and district CSR giving over time




## V. Methodology

### A. *The Effect of Senator Pivotality on Corporate Philanthropic and PAC Contributions*

This section explains the methodology used to investigate the relationship between a senator belonging to a pivot region and the amount of corporate philanthropic contributions to charities with which the senator is personally affiliated, and the extent to which it resembles PAC contributions. First, there is a straightforward theoretical argument that the pivotal measure of politician attractiveness is exogenous, which, if true, would imply that estimates from straightforward, cross-sectional regressions lend themselves to causal interpretations. Because pivotality is assigned based on a politician's position relative to those of other politicians on an ideological spectrum, whether a given senator is pivotal depends more on the preferences of the 99 *other* senators than it does on the specific preferences of the senator in question.

Nonetheless, counterarguments to exogeneity give us reason to include controls. One is that senators potentially choose their pivotality status strategically, and if factors relevant to these strategic concerns correlate with PAC or philanthropic contributions, then omitting controls would bias estimates. For instance, it might be that newcomers to the Senate tend to vote in a way such that they are not pivotal (perhaps they are less secure in their incumbency status than senior senators and therefore seek to avoid controversy or ire), and that a politician's seniority is related to their ability to attract PAC contributions or corporate donations to charities they are personally affiliated with. Additionally, political party might correlate with both corporate contributions and whether a senator falls in certain pivotal regions. To alleviate these concerns, we control for chamber seniority and political party. We estimate two primary specifications. The first is a straight regression, accounting for omitted variables which vary over senatorships but are constant across time by controlling for entity fixed effects, and is represented by the following equation:

$$\$\text{Contribution}_{i,t} = \beta_0 + \beta_1 \mathbf{1}_{\text{veto}_{i,t-1}} + \beta_2 \mathbf{1}_{\text{median}_{i,t-1}} + \beta_3 \mathbf{1}_{\text{filibuster}_{i,t-1}} + \boldsymbol{X'_{it}\gamma_{i,t}} + \alpha_i + u_{i,t}$$

$\$\text{Contribution}_{i,t}$ denotes either PAC contributions or philanthropic contributions to charities personally linked to the politician occupying senatorship $i$ in Congress $t$. Regressions will be run with the two different outcome variables to investigate whether they are sensitive in similar ways to the pivotal incentives. The lagged dummy variables indicate whether the politician occupying senatorship $i$ was in either a veto, median, or filibuster pivotal region in Congress $t-1$, as ideological scores are computed at the end of each Congress based on all Roll Call votes and so directing contributions accordingly should be expected to occur in the next period.[4] Note that the baseline category which pivotal regression coefficients will be interpreted in reference to are senators that are not pivotal in any of the three senses. Three regressions will be run with pivots determined by the time-varying measures of economic, social, and environmental ideology. When using environmental ideology, the outcome variable will be restricted to contributions from energy companies. We expect that economic pivotality will outperform social pivotality, given that corporations care more about economic issues. $\boldsymbol{X'_{it}}$ is a control vector including chamber seniority and an indicator variable for being a Republican so that the omitted baseline category is Democrat (plus a handful of observations for independent senators). $u_{i,t}$ is the error term which will be clustered by senatorships to account for correlation between observations across a given senatorship over time.

The second specification is precisely the same as the first but uses time lifetime, time invariant scores and does not control for entity fixed effects. Note that it is not necessary to include time fixed effects in any specification as there are no variables that vary over time but not across senatorships which can be correlated with pivotality status, given that pivotality is limited only to a small subset of senators each Congress that is fixed in size, precluding the possibility of Congresses in which everyone has an equally higher or lower likelihood of being pivotal. Finally, we report, as robustness checks in the appendix, regressions using other combinations of score selection (time-varying versus lifetime), lag choice (0 or 1), and fixed effects choice (yes or no).

*B. The Effect of Energy Corporations' Philanthropic and PAC contributions on House Representatives' Behavior Towards the Environment*

We use two different methodologies to investigate the second question of the paper about the causal effect of philanthropic (and PAC, for comparison) contributions on Roll Call voting on environmental bills and floor speeches on climate change in the House of Representatives. The main concern about causal estimates is that of bias from reverse causality: does the investment hypothesis or influence hypothesis prevail? Our first methodology seeks to attenuate this concern by accounting for temporal precedence, while our second uses two instruments which are related to contributions but which we argue are unrelated to politician behavior on the environment.

Our first methodology involves running two-way-fixed effects models—a 'typical' one as well as one in which one lead and lag of contributions are included to investigate dynamic effects and potential simultaneous causality. If the lag is significant, it implies that contributions in $\text{congress}_t$ lead to a change in representative behavior in $\text{congress}_{t+1}$, consistent with the influence hypothesis. Alternatively, if leads are significant, it implies that a representative exhibiting certain behavior in $\text{congress}_t$ predicts donations in $\text{congress}_{t+1}$, consistent with the investment hypothesis. This empirical specification is the following:

$$Y_{i,t} = \alpha_0 + \sum_{j=-1}^{1} \phi \$\text{Contribution}_{i,t+j} + \gamma_i + \delta_t + \epsilon_{it}$$

$Y_{i,t}$ is the outcome variable measuring the politician activity in congressional seat $i$ in congress $t$. This will either be the fraction of environmental Roll Call votes during the congress in which they took the pro-environmental position, or the proportion of their floor speeches on climate change devoted to anti-environmental topics. $\$\text{Contribution}_{i,t}$ will either be philanthropic donations to charities located in the district represented by congressional seat $i$, donations to the House member's PFD charities[5], or PAC contributions in Congress $t$. $\gamma_i$ and $\delta_t$ are congressional seat and congress fixed-effects, respectively. $\epsilon_{it}$ is the error term which will be clustered by congressional seat. We choose to include only one lag and lead because it is the least we can include to investigate temporal precedence. Each Congress is 2 years long, implying that including the lead, the lagged, and the present value spans 6 years. It seems unlikely that a contribution of any time will operate on a lag or lead of more than a few years, and to include more lags or leads would require dropping valuable observations. Lastly, I will also note when we find relationships which, despite falling short on strongly supporting a causal claim vis-à-vis the effect of contributions on representatives' behavior, seem to provide or fail to provide suggestive evidence that corporate philanthropy is being used for political influence.

However, if this analysis finds evidence that simultaneous causality is a problem, then our first methodology is limited in its ability to provide credible causal inference. This motivates our second methodology, which is to estimate a two-stage-least-squares two-way fixed effects model in

[4]The downside of lagging pivotal variables is that it waters down estimates of the effect of pivotality for observations in which a Senate seat has just been filled by a new senator. We act on the assumption that the theoretical argument for lagging pivotal measures outweighs this empirical downside, but also report regression results in the appendix whose pivotal measures are not lagged.

[5]The relevant difference between PFD and district CSR giving is that the potential "gearing" between philanthropic contributions and legislator behavior is probably tighter in the former, as politicians are probably more invested in their own charities than in charities in their district.

which PAC contributions and philanthropic contributions to charities are the endogenous variables, and the two instruments are exogenous measures of the PAC and philanthropic charitability of the set of companies which donated to the representative in seat $i$ in congress $t$. The "PAC instrument" is computed as the average amount of PAC contributions that the set of energy companies which gives to the representative in seat $i$ in congress $t$ give to all other representatives except for $i$ in congress $t$. Similarly, the "CSR instrument" is the average amount of philanthropic contributions that the oil companies which give to charities in seat $i$'s district in congress $t$ donate to all other districts except for seat $i$'s in congress $t$. For example, if the only energy corporations in our sample which donate during a given congress to Representative A are Company X, which donated $\$10$ to Rep. A, $\$10$ to Rep. B, and $\$20$ to Rep. C, and Company Y, which donated $\$10$ to Rep. A and $\$20$ to Rep. B., then the instrumental variable for Rep. A would be $\frac{1}{2}(\$30 + \$20) = \$25$.

Theoretically, there is good reason to think that the instruments are relevant. If the companies donating to a given representative are more generous on average during a given congress, then they are more likely to be generous to the representative at hand. Indeed, regressing energy PAC contributions on the two instruments and conducting an F-test for joint significance yields a Chi-squared statistic of $160.2$ ($p < 0.001$). When energy CSR contribution amount is the regressand, the statistic is $44.2$ ($p < 0.001$). And yet their computations, which involve all receiving politicians but $i$, make them likely exogenous insofar as they are unrelated to the conditions of district $i$ or its representative. Moreover, averaging, rather than summing, over the contributing companies makes the instruments unrelated to the total number of energy companies donating to the politician, which could be related to their actions on environmental issues.[6] Because we are exactly identified, we cannot perform formal tests for exogeneity. To further increase estimate plausibility, we control for time and fixed effects.

The first stage $(1.1, 1.2)$ and structural equations $(2)$, then, are:

$$\$\text{PAC}_{it} = \pi_0 + \pi_1 Z_{\text{PAC}_{it}} + \pi_2 Z_{\text{CSR}_{it}} + \rho_i + \tau_{it} + v_{it} \quad (1.1)$$

$$\$\text{CSR}_{it} = \xi_0 + \xi_1 Z_{\text{PAC}_{it}} + \xi_2 Z_{\text{CSR}_{it}} + \mu_i + \delta_t + u_{it} \quad (1.2)$$

$$Y_{it} = \beta_0 + \beta_1 \$\text{PAC}_{it} + \beta_2 \$\text{CSR}_{it} + \kappa_i + \lambda_t + e_{it} \quad (2)$$

$Y_{it}$ is the speech or Roll Call outcome variable measuring the politician activity in congressional seat $i$ in congress $t$. $\$\text{CSR}_{it}$ and $\$\text{PAC}_{it}$ are corporate philanthropic (to charities in the congressional district) and PAC contributions, respectively. $Z_{\text{CSR}_{it}}$ and $Z_{\text{PAC}_{it}}$ are the two instruments just explained. $\rho_i$, $\mu_i$, and $\kappa_i$ are congressional-seat or district fixed effects while $\tau_{it}$, $\delta_t$, and $\lambda_t$ are time-fixed effects. $v_{it}$, $u_{it}$, and $e_{it}$ are errors that will be clustered by congressional districts.

## VI. Results

### A. The Effect of Senator Pivotality on Corporate Philanthropic and PAC contributions

Table 4 presents results from regressions without fixed effects comparing the impact of (lagged) senator pivotality as measured by lifetime scores on philanthropic and PAC giving, while Table 5 reports regressions which use time varying ideology scores and include fixed effects.[7] Estimates and standard errors are reported in units of $1,000s. We first notice that coefficient signs are negative and positive for both PAC and CSR contributions across different pivotality measures, implying that neither PAC nor CSR giving universally increases with all types of pivotality, and that we therefore cannot make the straightforward comparison to PAC giving we initially intended to learn whether CSR follows political incentives. We also see that for only 10 out of the 18 pivot-ideology combinations reported in both tables do the PAC and CSR coefficients have the same sign. When we exclude columns for environmental pivots and energy contributions, however, this fraction becomes $\frac{8}{12}$. We therefore observe some weak evidence of similarity between the directions of the sensitivity of total PAC and CSR giving for economic and social pivots (1), whereas energy PAC and CSR giving differ sharply in response to environmental pivots (2). Let us hone in on these two separate cases.

**Table 4: Regressions using Lifetime Scores without Fixed Effects**

| | *Economic Ideology* | | *Social Ideology* | | *Environmental Ideology* | |
|---|---|---|---|---|---|---|
| $Contributions | $CSR | $PAC | $CSR | $PAC | $CSR (*energy*) | $PAC (*energy*) |
| Veto Pivot (0-1) | 140.5 | 37.2 | 676.2 | 178.5 | 12.7 | -20.5** |
| | (747.1) | (132.6) | (810.6) | (148.6) | (47.2) | (10.2) |
| Median Pivot (0-1) | 397.8 | -7.9 | -526.1*** | -175.8 | -29.6 | -5.6 |
| | (431.7) | (145.4) | (187.4) | (115.7) | (28.3) | (15.1) |
| Filibuster Pivot (0-1) | 396.9 | 376.1** | -186.2 | 71.0 | 403.6 | 19.1 |
| | (446.9) | (157.5) | (438.7) | (154.4) | (298.6) | (26.2) |
| Republican (0-1) | 552.0 | 78.8 | 550.4 | 76.8 | 73.9 | 27.1*** |
| | (370.2) | (63.4) | (378.3) | (63.4) | (53.2) | (13.1) |
| Seniority | 42.2* | -3.0 | 42.1* | -2.8 | -1.1 | 0.0 |
| | (22.1) | (3.2) | (22.5) | (3.1) | (3.7) | (0.6) |
| Constant | -47.4 | 609.1*** | -10.9 | 625.9*** | 37.2 | 27.1*** |
| | (312.5) | (64.7) | (294.0) | (69.3) | (45.0) | (7.3) |
| Observations | 379 | 650 | 379 | 650 | 332 | 665 |
| R-squared | 0.058 | 0.013 | 0.056 | 0.013 | 0.015 | 0.028 |

**Table 5: Regressions using Time-varying scores, with Fixed Effects**

| | *Economic Ideology* | | *Social Ideology* | | *Environmental Ideology* | |
|---|---|---|---|---|---|---|
| $Contributions | $CSR | $PAC | $CSR | $PAC | $CSR (*energy*) | $PAC (*energy*) |
| Veto Pivot (0-1) | 156.3 | 7.731 | -729.8** | 116.7 | 15.05 | -19.07** |
| | (450.4) | (132.5) | (322.2) | (164.9) | (25.09) | (7.635) |
| Median Pivot (0-1) | 18.07 | 258.2 | -174.3 | 76.75 | -40.58 | 63.25 |
| | (133.8) | (194.3) | (269.1) | (143.7) | (34.57) | (47.30) |
| Filibuster Pivot (0-1) | 348.6 | 476.5*** | 130.0 | 125.9 | 5.143 | -3.643 |
| | (824.0) | (160.5) | (479.8) | (171.4) | (74.82) | (12.22) |
| Republican (0-1) | 352.3 | 78.44 | 285.7 | 83.14 | 76.46 | 44.52*** |
| | (358.3) | (73.12) | (341.6) | (75.92) | (54.88) | (14.20) |
| Seniority | 54.95** | -2.999 | 57.51** | -4.023 | 2.644 | 0.0913 |
| | (25.01) | (3.690) | (24.78) | (3.625) | (4.659) | (0.542) |
| Constant | 175.9 | 630.0*** | 251.3 | 668.9*** | 76.46 | 20.60* |
| | (307.7) | (79.54) | (347.2) | (84.35) | (88.64) | (12.40) |
| Observations | 379 | 650 | 379 | 650 | 332 | 665 |
| R-squared | 0.058 | 0.013 | 0.056 | 0.013 | 0.015 | 0.028 |

Robust standard errors clustered by senatorships in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Units: $1000s

[6] A potential counterargument to our exogeneity reasoning might be that a representative receives contributions from companies which tend to give a lot to others may be related to that politician's behavior. We assume this to be a second order concern.

[7] Note that a small $R^2$ is not a problem given all but one of our covariates is an indicator variable.

For total corporate contributions and economic and social pivots, we find important differences in the general trends of PAC and CSR giving. First, PAC contributions seem to increase significantly with economic pivotality and might insignificantly increase with social pivotality. While 5 of 6 economic pivot coefficients reported in both tables are positive, we find that the economic filibuster is the most important. The first specification estimates that economic filibuster pivots receive \$376,000 ($p < 0.05$) more than non-economically pivotal senators in a given Congress, while the second specification estimate is \$476,500 ($p < 0.01$). These numbers are economically significant given that the average PAC money a Senator receives per Congress is \$630,000. Social pivotality, on the other hand, seems to matter less for PAC giving; 5 of 6 coefficients are positive, but none are significant. This difference is intuitive given that corporations likely care about economic issues (e.g., taxation) more so than they do about social issues.

When we look at CSR, however, the story is reversed. While being economically pivotal may increase CSR contributions (5 of 6 coefficients positive, none are significant at 10% level), the main finding is that being socially pivotal seems to decrease CSR giving; 4 of 6 coefficients are negative, with the first specification estimating that the social median pivot receives \$526,000 ($p < 0.01$) less than socially non-pivotal senators, and the second specification estimating that the social veto pivot receives \$730,000 ($p < 0.05$) less than socially non-pivotal senators. Despite not amounting to evidence of corporate philanthropy usage for political influence, it provides evidence that firms are likely highly aware of the political careers of the receiving charity's board members. This high awareness is more characteristic of the strategic CSR motive than the delegated philanthropy or insider initiated and therefore suggests that the former may play a significant role. Moreover, as social issues divide American opinion more than economic issues (Saad, 2021) it does not seem unreasonable to assume that socially pivotal senators may also tend to be more controversial (e.g., Joe Manchin). This may explain why corporations driven most by the CSR strategic motive potentially donate less to charities whose board members are socially pivotal senators, as such a donation would result in the opposite of its intended effect of placating public opinion and making the firm palatable to socially conscious investors, employees, and consumers. Therefore, despite falling short of evidence that firms use philanthropy for political influence, these results are consistent with the strategic CSR motive playing a significant role in corporate CSR decisions.

Regarding environmental pivots and energy contributions, interpretations are less clear. This discrepancy fails to evidence the importance of the Gray and Jenkins (2020) insight. While coefficient signs and magnitudes vary across specification and kind of contribution, one constant effect is that environmental veto pivots receive \$19,000 to \$20,500 ($p < 0.05$) less PAC money than non-environmental pivots. Another point of suggestive evidence is that ceteris paribus Republicans, who are known to be less pro-environmental, receive statistically significantly ($p < 0.01$) more energy PAC money (\$27,000 to \$45,000 depending on the specification) than Democrats or Incumbents, while they do not get statistically significantly more energy PFD charitable donations, although the economic estimate is large (\$74,500 – \$76,500). This suggests that CSR money may not be used in the same way as PAC money by energy corporations.

### *B. The Effect of Energy Corporations' Philanthropic and PAC contributions on Behavior Towards the Environment in the House*

This section presents results on how CSR district, CSR PFD, and PAC contributions affect voting on environmental issues and rhetoric on climate change in the House of Representatives. We present first the results of the 2FWE methodology investigating temporal precedence and then the results of the instrumental variable analysis. Table 6 reports regression outputs of environmental Roll Call voting on energy CSR donations. The first column of the left hand panel shows that when we control just for time fixed effects, we find a negative relationship implying that a \$1 million in energy CSR contributions to a representatives' PFD charity corresponds to a 4.1 percentage point decrease in their pro-environmental voting rate ($p < 0.01$). While claiming causality is not credible, this does tell us that corporations are systematically donating more to the personal charities of representatives who vote against the environment, and may be interpreted as suggestive evidence vis-à-vis our first question, given that omitted variables are less likely to bias estimates for PFD charity donations than they are for district charities.

However, both statistical and economic significance do not survive the inclusion of district fixed effects or of a lag and lead but retain their negative sign. The coefficient on the lead is negative while the lag is positive, but as both are of such negligible magnitudes and weak statistical significance, we are not informed about dynamics. Counterintuitively, all coefficients for district CSR giving are positive, which may be the result of omitted variable bias from factors varying over time and across district; such biases may be more prevalent at the district level than for PFD giving, given that the former is likely more affected by macro-conditions. Controlling only for time fixed effects implies that a \$1 million increase in district energy CSR donations would lead to a 3.7 percentage point increase ($p < 0.01$) in pro-environmental voting rate. Including district fixed effects makes this estimate 0.71 percentage points ($p < 0.1$). After including a lag and lead all estimates are of negligible magnitudes and significances, again not informing us of dynamics.

**Table 6: Pro-environmental Voting Rates and $CSR**

| **Proportion Pro-Env. Votes** | *Charities of Personal Affiliation* — *Basic Reg. w/ Congress dummies* | *2WFE* | *2WFE w/ lag & lead* | *Charities in Geographic District* — *Basic Reg. w/ Congress dummies* | *2WFE* | *2WFE w/ lag & lead* |
|---|---|---|---|---|---|---|
| $Contribution$_t$ | -4.10e-08*** | -1.61e-09 | -8.34e-10 | 3.69e-08*** | 7.05e-09* | 2.69e-09 |
| | (1.40e-08) | (1.20e-08) | (2.73e-09) | (9.77e-09) | (3.61e-09) | (2.81e-09) |
| $Contribution$_{t+1}$ | | | -5.49e-09 | | | 2.50e-09 |
| | | | (5.65e-09) | | | (1.93e-09) |
| $Contribution$_{t-1}$ | | | 3.84e-09 | | | 8.92e-09 |
| | | | (3.57e-09) | | | (5.57e-09) |
| Constant | 0.557*** | 0.512*** | 0.564*** | 0.480*** | 0.484*** | 0.488*** |
| | (0.0333) | (0.0265) | (0.0273) | (0.0183) | (0.0179) | (0.0187) |
| Observations | 1,348 | 1,348 | 781 | 3,387 | 3,387 | 2,451 |
| R-squared | 0.024 | | | 0.032 | | |
| # of cong. offices | | 285 | 199 | | 459 | 434 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7, which repeats Table 6 but for energy PAC contributions only, provides stronger results. Controlling just for time fixed effects, a mere $100,000 increase in energy PAC donation corresponds to a 70.2 percentage point decrease in pro-environmental voting rate ($t = -14.99$). Controlling for 2WFE changes that estimate to a 35 percentage point decrease ($t = -7.36$). After including a lag and lead, the estimate drops to a 29 percentage point decrease, still significant at the 1% level. Moreover, while the lag coefficient is positive, small (1.2 percentage points), and insignificant statistically, the coefficient on the lead is negative, relatively big (7.3 percentage points), and significant at the 5% level. This implies that an anti-environmental voting record in this Congress predicts energy PAC contributions in the next, presenting evidence consistent with Goldberg et al. (2020)'s findings and the investment hypothesis rather than the influence hypothesis for PAC giving.

**Table 7: Pro-environmental Voting Rates and $PAC**

| Proportion Pro-Env. Votes | *Basic Reg. w/ Cong. dummies* | *2WFE* | *2WFE w/ lag & lead* |
|---|---|---|---|
| $PAC$_t$ | -7.46e-06*** | -3.49e-06*** | -2.86e-06*** |
| | (4e-07) | (4.74e-07) | (5.56e-07) |
| $PAC$_{t+1}$ | | | -7.32e-07** |
| | | | (3.53e-07) |
| $PAC$_{t-1}$ | | | 1.22e-07 |
| | | | (5.99e-07) |
| Constant | 0.587*** | 0.505*** | 0.519*** |
| | (0.00888) | (0.0179) | (0.0193) |
| Observations | 3,387 | 3,387 | 2,451 |
| R-squared | 0.079 | | |
| # of cong. offices | | 459 | 434 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Tables 8 and 9 repeat Tables 6 and 7 for the outcome variable recording the average extent to which a representative's speeches contain topics in the anti-environmental category. Table 8 reports mixed and counterintuitive results; in the lag and lead specifications, the (insignificant) coefficient on PFD energy contributions implies that a $1 million increase in energy PFD contributions corresponds to a 4.8 percentage point increase in anti-environmental speech prevalence, while for the district CSR regression that number is 0.15 percentage points (also insignificant). The only significant coefficient ($p < 0.05$) is that on the lag of PFD contributions, which counterintuitively implies that a $1 million increase in energy PFD contributions in this Congress leads to a 20.1 percentage point decrease in anti-environmental topic prevalence in the next Congress. While the timing of this effect is consistent with the influence hypothesis, its direction is not. Next, Table 9 shows that energy PAC contributions correspond to more anti-environmental speech. The straightforward 2WFE regression implies that a mere $100,000 increase in energy PAC contributions corresponds to a 19.2 percentage point increase in anti-environmental speech prevalence ($p < 0.01$). After including a lag and lead, the estimate is 33.2 percentage point increase, but now only significant at the 10% level. Interestingly, for speech we have the opposite result regarding PAC contributions: now, the lag coefficient ($p = 0.15$) outperforms the lead ($p = 0.92$) in significance and size (29 versus 1.9 percentage point increase).

**Table 8: Anti-environmental Rhetoric Prevalence and $CSR**

| Proportion of Speech Anti-Env. | *Charities of Personal Affiliation* — *Basic Reg. w/ Congress dummies* | *2WFE* | *2WFE w/ lag & lead* | *Charities in Geographic District* — *Basic Reg. w/ Congress dummies* | *2WFE* | *2WFE w/ lag & lead* |
|---|---|---|---|---|---|---|
| $Contribution$_t$ | 1.12e-07 | 4.03e-08 | 4.76e-08 | 1.35e-09 | 1.46e-09 | 9.77e-10 |
| | (7.19e-08) | (1.04e-07) | (3.31e-07) | (6.18e-09) | (4.54e-09) | (1.61e-08) |
| $Contribution$_{t+1}$ | | | 6.59e-09 | | | 6.12e-09 |
| | | | (1.25e-07) | | | (3.12e-08) |
| $Contribution$_{t-1}$ | | | -2.01e-07** | | | -1.76e-08 |
| | | | (7.84e-08) | | | (2.94e-08) |
| Constant | 0.308*** | 0.256*** | 0.302*** | 0.352*** | 0.283*** | 0.366*** |
| | (0.0167) | (0.0633) | (0.0392) | (0.0117) | (0.0429) | (0.0473) |
| Observations | 181 | 181 | 117 | 431 | 431 | 55 |
| R-squared | 0.026 | | | 0.000 | | |
| # of cong. offices | | 95 | 65 | | 231 | 33 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 9: Anti-environmental Rhetoric Prevalence and $PAC**

| Proportion of Speech Anti-Env. | *Basic Reg. w/ Cong. dummies* | *2WFE* | *2WFE w/ lag & lead* |
|---|---|---|---|
| $PAC$_t$ | 3.52e-06*** | 1.92e-06*** | 3.32e-06* |
| | (7.07e-07) | (6.72e-07) | (2.01e-06) |
| $PAC$_{t+1}$ | | | 1.88e-07 |
| | | | (1.86e-06) |
| $PAC$_{t-1}$ | | | 2.91e-06 |
| | | | (2.01e-06) |
| Constant | 0.306*** | 0.265*** | 0.307*** |
| | (0.0130) | (0.0433) | (0.0463) |
| Observations | 431 | 431 | 55 |
| R-squared | 0.082 | | |
| # of cong. offices | | 231 | 33 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Finally, Table 10 presents the results of the 2WFE-2SL2 IV analysis for energy PAC and CSR district giving on voting and speech variables. If our IV assumptions hold, our results imply that a $100,000 increase in energy PAC contributions decreases a representative's pro-environmental voting rate by an outsized 36 percentage points ($z = -7.40$). On the other hand, we find that a $1,000,000 increase in CSR district contributions increases the pro-environmental voting rate by 0.6 percentage points, marginally significant at the 10% level, an estimate with a confusing sign, but incredibly small magnitude. For the speech variables, PAC estimates dominate again. A $100,000 increase in energy PAC contributions corresponds to a 20.4 percentage point increase in the anti-environmental speech rate ($p = 0.002$), 29.3 percentage point decrease in the pro-environmental speech rate ($p = 0.000$), and a noteworthy 7.4 percentage point increase in the neutral speech rate ($p = 0.06$). As for voting, note that the estimated amount of energy PAC money it takes to considerably change a representative's behavior on environmental issues seems unreasonably small. This should either caveat our methodology or point to the persistence of Tullock's Puzzle, or both. CSR estimates, alternatively, despite having the same signs as PAC estimates, are far smaller and insignificant: the effect from a \$1 million increase in energy CSR contributions on the anti-environmental, pro-environmental, and neutral speech categories in percentage points is 0.25 ($p = 0.56$), -0.31 ($p = 0.43$), and 0.036 ($p = 0.86$), respectively.

**Table 10: 2SLS-2WFE Regressions**

| | Pro-Env. Vote Fraction | Floor Speeches | | |
|---|---|---|---|---|
| | | *Proportion Anti-Env.* | *Proportion Pro-Env.* | *Proportion Neutral* |
| $PAC | -3.62e-06*** | 2.04e-06*** | -2.93e-06*** | 7.44e-07* |
| | (4.89e-07) | (6.66e-07) | (5.66e-07) | (3.98e-07) |
| $CSR | 6.12e-09* | 2.51e-09 | -3.09e-09 | 3.56e-10 |
| | (3.54e-09) | (4.25e-09) | (3.91e-09) | (2.05e-09) |
| Constant | 0.505*** | 0.263*** | 0.356*** | 0.382*** |
| | (0.0179) | (0.0429) | (0.0422) | (0.0368) |
| Observations | 3,387 | 431 | 431 | 431 |
| # of districts | 459 | 231 | 231 | 231 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## VII. Conclusion

In this paper we explore (I) whether large corporations use charitable giving strategically to influence politics and (II) the extent to which this philanthropy might be effective in changing legislator behavior. We fashion from the PPT a measure of senator importance we called 'pivotality,' and investigate whether economic or social pivotality increased total CSR and PAC giving. We replicate the same for environmental pivots and energy company contributions. Although the measures fall short in predicting PAC and CSR giving overall, and especially fail in the environmental/energy contexts (thus failing to evidence Gray and Jenkin (2017)'s argument), we find statistically significant evidence that PAC giving responded to economic pivotality, particularly for the filibuster, and is less affected by social pivotality, as expected given that corporations care more about economic issues. For CSR trends, we find insignificant increases in giving towards economic pivots; had these coefficients been significant, they would have provided stronger evidence of political CSR usage. But the most noteworthy CSR finding is a large, statistically significant decrease in giving for social pivots, consistent with the explanation that the strategic CSR motive plays an important role in guiding corporate philanthropic trends.

For our second question we implement two methodologies: a 2WFE regression which accounts for temporal precedence as well as 2SLS instrumental variable approach in which 2WFE are controlled for. Conducting the first methodology, we find little to weak evidence that CSR giving impacts voting or pro-environmental speech, as most lag, current, and lead coefficients are of small magnitude and significance in the most inclusive specification. Energy PAC contributions are found to have a much more robust "anti-environmental effect." Also, for energy PAC contributions, the data support the influence hypothesis for Roll Call voting while weakly supporting the investment hypothesis for climate change rhetoric, an interesting set of results. Some of the results of this first section also speak to our first question, but in mixed ways. On one hand, contributions to PFD charities are significantly negatively related to pro-environmental voting controlling for time fixed effects. On the other hand, district CSR giving is positively related with pro-environmental voting. It is arguable that the former result is more informative with respect to the first question, as the potential for omitted variable bias seems to pose more of a threat for congressional district CSR giving, which is presumably affected more by macro-conditions than a representative's personal charity. Finally, the instrumental variable analysis implies that CSR contributions have a very small, positive, statistically significant causal effect on pro-environmental voting, a counterintuitive result. Its impact on speech is negligible but has intuitive signs. Yet, energy PAC contributions significantly decrease pro-environmental voting and shift speech not just from pro-environmental to anti-environmental but also from pro-environmental to neutral. However, the estimated amount of energy PAC money it takes to drastically change a representative's behavior on environmental issues seemed unreasonably small, a result which should either caveat our methodology or harken to the persistence of Tullock's Puzzle, or both.

## REFERENCES


[1] Ansolabehere, S., de Figueiredo, J., Snyder, J. (2003). Why is there so little money in politics? https://doi.org/10.3386/w9409

[2] Baron, David P. 2001. "Private Politics, Corporate Social Responsibility, and Integrated Strategy." Journal of Economics Management Strategy 10 (1): 7–45. Benabou, Roland, and Jean Tirole. 2010. "Individual and Corporate Social Responsibility." Econom- ica 77 (305): 1–19.

[3] Bertrand, Marianne, Matilde Bombardini, Raymond Fisman, and Francesco Trebbi. (2020). Tax-Exempt Lobbying: Corporate Philanthropy as a Tool for Political Influence. American Economic Review, 110 (7): 2065-2102.

[4] Boche, A., Lewis, J.B., Rudkin, A. et al. The new Voteview.com: preserving and continuing Keith Poole's infrastructure for scholars, students and observers of Congress. Public Choice 176, 17–32 (2018). https://doi.org/10.1007/s11127-018-0546-0

[5] Brulle, R.J. The climate lobby: a sectoral analysis of lobbying spending on climate change in the USA, 2000 to 2016. Climatic Change 149, 289–303 (2018). https://doi-org.ezproxy.princeton.edu/10.1007/s10584-018-2241-z

[6] Charitable contribution deductions. (n.d.). Retrieved December 07, 2021, from https://www.irs.gov/charities-non-profits/charitable-organizations/charitable-contribution-deductions

[7] Check out the 2020 National Environmental Scorecard. (n.d.). Retrieved December 03, 2021, from https://scorecard.lcv.org/

[8] Chiou, F.-Y. and Rothenberg, L. S. (2003) When Pivotal Politics Meets Partisan Politics. American Journal of Political Science 47(3): 503–522

[9] Clinton, J. D. (2007) Lawmaking and Roll Calls. Journal of Politics 69(2): 457–469

[10] Collingwood, L., Wilkerson, J. (2012). Tradeoffs in accuracy and efficiency in supervised learning methods. Journal of Information Technology Politics, 9(3), 298–318. https://doi.org/10.1080/19331681.2012.669191

[11] Covington, C. R. and Bargen, A. A. (2004) Comparing Floor-Dominated and Party-Dominated Explanations of Policy Change in the House of Representatives. Journal of Politics 66(4): 1069–1088. Downie, C. (2017). Business actors, political resistance, and strategies for policymakers. Energy Policy, 108, 583-592. doi:10.1016/j.enpol.2017.06.018

[12] Enke, Benjamin. "Moral values and voting." Journal of Political Economy 128.10 (2020): 3679-3729.

[13] ExxonMobil lobbyists filmed saying oil giant's support for carbon tax a PR ploy. (2021, June 30). Retrieved December 05, 2021, from https://www.theguardian.com/us-news/2021/jun/30/exxonmobil-lobbyists-oil-giant-carbon-tax-pr-ploy

[14] Fisman, Raymond. (2020, December 23). Spotting the political calculus behind some acts of Corporate Charity. Retrieved December 06, 2021, from https://theconversation.com/spotting-the-political-calculus-behind-some-acts-of-corporate-charity-95743

[15] Ferguson, Thomas. 1995. Golden Rule: The Investment Theory of Party Competition and the Logic of Money- Driven Political Systems. Chicago, IL: University of Chicago Press.

[16] Gilens, M., Page, B. I. (2014). Testing theories of American Politics: Elites, interest groups, and average citizens. Perspectives on Politics, 12(3), 564–581. https://doi.org/10.1017/s1537592714001595

[17] Goldberg, M. H., Marlon, J. R., Wang, X., Van der Linden, S., Leiserowitz, A. (2020). Oil and gas companies invest in legislators that vote against the environment. Proceedings of the National Academy of Sciences, 117(10), 5111-5112. doi:10.1073/pnas.1922175117

[18] Gray, T. R., Jenkins, J. A. (2017). Pivotal politics and the ideological content of landmark laws. Journal of Public Policy, 39(1), 115-142. doi:10.1017/s0143814x1700023x

[19] Hacker, Jacob S., and Paul Pierson. 2010. Winner-Take- All Politics: How Washington Made the Rich Richer—and Turned its Back on the Middle Class. New York: Simon Schuster.

[20] Hall, R. L., Wayman, F. W. (1990). Buying time: Moneyed interests and the mobilization of bias in Congressional committees. American Political Science Review, 84(3), 797-820. doi:10.2307/1962767

[21] Kalla, J. L., Broockman, D. E. (2016). Campaign Contributions Facilitate Access to Congressional Officials: A Randomized Field Experiment. American Journal of Political Science, 60(3), 545–558. http://www.jstor.org/stable/24877480

[22] Kang, K. (2015). Policy influence and private returns from lobbying in the energy sector. The Review of Economic Studies, 83(1), 269-305. doi:10.1093/restud/rdv029

[23] Krehbiel, K. (1998). Pivotal politics: A theory of U.S. lawmaking. Chicago: Univ. of Chicago Press.

[24] Krehbiel, K., Meirowitz, A. and Woon, J. (2005) Testing Theories of Lawmaking. In Austen-

[25] Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2021). Voteview: Congressional Roll-Call Votes Database. https://voteview.com/

[26] Lindblom, Charles E. 1977. Politics and Markets. New York: Basic

[27] McConnell, Grant. 1966. Private Power and American Democracy. New York: Random House.

[28] Mixon, F.G., Crocker, C.C. Black, H.T. Pivotal power brokers: Theory and evidence on political fundraising. Public Choice 123, 477–493 (2005). https://doi.org/10.1007/s11127-005-7172-3

[29] Powell, L. W. (2014). The influence of campaign contributions on legislative policy. The Forum, 11(3). https://doi.org/10.1515/for-2013-0047

[30] Richman, J. (2011) Parties, Pivots, and Policy: The Status Quo Test. American Political Science Review 105(1): 151–165.

[31] Roscoe, D. D., Jenkins, S. (2005). A Meta-Analysis of Campaign Contributions' Impact on Roll Call Voting. Social Science Quarterly, 86(1), 52–68. http://www.jstor.org/stable/42956049

[32] Saad, L. (2021, November 20). Americans more divided on social than economic issues. Gallup.com. Retrieved April 18, 2022, from https://news.gallup.com/poll/351494/americans-divided-social-economic-issues.aspx

[33] Schlozman, Kay Lehman, Sidney Verba, and Henry E. Brady. 2012. The Unheavenly Chorus: Unequal Political Voice and the Broken Promise of American Democracy. Princeton, NJ: Princeton University Press.

[34] Smith D. and Duggan J. (eds.), Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks. Berlin: Springer-Verlag, 249–268.

[35] Snyder, James M. 1992. "Long-Term Investing in Politicians: Or, Give Early, Give Often." Journal of Law and Economics 35 (1): 15–43.

[36] Stiglitz, E. H. and Weingast, B. R. (2010) Agenda Control in Congress: Evidence from Cutpoint Estimates and Ideal Point Uncertainty. Legislative Studies Quarterly 35(2): 157–185.

[37] Tabuchi, H. (2021, June 30). In video, Exxon lobbyist describes efforts to undercut climate action. Retrieved December 04, 2021, from https://www.nytimes.com/2021/06/30/climate/exxon-greenpeace-lobbyist-video.html

[38] Tullock, Gordon. (1972). The Purchase of Politicians. Western Economic Journal 10: 354–55.

[39] Volden, C., Wiseman, A. E. (2018). Legislative Effectiveness in the United States Senate. The Journal of Politics, 80(2), 731–735. https://doi.org/10.1086/697121

[40] Wawro, G. J. and Schickler, E. (2004) Where's the Pivot? Obstruction and Lawmaking in the Pre-Cloture Senate. American Journal of Political Science 48(4): 758–774.

[41] Woon, J. and Cook, I. P. (2015) Competing Gridlock Models and Status Quo Policies. Political Analysis 23(3): 385–399

[42] Zingales, L. (2017). Towards a political theory of the firm. https://doi.org/10.3386/w23593

# VIII. Appendix

Figure IV.A
Energy PFD CSR giving over time

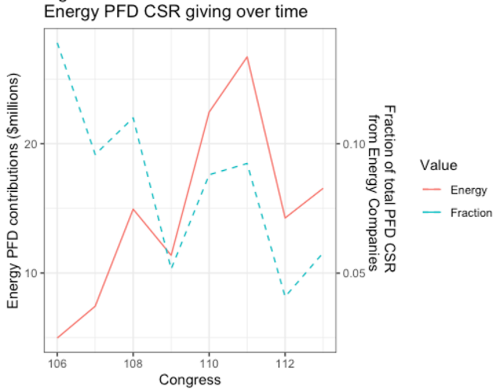



**Table A1: Time varying scores, cros sectional regression, lagged**

| | *Economic Ideology* | | *Social Ideology* | | *Environmental Ideology* | |
|---|---|---|---|---|---|---|
| $Contributions | $CSR | $PAC | $CSR | $PAC | $CSR (*energy*) | $PAC (*energy*) |
| Veto Pivot (0-1) | 880.5 | 7.3 | -593.4** | 114.1 | -6.6 | -19.1** |
| | (672.3) | (131.9) | (233.4) | (164.7) | (40.8) | (7.6) |
| Median Pivot (0-1) | 147.8 | 258.9 | -598.5** | 69.0 | -49.0 | 63.2 |
| | (379.8) | (190.3) | (292.7) | (142.5) | (29.9) | (47.3) |
| Filibuster Pivot (0-1) | 1,779.2 | 490.7*** | 34.5 | 117.9 | 155.6 | -3.6 |
| | (1,622.1) | (159.0) | (516.5) | (174.1) | (178.9) | (12.2) |
| Republican (0-1) | 593.7 | 79.8 | 621.7 | 83.2 | 85.9 | 44.5*** |
| | (409.4) | (72.8) | (426.8) | (76.1) | (61.4) | (14.2) |
| Seniority | 43.3* | -3.0 | 42.2* | -4.0 | -0.6 | 0.1 |
| | (23.0) | (3.7) | (24.4) | (3.6) | (3.6) | (0.5) |
| Constant | -60.1 | 628.9*** | 114.6 | 670.4*** | 55.1 | 20.6* |
| | (293.3) | (78.4) | (318.5) | (84.7) | (53.5) | (12.4) |
| Observations | 379 | 650 | 379 | 650 | 332 | 665 |
| R-squared | 0.058 | 0.013 | 0.056 | 0.013 | 0.015 | 0.028 |

Robust standard errors clustered by senatorships in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table A2: Lifetime scores, Cross-sectional regression, not lagged**

| | *Economic Ideology* | | *Social Ideology* | | *Environmental Ideology* | |
|---|---|---|---|---|---|---|
| $Contributions | $CSR | $PAC | $CSR | $PAC | $CSR (*energy*) | $PAC (*energy*) |
| Veto Pivot (0-1) | 140.5 | 37.2 | 676.2 | 178.5 | 12.7 | -20.5** |
| | (747.1) | (132.6) | (810.6) | (148.6) | (47.2) | (10.2) |
| Median Pivot (0-1) | 397.8 | -7.9 | -526.1*** | -175.8 | -29.6 | -5.6 |
| | (431.7) | (145.4) | (187.4) | (115.7) | (28.3) | (15.1) |
| Filibuster Pivot (0-1) | 396.9 | 376.1** | -186.2 | 71.0 | 403.6 | 19.1 |
| | (446.9) | (157.5) | (438.7) | (154.4) | (298.6) | (26.2) |
| Republican (0-1) | 552.0 | 78.8 | 550.4 | 76.8 | 73.9 | 27.1*** |
| | (370.2) | (63.4) | (378.3) | (63.4) | (53.2) | (13.1) |
| Seniority | 42.2* | -3.0 | 42.1* | -2.8 | -1.1 | 0.0 |
| | (22.1) | (3.2) | (22.5) | (3.1) | (3.7) | (0.6) |
| Constant | -47.4 | 609.1*** | -10.9 | 625.9*** | 37.2 | 27.1*** |
| | (312.5) | (64.7) | (294.0) | (69.3) | (45.0) | (7.3) |
| Observations | 379 | 650 | 379 | 650 | 332 | 665 |
| R-squared | 0.058 | 0.013 | 0.056 | 0.013 | 0.015 | 0.028 |

Robust standard errors clustered by senatorships in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table A3: Time Varying Cross sectional regression (not lagged)**

| | *Economic Ideology* | | *Social Ideology* | | *Environmental Ideology* | |
|---|---|---|---|---|---|---|
| $Contributions | $CSR | $PAC | $CSR | $PAC | $CSR (*energy*) | $PAC (*energy*) |
| Veto Pivot (0-1) | 777.0 | -49.7 | -14.6 | -72.9 | -42.1 | -7.5 |
| | (583.9) | (103.3) | (445.8) | (105.0) | (42.7) | (9.9) |
| Median Pivot (0-1) | 80.6 | 35.8 | -652.6*** | 103.7 | -48.9 | 19.7 |
| | (365.0) | (137.9) | (232.4) | (153.5) | (34.6) | (20.3) |
| Filibuster Pivot (0-1) | 874.3 | 173.8 | 516.3 | 185.0 | -3.5 | -17.4 |
| | (685.9) | (146.3) | (861.3) | (169.8) | (87.9) | (15.3) |
| Republican (0-1) | 512.5 | 75.8 | 554.1 | 85.7 | 90.4 | 46.7*** |
| | (358.5) | (65.3) | (367.3) | (66.8) | (66.0) | (12.7) |
| Seniority | 40.6* | -2.6 | 41.8* | -3.0 | -0.2 | 0.1 |
| | (21.0) | (3.2) | (22.6) | (3.2) | (3.3) | (0.5) |
| Constant | -30.1 | 620.7*** | 8.1 | 617.2*** | 62.1 | 25.1*** |
| | (298.4) | (66.5) | (283.0) | (69.1) | (56.7) | (8.0) |
| Observations | 379 | 650 | 379 | 650 | 332 | 665 |
| R-squared | 0.058 | 0.013 | 0.056 | 0.013 | 0.015 | 0.028 |

Robust standard errors clustered by senatorships in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table A4: Time Varying measures, Fixed Effects (not lagged)**

| | *Economic Ideology* | | *Social Ideology* | | *Environmental Ideology* | |
|---|---|---|---|---|---|---|
| $Contributions | $CSR | $PAC | $CSR | $PAC | $CSR (*energy*) | $PAC (*energy*) |
| Veto Pivot (0-1) | 179.0 | -49.7 | -111.9 | -72.9 | -25.0 | -7.5 |
| | (296.9) | (103.3) | (430.6) | (105.0) | (28.6) | (9.9) |
| Median Pivot (0-1) | -135.1 | 35.8 | -286.1 | 103.7 | -40.7 | 19.7 |
| | (86.53) | (137.9) | (273.6) | (153.5) | (30.7) | (20.3) |
| Filibuster Pivot (0-1) | 426.6 | 173.8 | 832.6 | 185.0 | 11.9 | -17.5 |
| | (364.2) | (146.3) | (919.1) | (169.8) | (45.95) | (15.3) |
| Republican (0-1) | 113.2 | 75.8 | 130.6 | 85.7 | 59.6 | 46.7*** |
| | (313.0) | (65.32) | (318.0) | (66.8) | (49.2) | (12.8) |
| Seniority | 64.09** | -2.6 | 63.84** | -2.98 | 2.5 | 0.09 |
| | (25.6) | (3.2) | (25.2) | (3.2) | (4.3) | (0.5) |
| Constant | 108.9 | 620.7*** | 91.9 | 617.2*** | 78.1 | 25.14*** |
| | (329.7) | (66.5) | (344.0) | (69) | (85.2) | (7.96) |
| Observations | 379 | 650 | 379 | 650 | 332 | 665 |
| Number of senatorships | 66 | 97 | 66 | 97 | 65 | 98 |

Robust standard errors clustered by senatorships in parentheses
*** p<0.01, ** p<0.05, * p<0.1

# Loss Aversion in Dictator Games

Anthony Khaiat, Camille Burton. Gabe Moos, and Kimberly Stafford

*University of Chicago*

*Abstract*— **Dictator Games consist of two players, an allocator and a recipient, in which the allocator governs the distribution of an endowment between themselves and the recipient. Widely replicated Dictator Game studies by Kahneman et al. (1986) and Kuang et al. (2006), among others, reveal that altruistic behavior among allocators can be influenced by variables that regulate models of human inequity aversion. This study presents a randomized, modified Dictator Game experiment with loss aversion and transparency manipulations. We find that loss aversion provides no significant effect on a Dictator's willingness to give, yet transparency between players nudges allocators to be more generous. Additionally, we find that individuals believe they are more altruistic than others under loss aversion coupled with transparency.**

## I. Introduction and Related Literature

Standard neoclassical economic theory relies on the assumption that individuals have stable, well-defined, and rational preferences. The classic utility function in economics captures these predicted preferences via a mathematical representation of how much utility, $U(x)$, a good or bundle of goods, provides individuals, thereby establishing a normative economic theory of human decision-making. Yet in practice, people often engage in behaviors that systematically deviate from the classic model of rational utility-maximizing behavior (Kahneman & Tversky, 1979). Behavioral economics (BE) seeks to more accurately describe human behavior by integrating standard economic models with psychological insights that account for the inherent biases and heuristics shaping real-life decision-making under risk. This paper embraces the mission of BE by asking an important question: how does the presence of loss aversion and transparency—two well-studied, robust decision-making biases in the literature—interact with standard economic models of social preferences and inequity aversion as measured by the Dictator Game?

Answering this question requires a preliminary understanding of loss aversion as established by Kahneman and Tversky's Prospect Theory: An Analysis of Decision under Risk (1979). Prospect theory is characterized by three main features: (1) sensitivity to relative changes around a reference point, (2) aversion to losses, and (3) probability weighting. Loss aversion is a result of the value function (Figure 1) being significantly steeper for losses than gains, such that individuals are expected to be risk-seeking over the loss domain and risk-averse over the gain domain. Furthermore, loss aversion imposes various implications on economics literature in addition to real-life market behavior.

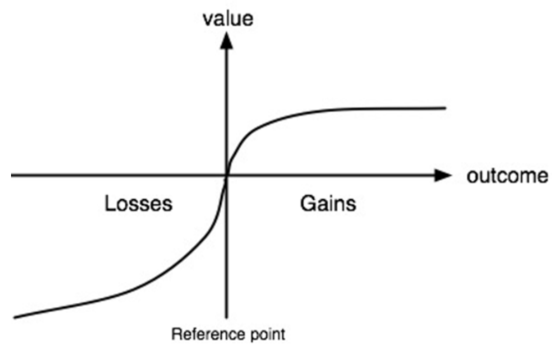



Fig. 1: Prospect Theory's Value Function

In accordance with Thaler's (1999) theory on mental accounting, experiencing a recent loss may prompt increased aversion toward risk-taking behaviors due to the larger negative hedonic impact of such losses. Mental accounting describes the cognitive organization of monetary accounts into separate mental brackets which each have its own relative reference point and marginal propensity to spend. Thaler (1999) introduces three key components of mental accounting: (1) the differential perception and experience of outcomes, (2) the assignment of monetary activities to specific accounts, and (3) the differential frequency with which these accounts are evaluated. Although Prospect Theory predicts that prior loss priming should prompt a decrease in risk-taking attitudes, the empirical literature on this topic is divided: certain studies confirm a trend of increased loss aversion (i.e. avoiding further risk following the prior loss), whereas other studies reveal loss-chasing behavior (i.e. taking on greater subsequent risk to try to regain the prior loss).

Loss aversion is highly dependent on the context of its application; the Realization Effect (Imas, 2016) provides a salient example of loss aversion's profound ability to interact with models of cognitive preferences to shape actual behavior in markets with risky assets. Imas reconciles the inconsistencies in the literature highlighted above by introducing an important distinction between paper and realized losses. In experimental settings, individuals consistently avoid risk-taking (i.e. exhibit loss aversion) after experiencing a realized loss but exhibit greater risk-seeking behavior following a paper loss of identical magnitude. Mental accounting mediates this effect because paper and realized losses receive different cognitive accounting treatments. When a loss is realized and cash leaves an individual's account, the mental account closes and the reference point resets to zero. However, the mental account remains open following a paper loss such that a subsequent high-risk/high-reward prospect gets integrated and evaluated jointly with the prior loss rather than independently.

Shifting now to the BE literature on loss aversion in the context of social preferences, we must first introduce the well-documented Ultimatum Game. In this game, two players are allotted a sum of money, in which the first player, denoted as the Proposer, offers some portion of the money to the second player, denoted as the Responder. The Responder has the ability to either accept or reject the offer. If both players are rational income maximizers as normative economic theory predicts, the Proposer should consistently offer the smallest unit of currency available, and the Responder should always accept these positive, albeit small, offers. However, experimental findings reveal this is not the case; offers typically average around 30–40% of the total allotment, with a 50-50 split often the mode, and offers less than 20% frequently rejected (Camerer & Thaler, 1995). While these findings provide some degree of evidence for models of global social preferences, the Dictator Game pushes these insights a step further by eliminating risk aversion from the experiment due to the fact that offers are unable to be rejected.

In the Dictator Game, a derivative of the classic Ultimatum Game, allocators are similarly given a fixed amount of money to distribute amongst themselves and a receiver. This amount of wealth in this asset account is typically $10. However, the receiver lacks the ability to reject the proposed offer, and must always accept the allotment of funds. The literature is saturated with empirical evidence from Dictator Game replications that manipulate a host of explanatory parameters such as demographics, framing, and subject anonymity. In a recent meta-study that includes multiple regressions of a dataset of over 100 Dictator games, Engel (2011) finds that dictators on average give 28.35% of the initial allocation and are more likely to allocate a less equal split than in the Ultimatum Game, as demonstrated by a left-skewed distribution of means. Furthermore, certain variables and demographics predictably influence the generosity of dictators: if the dictator is older or their identity becomes transparent, they tend to be more generous, while children and groups of dictators typically allocate less to the receiver (Engel, 2011).

Normative economic predictions about rational self-maximizing fail to adequately explain human behavior in these contexts: under the Dictator game framework, people consistently share part of the monetary allotment despite having rational incentives to claim the entire allocation. Moreover, is inherent altruism in the form of generosity driving these effects? Dana, Weber, and Kuang (2007) investigate this compelling question by imposing treatments on a binary choice Dictator game that grants dictators the option to introduce uncertainty into the relationship between their allocation decision and resulting outcomes. This uncertainty treatment thus bestow subjects with the moral "wiggle room" to potentially behave more in their self-interest. The results of this study find that dictators engage in significantly less generous behavior in these manipulations relative to baseline, suggesting that subjects may generally share more of their initial allocation in the classic Dictator game principally because they dislike appearing unfair to themselves or others (Dana et al., 2007). In this case, transparency was introduced in relation to the uncertainty imposed between the allocator's decision and the final distribution outcome; high uncertainty implies low transparency, while low uncertainty implies high transparency (i.e. the relationship between actions and outcomes is transparent). Given the correlation between transparency and Dictator game outcomes, we introduce a related transparency variable in our following study by manipulating asymmetric information regarding the amount of wealth that the dictator starts with. This imposed transparency may similarly influence dictator generosity by creating wiggle room for dictators.

An alternative, but somewhat related, explanation for classic Dictator game findings stems from Camerer & Thaler (1995), who attribute the cause to learned manners rather than altruism or utility functions. This view describes behavior in these decision-making contexts as the cumulative manifestation of rules of reciprocity that people learn in everyday life. In other words, people learn to consistently treat others fairly while punishing those who behave unfairly because these behaviors align with the inherent structure of long-term sociality within human societies. Over the course of many repeated encounters, these behavioral instincts become ingrained. Although the long-run benefits of reciprocity and "manners" are virtually irrelevant to the framework of one-shot Ultimatum and Dictator Games, Camerer & Thaler ascertain that subjects in such experimental settings may still implicitly rely on these repeated-game impulses out of habit.

The seminal behavioral theory to explain the social preferences illustrated by Dictator games that we will explore in this paper is Fehr and Schmidt's 1999 model of inequity aversion (see Equation 1), which rejects the classical economic notion of utility-maximizing behavior; rather, individual utility appears to decrease both when subjects receive an advantageous and disadvantageous inequitable division of resources. Importantly, loss aversion interacts with social comparison in these contexts such that individuals experience greater disutility from disadvantageous inequity (where $\alpha_i > \beta_i$) than advantageous inequity. As discussed above, empirical results of Dictator games in the literature suggest confirmation for this model, which integrates the abstract fairness concept associated with prosocial behavior (as evidenced by allocators sharing money with recipients) into the 2-player model below:

$$U_i = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\} \quad (1)$$

where $i \neq j$ and $\alpha_i \geq \beta_i$.

Now that we have developed an underlying framework to interpret the social preferences guiding the results of our Dictator game experiment, we can analyze loss aversion's interaction on these preferences with greater detail. One relevant and highly referenced variable in the literature is gain versus loss framing: perceptions of fairness strongly depend on whether a stimulus is framed as a reduction in a gain or an actual loss of identical magnitude (Kahneman, Knetsch, & Thaler, 1991). Implementing a modified Dictator Game experimental design, Boun My and colleagues (2008) applied such framing manipulations to Fehr and Schmidt's model of human equity preferences in a recent study. Their data reveal that subjects display significantly lower utility for advantageous inequality when outcomes are framed as losses than when outcomes are framed as gains. In other words, the amount of personal payoff that subjects are willing to sacrifice to grant others—a measure of advantageous inequality—is smaller under a loss frame than under a gain frame (Boun My et al., 2008). Thus, priming subjects to enter a loss frame via an imposed loss treatment may have tangible implications on their generosity in Dictator Games.

Another salient parameter shown to interact with models of inequity aversion are manipulations of the wealth assets used in Ultimatum and Dictator Games. Halvorsen (2015) recently replicated a Dictator Game experimental design in which allocators were asked to distribute some portion of their earnings to a receiver either before or after "earning" the wealth. However, the results reveal that whether the money is already earned or not has no statistically significant effect on the sharing behavior of dictators, indicating that loss aversion is not present with respect to status quo wealth levels. However, Cherry, Frykblom, & Shogren (2002) argue that an alternative Dictator Game setting in which the assets for allocation are legitimate (i.e. earned, rather than bestowed) is more likely to produce "rational" behavior that aligns more closely with normative self-maximizing models. In contrast to previous Dictator Game studies, their experimental design instructed allocators to distribute earned wealth rather than unearned wealth granted by the experimenter. Their results confirm that dictators allocating earned wealth behave more self-interestedly, and when complete anonymity of the dictator is introduced, generous behavior essentially disappears (Cherry et al., 2002).

This analysis of the existing economic literature on loss aversion, inequity aversion, and Dictator Games provides crucial insight into the psychological factors underlying human behavior within these contexts. Our present research seeks to further extend the literature by investigating the effects of loss aversion priming on the dictator's initial asset balance and transparency of initial asset information on the results of the classic Dictator Game.

## II. Methods and Hypotheses

### A. Experimental Design

Our experiment aimed to determine whether loss aversion impacts allocators' willingness to distribute and receive money. To determine the effects of these two factors, we conducted a modified version of the Dictator Game. This was done via Qualtrics using a between-subjects design. Of the 201 individuals that took part in the experiment, each was randomly assigned to one of four groups, labeled Group 1, Group 2, Group 3, and Group 4. All subjects played the role of the allocator, or "dictator", and were given a hypothetical initial balance of either $10 or $12. Half of the groups starting with an initial balance of $12 comprised the loss aversion treatment, and subsequently "lost" $2 which resulted in a $10 balance to allocate. All allocators were then asked to record how much money they would be willing to give a pseudo-receiver (i.e. a random individual). As a follow-up, participants then indicated the minimum offer they would be willing to receive if their role was flipped as the recipient. The questions displayed answer options on a slider with whole number options ranging from 0-10. In half the groups (2 & 4), the allocator was told that the receiver knew about their initial and current balances to allocate. This was an effort to promote transparency of information amongst the players of the game.

Group 1, the baseline case, emulated the classic Dictator Game, where the allocator started off with $10, and the receiver was unaware of the "size of the pie." In Group 2, we incorporated transparency into the Dictator game: the allocator started off with $10, but was told that the initial allocation was revealed to the recipient. In Group 3, labeled as the loss aversion group, allocators started with $12 but were told they lost $2. Allocators were informed that recipients were unaware of this loss and the size of the pie.

In Group 4 we combined loss aversion with transparency: allocators were told that recipients knew the allocator lost $2, and that their current balance was less than their initial balance. Table 1 further illustrates each group's breakdown.

We supplemented the text on Qualtrics with images of $10 and $1 bills [see appendix for reference] to make the experiment easier to digest and visually interesting. However, we do not think that the presence of images influenced subjects in any significant way.

### B. Hypothesis

Our understanding of loss aversion and transparency in social preferences informed several hypotheses regarding the expected results. The hypotheses are twofold, one concerning the effects of loss aversion and the other of transparency. First, we expect the allocator in Group 1 to offer a larger allocation to the receiver than in Group 3 due to loss aversion. This is because allocators in these conditions will have different reference points ($10 in Group 1 vs. $12 in Group 3) when making their decision. Thus, Group 3 allocators may feel they "deserve" a larger portion of the initial endowment since they previously suffered a loss. The inequity aversion model for a two-player case also suggests that the dictators in the loss-aversion groups will start with disadvantageous inequity because they lost part of their endowment while the receiver's initial allocation remains unchanged. Additionally, we devised one of two possible outcomes relating to transparency. Firstly, per David Camerer and Richard Thaler's Economics of Manners, transparency might have caused allocators to feel more shame if they

**Table 1. Experimental Group Design**

| | **Group 1** (Baseline) | **Group 2** (Baseline + Transparency) | **Group 3** (Loss Aversion) | **Group 4** (Loss Aversion + Transparency) |
|---|---|---|---|---|
| Allocator | Start with $10 | Start with $10 | Start with $12, lose $2 | Start with $12, lose $2 |
| Receiver | Doesn't know the size of the initial allocation | Knows allocator didn't lose $2 | Doesn't know allocator lost $2 | Knows allocator lost $2 |

were perceived as less equitable (Camerer & Thaler, 1995). This is in line with the notion that transparency increases allocators' charitability. If this is the case, we would expect that the average allocator in Group 4 may provide more to the receiver than in Group 3. Conversely, the allocator might expect the receiver to be more empathetic in receiving a smaller amount in the transparency condition after witnessing their loss, which may lead the allocator in Group 4 to bestow less to the receiver than in Group 3. This hypothesis is informed by a behavioral model from Nagel (1995) and Ho et al. (1998) based on levels of reasoning: the dictator, thinking that the receiver would be more empathetic to the dictator's revealed loss, might allocate a lower partition of the initial allocation to the receiver. Thus, we believe that level 2 reasoning would occur in this instance. We predict these two sub-hypotheses regarding transparency to be equally likely.

Combining our hypotheses, we expect that transparency coupled with loss aversion, as demonstrated with Group 4, will either amplify or reduce the dictator's willingness to give in light of a loss. That is, we may expect Group 4 subjects to offer more than Group 3 if our first hypothesis on transparency is correct. Alternatively, Group 4 subjects may offer less than Group 3 if our second hypothesis is correct. Comparing the results of Group 4 to the initial baseline game in Group 1 can also tell us whether loss aversion and transparency negate, amplify, or have no effect on each other.

## III. Results

To run our regression analysis, we regressed the dependent variables (willingness to give and willingness to receive) on two independent variables, Transparency, and Treatment (loss aversion). In our model, we took into account the main effects from both Transparency and our Loss Aversion treatment as well as adding interactions between both independent variables. By performing a two-way anova test, we are modeling the willingness to give and the willingness to receive as functions of the transparency and our treatment (loss aversion), as shown in Equations 2 and 3:

$$WTG_i = \beta_{(0,0)} + \beta_{(1,0)}X_{(1,0)i} + \beta_{(0,1)}X_{(0,1)i} + \beta_{(1,1)}X_{(1,0)i}X_{(0,1)i} + \epsilon \tag{2}$$

$$WTR_i = \beta_{(0,0)} + \beta_{(1,0)}X_{(1,0)i} + \beta_{(0,1)}X_{(0,1)i} + \beta_{(1,1)}X_{(1,0)i}X_{(0,1)i} + \epsilon \tag{3}$$

where for $i = n$ observations,

- $WTG_i$ = dependent variable for equation 2
- $WTR_i$ = dependent variable for equation 3
- $X_{(1,0)i}$ = the explanatory variable for the transparency treatment
- $X_{(0,1)i}$ = the explanatory variable for the loss aversion treatment
- $\beta_{(0,0)}$ = the y-intercept (baseline case)
- $\beta_{(1,0)}$ = slope coefficient for the transparency treatment
- $\beta_{(0,1)}$ = slope coefficient for the explanatory variable loss aversion treatment
- $\beta_{(1,1)}$ = slope coefficient for the interaction term $X_{(1,0)i}X_{(0,1)i}$
- $\epsilon$ = the model's error term (residuals)

Our regression results are presented in Figures 2 and 3 below.

```
                       Df Sum Sq Mean Sq F value Pr(>F)
Transparency            1   18.9  18.945   4.626 0.0327 *
Treatment               1    0.0   0.030   0.007 0.9319
Transparency:Treatment  1    0.3   0.281   0.069 0.7938
Residuals             197  806.8   4.095
```

Fig. 2: Willingness to Give Regression

```
                       Df Sum Sq Mean Sq F value Pr(>F)
Transparency            1   12.2  12.248   2.773 0.0974 .
Treatment               1    0.1   0.085   0.019 0.8901
Transparency:Treatment  1   20.3  20.267   4.589 0.0334 *
Residuals             197  870.0   4.416
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig. 3: Willingness to Receive Regression

Figure 4 performs a two-sided ANOVA on the difference between the willingness to give and the willingness to receive within subjects.

```
                      Df Sum Sq Mean Sq F value  Pr(>F)
Transparency           1   61.7   61.66   8.451 0.00407 **
Treatment              1    0.0    0.01   0.002 0.96533
Transparency:Treatment  1   15.8   15.78   2.163 0.14300
Residuals            197 1437.4    7.30
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Fig. 4: Difference between WTG and WTR Regression

Figures 5 to 7 compare the willingness to receive with the willingness to give for each subject in each treatment group.

| Within-Subject | Give More | Receive More | Same | % Give More | % Receive More | % Same |
|---|---|---|---|---|---|---|
| *Base* | 15 | 16 | 20 | 29% | 31% | **39%** |
| *Loss* | 19 | 9 | 20 | 40% | 19% | **42%** |
| *Transparency* | 19 | 14 | 20 | 36% | 26% | **38%** |
| *Loss + Transparency* | 26 | 9 | 20 | **47%** | 16% | 36% |
| *Total* | 79 | 48 | 80 | 38% | 23% | **39%** |

Fig. 5: Within Subject Analysis

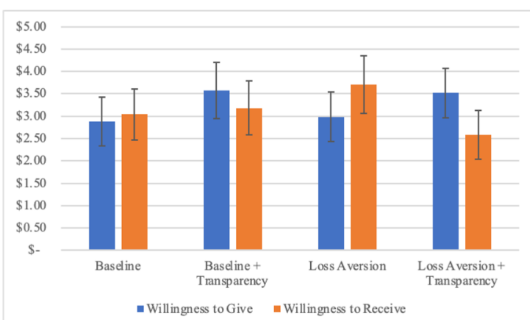



Fig. 6: Willingness to Give and Willingness to Receive Bar Chart

| Group | Allocators | | Receivers | | Difference | |
|---|---|---|---|---|---|---|
| | Average | Std Dev | Average | Std Dev | Average | Std Dev |
| *Base* | 2.88 | 1.96 | 3.04 | 2.04 | -0.16 | 2.54 |
| *Loss* | 2.98 | 1.97 | 3.71 | 2.31 | -0.73 | 2.73 |
| *Transparency* | 3.57 | 2.20 | 3.18 | 2.12 | 0.39 | 2.95 |
| *Loss + Transparency* | 3.52 | 1.96 | 2.58 | 1.92 | 0.94 | 2.66 |

Fig. 7: Willingness to Give and Willingness to Receive Table

Figure 8 checks to see whether the model fits the assumption for homoscedasticity. Looking at the diagnostic plots below, there are no large outliers that would cause bias in the model, and the Normal Q-Q plot's slope is fairly close to a linear slope of 1. From the diagnostic plots, we can conclude that the model fits the assumptions for homoscedasticity.

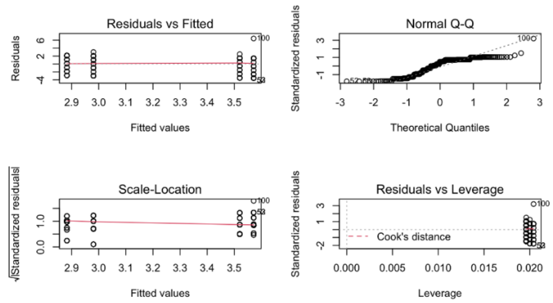



Fig. 8: Homoscedasticity Tests

To measure differences between group means, we performed a Tukey's Honestly Significant Difference post-hoc test to compare pairs from all groups, as shown in Figures 9 to 11.

```
$`Transparency:Treatment`
               diff        lwr       upr     p adj
1:0-0:0  0.68907563 -0.3598330 1.7379843 0.3252760
0:1-0:0  0.09803922 -0.9403274 1.1364058 0.9948396
1:1-0:0  0.63764706 -0.4058984 1.6811926 0.3905291
0:1-1:0 -0.59103641 -1.6399451 0.4578722 0.4636968
1:1-1:0 -0.05142857 -1.1054643 1.0026072 0.9992759
1:1-0:1  0.53960784 -0.5039377 1.5831533 0.5387306

$Transparency
        diff        lwr      upr     p adj
1-0 0.614082 0.05103811 1.177126 0.0327066

$Treatment
          diff        lwr       upr     p adj
1-0 0.02442394 -0.5385642 0.5874121 0.9319078
```

Fig. 9: Post-Hoc Test for Willingness to Give

```
$`Transparency:Treatment`
              diff        lwr         upr     p adj
1:0-0:0  0.1444578 -0.9448128  1.23372833 0.9860108
0:1-0:0  0.6666667 -0.4116562  1.74498949 0.3799017
1:1-0:0 -0.4592157 -1.5429167  0.62448534 0.6911959
0:1-1:0  0.5222089 -0.5670617  1.61147943 0.6008027
1:1-1:0 -0.6036735 -1.6982684  0.49092149 0.4828013
1:1-0:1 -1.1258824 -2.2095834 -0.04218133 0.0383321

$Transparency
         diff       lwr        upr    p adj
1-0 -0.4937611 -1.078471 0.09094862 0.097436

$Treatment
         diff        lwr     upr     p adj
1-0 0.0410081 -0.5436438 0.62566 0.8901259
```

Fig. 10: Post-Hoc Test for Willingness to Receive

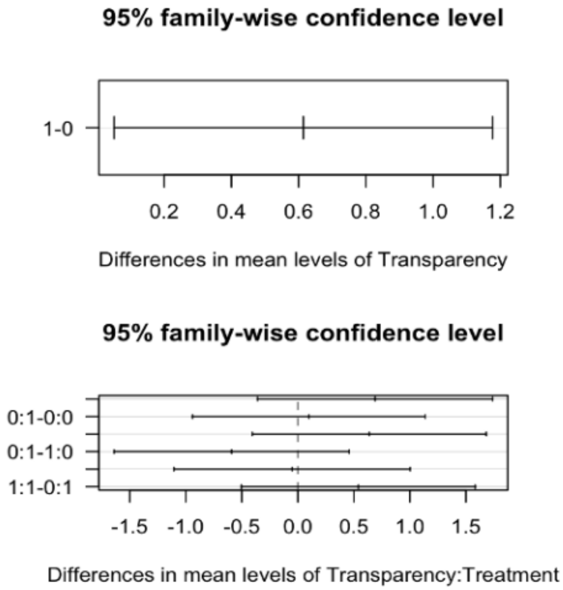



Fig. 11: Summary

## IV. Discussion

### A. Interpreted Data

As stated in the methods and hypotheses section, we wanted to test two main variables with this experiment; transparency and loss aversion. After running a two-sided analysis of variance (ANOVA) test on our data set, we found that only two of our test groups garnered statistically significant results, our Transparency group (giving: p-value = 0.0327, receiving: p-value = 0.0974) and our Transparency and Treatment group (receiving only: pvalue = 0.0334). Further, looking at our within-subject regression both the Transparency group and the Transparency + Treatment group led to significant results (p-value = 0.00407 and p-value = 0.1430 respectively). Thus, only the transparency and transparency + loss aversion groups led to any significant results. The conclusion that significant results occurred under transparency coupled with the treatment effect was shown for the dictator's willingness to receive in figure 10 (p-value = 0.0383). The loss aversion group did not lead to any statistically significant results, and potential reasons for this phenomenon are explored below.

### B. Hypothesis 1: Loss Aversion

In our experiment, we expected the allocator in our loss aversion group to give relatively less. However, we saw an average 21% increase in giving by the allocator when transparency was implemented, yet no significant change in allocations when a loss was induced. Oddly enough, we saw an increase in giving, albeit small, when there was no transparency but a loss occurred. There are various explanations for why loss aversion was not observed in the allocation portion of our experiment. A few reasonable explanations are that the respondents did not feel as if they had lost anything because the experiment was conducted virtually, participants may have not spent enough time to understand the nuance of the question (average survey time was 39 seconds), or the game was not believable since there was no real receiver. Physical experiments would allow us as researchers to raise the stakes and hand individuals $12 and then take $2 from them at a later time, however, we virtually demonstrated loss aversion as a hypothetical question, which might have been less convincing. Additionally, a physical experiment takes more time. Between verbally explaining the rules of the game, distributing money, taking away money, and asking follow up questions, participants would spend much more time thinking about their decision throughout these steps. Furthermore, subjects would be less incentivized to rush through the study, especially if we promised the resulting money to the participants.

Besides experimental design, we may have not seen loss aversion for a few other reasons. Cherry et al (2002) found that in their version of the dictator game, sums of money endowed to the dictator led to less loss aversion than earned sums of money. Adding a game that participants would play to earn their money could potentially lead to more loss aversion as the allocators would use an "earned mental account" instead of an "endowed mental account".

### C. Hypothesis 2: Transparency

As stated in the interpreted results section, transparency was significant in determining how much the allocator decided to offer the receiver. Corroborating the findings in the Thaler & Camerer paper, we concluded that transparency does lead to an increase in allocations, thus as allocators, participants' social preferences were geared toward being more generous. We believe that transparency removed asymmetric information from the game which nudged allocators to make larger allocations. For both the transparent + loss and transparent only group, transparency led to a statistically significant increase in allocations. The pseudo-receiver question for Group 4 resulted in a significant result such that participants expected the pseudo-allocator to believe that they would be empathetic to them as a receiver, thereby leading to a decrease in expected allocation by 19% (p-value = 0.0383). As allocators, there was no statistical significance for Group 4, however, there was strong statistical significance such that 84% of participants expected to receive the same or less than what they offered playing the allocator. Transparency had a significant effect in the post-hoc test for willingness to receive (p-value = 0.0974). This indicates that the participants expected other allocators to be loss averse, whereas allocators themselves were not. Further, Imas (2016), found that in the stock market realized losses led to loss aversion whereas paper losses often led to loss chasing. In our study, it could be hypothesized that transparency acted as a mechanism to 'realize' a loss. Similar to the realization effect, willingness to receive was significantly greater than willingness to give in the loss aversion treatment (e.g. the loss was viewed as not being realized) compared to the loss aversion and transparency treatment willingness to receive was significantly less than willingness to give, thus participants expected to receive less (e.g. the loss was viewed as being realized). This dichotomy indicates that transparency acted as a mechanism for allocators to realize a loss and thus enter the value into a new mental account.

### D. *Within-Subject Analysis*

Standard economic theory states that in the classic Dictator Game we would expect the dictator to offer \$0.01 and the receiver should accept anything above \$0.00. However, as mentioned earlier, past BE papers have found this to be untrue. In our study, this effect was still apparent: the average offer from the allocators across all treatment groups was \$3.23, meaning that inequity aversion still occurred in our study. If participants had homogeneous preferences, we would expect participants to offer the same amount as an allocator and expect the same amount as a receiver. Participants in the loss treatment without transparency expected to receive more than the baseline case, however, participants in the loss treatment with transparency case expected to receive less than the transparency only case. As stated above, perhaps an explanation for this outcome is that individuals did not realize the \$2 loss in their mental account and thus they would expect the allocator to make an allocation based on the \$12 mental account. In the transparent loss treatment, the \$2 loss is realized because the receiver witnesses the loss, and thus the mental account is set to \$10 and loss aversion with a realized loss occurs.

### E. *Experimental Design and Next Steps*

Our experiment consisted of both between-subject (treating subjects with a loss or transparency) and within-subject design (asking subjects how much they would be willing to allocate and how much they would expect to receive). The between-subject design consisted of a few flaws that potentially led to the surprising lack of loss aversion-induced inequity aversion found in our experiment. We believe that there are various factors that lead to our experiment resulting in only paper losses and never any realized losses, thus, in line with the Realization Effect (Imas, 2016). Some factors that may have contributed to the lack of realization in our study include lack of physical money, issues with real-life emulation, and brevity of the online experiment. Moreover, we could have received better data if the intervals of distribution were smaller. In our study, the intervals were 10% (\$1), but more robust experiments could have been run with 1% (\$0.10) distribution intervals. Without a physical receiver present, the effectiveness of transparency was likely lowered. For the within-subject experiment, subjects were both the allocator and the receiver but were always the allocator first. Thus, this could have led to a bias in the expected allocations received as participants could have used their previous experience as a reference point in the next question. These flaws would likely be reflected in the error term for the willingness to receive regression analysis. We would suspect the error term for the willingness to receive regression to be greater than the willingness to give regression due to response bias from the question placement. The residual sum of square for the WTR regression (870.0) was greater than for WTG regression (806.8). Thus, the flaw in not randomizing the position of our questions led to a larger error term for WTR.

In a broader context, this experiment could be applied to policy-making, sociology, or other social sciences as demographic questions could be included in the experiment to test if there are innate biases that surround immutable and mutable constructs like gender, age, political affiliation, or any other demographic attribute. We do not expect statistically significant data to come from these offshoots, however, it could be interesting for other disciplines to study.

It could also be insightful to test the observed one-sided altruism. We hypothesize that this observation is just an expectation that displays inequity aversion with a bias towards advantageous inequity aversion, however, (Xiao, 2021) discusses the "better than average" effect which finds that individuals often overestimate their own abilities. This could be analogous to participants overestimating the fact that they would demonstrate more generosity than another participant. Another interesting modification to the experiment would be to replace cash with another item such as a mug or a jacket; loss aversion could be modeled by damaging the item. Thus, we could test if the endowment effect is concurrent with loss aversion or transparency, which would build on the endowment effect.

## V. Conclusion

Under certain conditions, people behave in ways that normative economics considers anomalies. For instance, an individual's tendency towards altruistic behavior may depend on events leading to that moment, which may cause them to act less generously in cases of a prior loss. The transparency of an exchange can also play a role in generosity; symmetric information between the allocator and the receiving party can influence the giver's strategy. We believe this area of study is significant because it can provide insight into why many organizations act exceedingly frugal during economic downturns (Sharples, 2011). In this study, we conducted an experiment analyzing the separate and conjoined effects of loss aversion and transparency in a modified Dictator Game. Our results showed us that givers were relatively more generous under transparency. However, this was not the case coupled with a loss; participants were neither more generous nor selfish relative to the baseline Dictator Game. In that particular scenario, participants expected they would receive significantly less than what they gave themselves if they played the role of the recipient, suggesting there was an asymmetry between actions and expectations. We suggest future research implementing loss aversion and transparency treatments on alternate economic games such as Ultimatum and coalitional games, real-life scenarios, and money substitutes to see whether those results would be consistent. This type of research could be valuable in adding knowledge to the economic literature as well as other academic fields including charitable giving during a recession or social effects on charitable giving.

## References


[1] Boun My, K., Lampach, N., Lefebvre, M. & Magnani, J. (2018). Effects of gain-loss frames on advantageous inequality aversion. *Journal of the Economic Science Association*, 4(2), 99-109.

[2] Camerer, C., & Thaler, R. H. (1995). Anomalies: Ultimatums, Dictators, and Manners. *Journal of Economic Perspectives*, 9(2), 209-219.

[3] Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the Dictator. *The American Economic Review*, 92(4), 1218-1221.

[4] Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33, 67-80.

[5] Engel, C. (2011). Dictator games: a meta study. *Experimental Economics*, 14, 583-610.

[6] Halvorsen, T. U. (2015). Are dictators loss averse? *Rationality and Society*, 27(4), 469-491.

[7] ] Imas, A. (2016). The Realization Effect: Risk-Taking after Realized versus Paper Losses. *American Economic Review*, 106(8), 2086-2109.

[8] Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), 193-206.

[9] Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2), 263-292.

[10] Thaler, R. H. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12, 183-208.

[11] Thaler, R. H. (1999). Mental Accounting Matters. *Journal of Behavioral Decision Making*, 12, 183-208.

[12] Sharples, A. (2011). The Great Recession and its Effect on Small Businesses and Unemployment Rates.

[13] Xiao, Y., Wong K., Cheng, Q., & Yip, P. S. F. (2021). Understanding the Better Than Average Effect on Altruism. *Frontiers in Psychology*.

## APPENDIX A: QUALTRICS SURVEY

Group 1 (left) & 2 (right): Note the second row was asked in both groups.

You are starting off with **$10** and must decide how much of the funds to distribute. The person who will be receiving the distribution of the funds knows that you are starting with **$10**.

How much money are you willing to give to the other person?

0 1 2 3 4 5 6 7 8 9 10

U.S. Dollars ($)

You are starting off with **$10** and must decide how much of the funds to distribute. The person who will be receiving the distribution of the funds does not know how much you are starting with.

How much money are you willing to give to the other person?

0 1 2 3 4 5 6 7 8 9 10

U.S. Dollars ($)

What is the lowest amount you would be willing to receive if you were the receiver?

0 1 2 3 4 5 6 7 8 9 10

U.S. Dollars ($)

Group 1 (left) & 2 (right): Note the first and third rows were asked in both groups.

You are starting off with **$12**.

Imagine that you suddenly lose **$2** and now have **$10**. You must decide how much of the funds to distribute. The person who will be receiving the distribution of the funds does not know how much you are starting with.

How much money are you willing to give to the other person?

0 1 2 3 4 5 6 7 8 9 10

U.S. Dollars ($)

Imagine that you suddenly lose **$2** and now have **$10**. You must decide how much of the funds to distribute. The person who will be receiving the distribution of the funds knows that you started with $12, but then lost $2.

How much money are you willing to give to the other person?

0 1 2 3 4 5 6 7 8 9 10

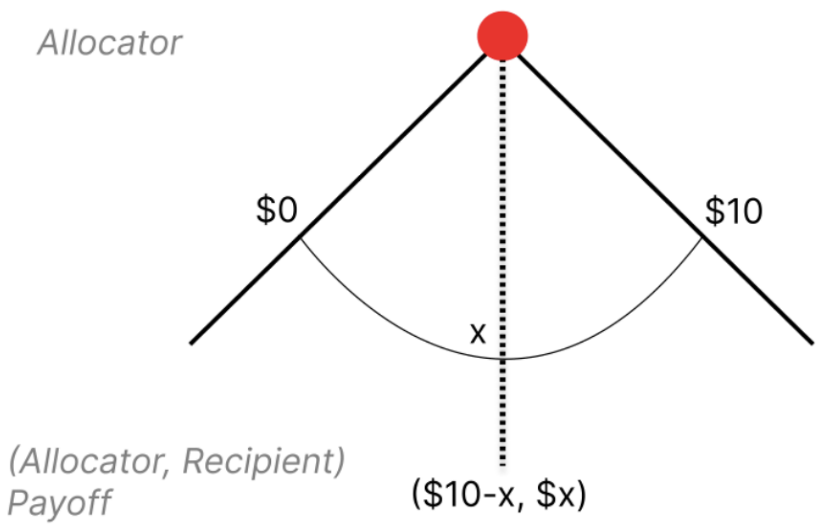
U.S. Dollars ($)

What is the lowest amount you would be willing to receive if you were the receiver?

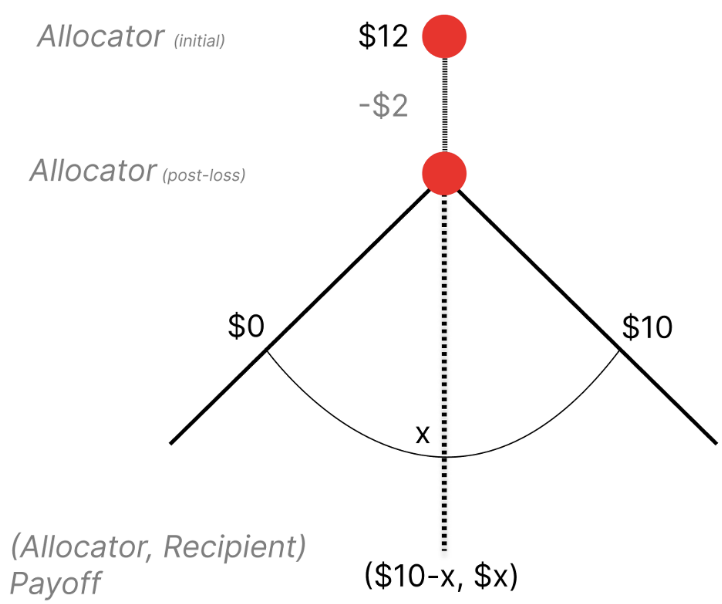
0 1 2 3 4 5 6 7 8 9 10

U.S. Dollars ($)

## I. Dictator Game




## II. Dictator Game with Loss Aversion



# The Effect of Primary Care Provider Supply on Downstream Emergency Department Visits for Pediatric Asthma in California

Jashdeep Dhillon

*University of California, Berkeley*

*Abstract*— **California is facing a primary care workforce shortage. Although this has been studied at the state level, it is important to understand the downstream health implications: how will this affect the management of chronic illnesses or how will this affect access to primary care? Due to the significance of investing in childhood health, I exploit different healthcare access features (i.e. the number of primary care physicians) across California's counties with a variety of socioeconomic variables to investigate the effect of the supply of primary care providers on emergency department visits for pediatric asthma. I find that measures of access (in the number of providers) have a negative relationship with the downstream ED visit rates for pediatric asthma.**

## I. Introduction

### A. Background

In this paper, I investigate the impact of California's primary care provider supply on the rates of emergency department visits for asthma in the pediatric population at a county level.

Most population-level research unequivocally demonstrates the powerful influence of economic conditions and various social characteristics on a community's health status. One of the common diseases that burden large community groups in the United States is asthma, a chronic inflammatory disorder of the respiratory system. Several studies have clearly established that asthma is prevalent among children in areas with a high density of environmental exposure (i.e. urban regions), which is compounded by racial disparities (Hasegawa et al. 2014). In addition, a study conducted in a section of New York undergoing structural changes found that socioeconomic determinants, including household income and health insurance, were key predictors of emergency department utilization for asthma complications (Eum, Youngseob, EunHye, and Bowen, 2019). It is apparent that asthma healthcare utilization disproportionately affects children of those living in low socioeconomic status (SES) neighborhoods, especially in historically divested communities. According to the Children's Advocacy Institute, asthma is one of the single most influential chronic disorders that exacerbates school absenteeism, which negatively impacts academic and social growth. In 2012, Largent et.al demonstrated that access to quality healthcare is a significant barrier to adequate asthma treatment. Their work showed that asthma has a significant impact on the rates of emergency department (ED) visits, with the results being most prevalent in children's age group of 0-4 years. Thus, it is important to determine and analyze how emergency department visits for children differ, as a function of their underlying demographics, across California in order to begin formulating programs and policies to improve the health and well-being of local communities and statewide.

In another dimension, California's composition of medical providers reveals a shortage of primary care clinicians in the next decades. Coffman et.al have demonstrated that despite a 21% growth rate of primary care physicians (PCPs) across California between 2006 and 2018, the "adequate" supply is patchwork. Small and rural counties confront shortages of primary care physicians, highlighting the asymmetric distribution across the state's 58 counties. The latest report by Coffman et al. shows the dire state of the healthcare landscape - the current supply of primary care physicians just barely met the minimum per capita ratio recommended by the Council of Graduate Medical Education. Recent statistics signal more macabre projections with an estimated shortfall of 10,500 primary care clinicians by 2030, representing a gap of 4,100 additional primary care physicians. In a report that detailed the forecasted supply and demand of medical professionals in the state, Coffman endorsed: "If we continue along our current path, more and more Californians will need to visit the emergency room for conditions like asthma, ear infections or flu because they lack a primary care provider" (Coffman, Geyn, Himmerick, 2017).

Prior assessments also have forecasted the supply of mid-level providers, such as nurse practitioners (NP) and physician assistants (PA), with primary care deficits worsening due to the retiring demographic in the NP aggregate. Despite the number of NPs and PAs growing between 2004 and 2016, their numbers still lag behind that of PCPs, reinforcing the point that most regions of California still do not have an adequate primary care workforce. Even amongst the mid-level providers' population, there is a scattered distribution in the supply of primary care services: 22% of PAs and 50% of NPs provide primary care, in contrast to 36% of active full-time equivalent physicians (Coffman, Geyn, Himmerick, 2017).

The specific geographical focus on California was chosen

I would like to thank my thesis advisor, Professor Benjamin Handel, for their support in the design and development of this paper. I would also like to thank Professor Ryan Edwards for facilitating the preliminary stages of the research process and helping me in the construction of a prospectus. In addition, I would like to also thank my colleague and incoming Berkeley MIDS candidate, Dhangur Singh, for their robust technical guidance in the use of Python for data wrangling and their support throughout the data crunching process.

because of the state's asymmetrically distributed collection of physicians, where most providers are generally practicing in dense populations e.g. Bay Area. In addition, most primary care residencies (physician training programs) are in hospitals within urban settings. As a result, both rural and some under-invested urban settings are home to low-income communities that are disproportionately uninsured and reliant on social insurance programs (i.e. MediCal). With this portrait of primary care clinicians in California, it must be apparent that the shortage of medical care is directly measured by the rate of illnesses and diseases in these areas. If California continues to trend in this direction, it is likely that more individuals will delay visits for preventative care and only utilize emergency services for acute and chronic conditions when they become devastatingly grave. This is also why pediatric asthma was isolated for this study. These asymmetries will ultimately mean managing diseases, like asthma, high blood pressure, and high cholesterol, will become much more difficult, leading to unnecessary complications and death. In addition, improving childhood health is important for developing a generation of productive young adults: If high-quality care cannot be ensured for California's pediatric population, this will significantly compromise the healthcare system and generations of adults beyond.

### B. Research Question and Hypotheses

The purpose of this paper is to determine whether the number of primary care providers (PCPs, NPs, PAs) has an impact on downstream ED rates for pediatric asthma in California's counties. I hypothesize that increasing the number of medical providers specializing in primary care will decrease the ED visit rates for pediatric asthma. For the types of medical providers, I look specifically at the number of full-time equivalent (FTE) primary care physicians, family nurse practitioners, and physician assistants in California counties. For ED outcomes, I looked at the number of emergency room cases pertaining to the pediatric asthma population (ICD-10-CM). Based on current research, I developed a series of hypotheses for my question:

TABLE I: Hypotheses

| |
|---|
| *H1*: Increasing the number of primary care physicians decreases pediatric asthma ED rates |
| *H2*: Increasing the number of family nurse practitioners decreases pediatric asthma ED rates |
| *H3*: Increasing the number of physician assistants decreases pediatric asthma ED rates |

### C. Study Area

My thesis is a retrospective analysis of asthma-related ED visits longitudinally over a 5 year period from 2015 to 2019 using data from the California Health and Human Services. The study area consisted, concurrently, of a total of 58 counties in California. I extracted ED visits due to asthma symptoms and diagnosis based on patient age, sex, race/ethnicity, and primary diagnosis code (ICD-10) between 2015 and 2019, in order to compare changes over a span of time. Based on the reasons stated in my previous section, I focused on the pediatric age cohort (ages 0-17) in the regression analyses.

## II. Literature Review

### A. Relevant Literature

There is a section of public health research establishing the various impacts of socioeconomic conditions on the predictors of health, primarily related to structural factors that include race, educational gaps, poverty, and access to social resources. There is a subset of literature that is analyzing how these same factors, coupled with healthcare characteristics (i.e. insurance coverage, access to a physician), affects the health of children and adults, alike, in rates of acute and chronic diseases. First, I will review literature studying the quality of care for children with asthma and the reported use of emergency visits for pediatric asthma. Then, I will look at the economic literature discussing the broader demographic features that contribute to the demographic differences in ED visits and hospitalizations for asthma. Finally, I will review the meta-literature discussing the necessity of closing the primary care workforce gap.

There is extensive literature discussing the impact of healthcare coverage in ensuring children have appropriate quality care and consistent access to medical providers in order to prevent asthma-related morbidity. Literature has shown that although efforts have been made to reduce health-related disparities, asthma continues to persist in low-income populations and makes significant economic dents. According to data from the Medical Expenditure Panel Survey, asthma costs the US economy approximately $80 billion per year due to issues such as school absenteeism, loss of work time, etc. This is directly a healthcare coverage problem as a study found that a lack of consistent coverage is associated with the poor overall quality of care for children with asthma, claiming that even brief periods of discontinued coverage can prevent children from receiving timely interventions (Halterman, Montes, Shone, Szilagyi 2008). The study used univariate chi-square analyses to establish differences between insurance groups in terms of healthcare outcomes using data from the National Survey of Children's Health conducted between January 2003 and July 2004. In addition, this study determined that such differences were cemented across racial lines: Caucasian children were more likely to report full-insured private coverage than minority ethnic groups, and thus face less asthma burden. Also, the authors of this study conducted an odds ratio for unmet medication needs with the continuity of care finding that the odds of having an unmet need (or untreated asthma) were 6.5 and 14.02 for the gained insurance and uninsured group, respectively.

Multiple studies have specifically looked at the socioeconomic distribution and demographic differences in emergency department (ED) visits and hospitalizations for pediatric asthma. Largent et al. developed a retrospective population-based and cross-sectional analysis of asthma hospital admissions and ED visits for Orange County (California) children and adolescents between 2000 and 2007. They used a logistic regression model with children ages 0-14 to determine the association between hospital admission of certain racial/ethnic groups and socioeconomic status (using metrics, such as median household income, for standardization). This allowed them to see whether there was a relationship between the prevalence of ED rates with hospital admissions, in order to determine the severity of cases that necessitated further medical intervention. This paper found the greatest declines in admission rate occurred in the 0-4 year age cohort. In addition, lower SES groups (ie, African Americans) had the highest rate of ED visits and hospital admissions, revealing that certain groups are disadvantaged and are at higher risk. Despite these results, researchers acknowledge that only focusing on a single county in California lends itself to possible under-or-over estimation of hospital admission rates. Additionally, a study analyzing California's San Joaquin Valley observed that asthma utilization in the emergency department as measured by insurance coverage was influenced by community-level socioeconomic factors, such as concentrated poverty (Alcala, Emanuel, Ricardo, Capitman, 2018). Such "confounding" factors can explain the results from the study by Largent et al.

Similarly, in a study looking at the distribution of asthma-related ED visit rates in North Carolina, Dieu et al. discovered that the larger burden of asthma-related visits was shouldered by rural and impoverished counties in North Carolina. The study was a retrospective data analysis of North Carolina patients from 120 hospitals who were diagnosed with asthma in the emergency department (ED) between 2010 and 2014. Researchers found a greater density of asthma-related ED visits in rural counties (particularly in the east) as compared to suburban and urban counties. Among these counties, they found that young children between the ages of 5 to 9 experienced the highest percentage of ED visits. Although there was no data on the disaggregation of race and ethnicity within the study, the counties with the highest prevalence of asthma cases were predominantly black, non-Hispanic populations. The following variables were used in the analysis: ED visit date, age-specific groups, sex, and patient county residence. The study suggests that hospitals in rural counties are shouldering the cost of excessive asthma-related visits; according to researchers, it is apparent that asthma education programs and medication compliance are not adequate.

Equally as important as the research that exploits data to run various linear regression and difference analyses, the broader meta-literature discussing the role of primary care is imperative to understanding the crux of the paper I have presented. Shi expounds on the significance of primary care as a cornerstone of the healthcare system in providing equitable access to care in a comprehensive and longitudinal manner. His study reveals how increasing the supply of primary care can alleviate negative economic conditions, such as income inequality that mitigates access for various racial/ethnic minorities. In addition, Shi showcases how primary care fills in the gap in medical care for under served populations, highlighting the need for robust policy making for expanding primary care access. Also, McGovern et al. conducted a longitudinal retrospective review of children diagnosed with asthma that scheduled a primary care appointment between 2010 and 2012. They found that 2.7% of the children who did not complete a primary care appointment with their physicians ended up having a subsequent emergency department visit for their asthma symptoms whereas none of the children, who were compliant with scheduled primary care appointments, had contact with the emergency department. This speaks to the efficacy of primary care intervention at reducing preventable ED visits.

Another study examined the physician shortage upon which the Association of American Medical Colleges (AAMC) has issued a clarion call to develop mechanisms to narrow the gap between physician supply and demand. Morgan 2019 provides projection models that demonstrate including other primary care providers such as family nurse practitioners (NPs) and physician assistants (PAs) can have a large impact on closing that deficit and, in fact, work at the same productivity level as a physician. Developing the future workforce with NPs and PAs in consideration of their contributions to healthcare delivery can alleviate the shortfalls of the physician shortage and even lower the costs imposed upon consumer society.

## III. Data

### A. Data Source

My primary data source was the California Department of Health and Human Services, a department that oversees state entities that perform work on the social service and public health needs of Californians. In addition, another data source that enhanced the profile of my regression models came from the 5-Year Narrative Profiles of American Community Survey (ACS), which is released yearly to provide data on social, economic, housing, and demographic characteristics of different geographical units across the United States. I specifically selected data for California counties between 2015-2019.

The dataset for pediatric asthma contained the rates of asthma emergency department visits per 10,000 residents among Californians across all 58 counties and age groups (0-17, 18+, and all ages). I specifically focused on the pediatric cohort from 0-17. Per the source dataset, the data is reflective of emergency department visits from licensed hospitals in California that use the ICD10-CM diagnosis code for asthma. Although the dataset contains 58 counties, in my primary analysis (i.e. time fixed effects), I only use 50 counties due to poor data quality. These counties include Alpine, Amador, Inyo, Mariposa, Mono, Plumas, Sierra, and Trinity. The larger pattern is that these counties are rural; for example, Alpine is California's least populous county which thus, makes my dataset susceptible to selection bias. I would like to acknowledge that the results of my statistical analysis may not be generalizable to the counties left out.

I selected healthcare characteristics from the Primary Care Clinic Annual Utilization Data from the California Health and Human Services. Health features include the number of full-time equivalent(FTE) primary care physicians, family nurse practitioners, and physician assistants, which tracked utilization information for different primary clinics across the state.[1] Other information included encounters by specific medical service, procedure codes, gross revenue of clinics, etc.

I used several economic and housing characteristics, such as health insurance coverage, income, and poverty status from the 5-year estimates of the ACS. The data profiles of the county were presented in both estimates and percentages. I selected these variables to cover the socioeconomic profile of the different California county units in my regression analysis.

## IV. Model and Empirical Strategy

### A. Methodology

To investigate the relationship between the primary care provider supply and potential downstream effects on pediatric asthma ED rates, I performed regression analyses on three separate models that employ the different datasets. The models are listed below:

*1) Model 1: Healthcare Access Features:* In this model, I solely look at the influence of explanatory variables that measure healthcare access or supply in order to determine the influence of these variables on our ED rate outcome. This scenario demonstrates the effect of the healthcare supply characteristics on the health implications for asthma emergency department visits, independent of other interactions or social structural factors. As will be explained later on, this model is not meant to be a truly robust explanation for the observed effects since it does not fully capture the different interactions with other non-health variables. The standard default regression equation for this model is

$$Y_{i,t} = \beta_0 + \beta_1(PCPs10K)_{i,t} + \beta_2(PA10K)_{i,t} \quad (1)$$
$$+ \beta_3(FNP10K)_{i,t} + \beta_4(Encounter10K)_{i,t} \quad (2)$$
$$+ u_{i,t} \quad (3)$$

Model 1 is an OLS regression model where *PCPs10K* is the estimated number of FTE primary care physicians per 10,000 people in county *i* at time *t*. *PAs10K* is the estimated number of FTE physician assistants per 10,000 people in county *i* at time *t*.*FNPs10K* is the estimated number of FTE nurse practitioners per 10,000 people in county *i* at time *t*. *Encounter10K* is the estimated total number of preventive care appointments for infants, children, adolescents, and adults per 10,000 people in a given county *i* at time *t*.

*2) Model 2: Socioeconomic Factors:* In this model, I study the relationship between several county-level socioeconomic variables upon downstream pediatric asthma ED rates. My goal was to differentiate how healthcare characteristics and socioeconomic characteristics of counties across different years influence the response variable in my study. As stated above, isolating only the social features of the different counties does not capture the full story. The standard default regression for this model is

$$\begin{aligned} Y_{i,t} = \beta_0 &+ \beta_1(NoHI10K)_{i,t} + \beta_2(Unemployed10K)_{i,t} \\ &+ \beta_3(Poverty10K)_{i,t} + \beta_4(PerCapIncome)_{i,t} \\ &+ \beta_5(MedHouseValue)_{i,t} + u_{i,t} \end{aligned} \quad (4)$$

Model 2 is an OLS regression model where *NoHI10K* is the number of uninsured individuals per 10,000 people in county *i* at time *t*. *Unemployed10K* is the number of unemployed individuals per 10,000 people in county *i* at time *t*.*Poverty10K* is the number of individuals in poverty per 10,000 people in county *i* at time *t*. *PerCapIncome* is the mean income for every individual in county *i* at time *t*. *MedHouseValue* is the median value for real estate prices in county *i* at time *t*.

*3) Healthcare Access and Socioeconomic Factors:* This model explores the association of healthcare features and economic performance across California counties and the influence on asthma ED rates. My aim is to provide insights as to how we can dynamically evaluate economic factors in conjunction with healthcare performance to see how they work together to provide a portrait of pediatric asthma ED incidence rates. The standard default regression for this model is

$$\begin{aligned} Y_{i,t} = \beta_0 &+ \beta_1(PCPs10K)_{i,t} + \beta_2(PA10K)_{i,t} \\ &+ \beta_3(FNP10K)_{i,t} + \beta_4(Encounter10K)_{i,t} \\ &+ \beta_5(NoHI10K)_{i,t} + \beta_6(Unemployed10K)_{i,t} \\ &+ \beta_7(Poverty10K)_{i,t} + \beta_8(PerCapIncome)_{i,t} \\ &+ \beta_9(MedHouseValue)_{i,t} + u_{i,t} \end{aligned} \quad (5)$$

### B. Pooled OLS

I used a Pooled OLS regression model to define the relationship between the number of primary care providers in a Californian county (total of 50 counties observed) and the annual rate of emergency department visits for pediatric asthma for that county between 2015 and 2019. Since the following data can be presented as a time series panel, it was fitting to begin with Pooled OLS in order to derive estimates of the parameters in the presence of time attributes. I checked the suitability of the regression model for the panel data sets and discovered that it met 1 of 2 conditions for Pooled OLS: auto-correlation but not homoscedasticity. In addition, it is important to note that the residual errors are not normally distributed; however, the pooled OLS estimator for the parameters still is the best linear unbiased estimator for this panel series. The non-normality of the residuals is not a significant deterrent to estimation; however, I cannot build reliable confidence intervals. This issue is also contributed by heteroscedasticity since the standard errors in confidence intervals are going to be biased.

[1]The data sets used numbers calculating FTE for the various providers. For example, in the year 2019, there were 140.17 FTE PCPs operating in Alameda County, suggesting some physicians were not working the standard 40 hour work-week.

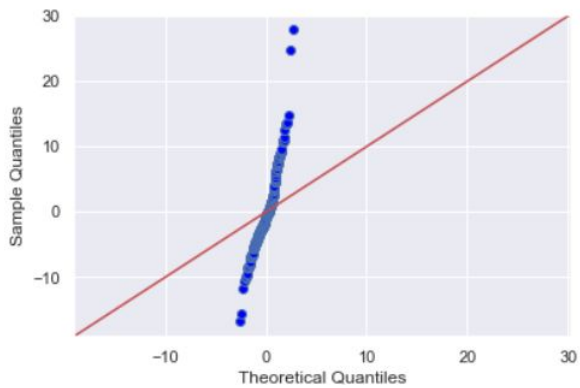



Fig. 1: Non-normality of Residuals for Model 3

### C. Time-Fixed Effects

I used a time-fixed effects model to capture the effects of the parameters that do not change over time in order to eliminate the risk of bias due to omitted factors over time. Hence, I clustered the covariance by time to control for correlation between counties across the 5-year time period. I used the following fixed effects regression model:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \delta_2 B2_t + \cdots + \delta_T BT_t + u_{it} \quad (6)$$

The subscript $i$ denotes the 50 individual counties being observed (1,...,n). The subscript $t$ denotes the time period. The error term contains non-time-invariant omitted variables and the portion of the outcome variable that cannot be explained by the regression model.

Throughout all these regression analyses, there are 50 observations (or counties) across the 5 year time period; these counties (including Alpine, Amador, Inyo, Mariposa, Mono, Plumas, Sierra, and Trinity) have been removed from the analyses due to lack of empirical evidence in the datasets (Note: some of the stated counties are rural and may lack the breadth of comprehensive resources to do detailed reporting of asthma emergencies).

## V. Results

### A. Pooled OLS

Table 2 displays the regression estimates for the asthma emergency department outcome variables for all 3 models using the Pooled OLS method. In Model 1,the particular estimator of interest is *PCPs10K* due to California's current deficit of primary care physicians; however, it is important to note that these other parameters are equally vital in measuring their effect of the downstream outcomes on asthma emergency service use. Table 2 indicates that, indeed, the physician provider supply is statistically significant at the 1% level, particularly in the estimated reduction of the number of pediatric asthma ED rates due to an increase in 1 FTE PCP per 10,000 is 2.88 patients per 10,000 people. This validates my hypothesis that increasing the supply of PCPs will reduce the level of emergency service use. However, we do see that the number of primary care encounters is statistically significant at the 1% level but marginally trends in the positive direction that, in fact, increased utilization of out-patient primary care services increases contact with emergency departments in pediatric populations. The sign of the coefficient did not align with my initial hypothesis. In addition, I ran a Pearson's test of correlation to determine if there is any association between the residual errors and my outcome variable and discovered the correlation to be 94%. It suggests that the Pooled OLS model is missing some variables that could have explained this correlation, hence, leaking into the residual errors. Another important finding to note is that the adjusted R-squared value was quite low indicating that in the aggregate, our predictor variables are not significant in this regression model.

Table 2 also displays trends in the outcome variables due to socioeconomic explanatory variables, including the level of uninsurance, unemployment, poverty, and the median house value. This is isolated from Model 1 in order to see how socioeconomic characteristics, individually, affect our downstream ED rate outcome. Table 2 indicates that increasing the unemployment rate by 1 individual approximately increased the number of pediatric asthma ED rates by 0.0128 patients per 10,000 people, which is statistically significant at the 1 percent level. In addition, the results reveal that poverty also has a statistically significant effect at the 1 percent level, particularly increasing poverty by 1 individual (per 10,000) increases the number of pediatric asthma ED rates by 0.0119 people per 10,000 people. Also, the coefficient on *MedHouseValue* is positive and significant. The interpretation is that the estimated increase in the number of pediatric asthma ED rates due to an increase in the value of the house by $1 is 6.06E-6 per 10,000 people. I was surprised to find that the uninsurance rate did not have any statistically significant effect on the outcome of the ED rate, nullifying my initial inclination to believe the decreased provision of health insurance would lead individuals to solicit emergency care services, in the case of acute asthma attacks (or any-asthma related pathology). One interesting finding was that the adjusted R-squared value in this model was roughly 5 times greater than the adjusted R-squared value in Model 1, suggesting the socioeconomic predictor variables are adding marginally more value to our regression model than the healthcare characteristics. Similar to Model 1, the correlation between the residual errors and the response variable was quite significant at 72% which suggests that this model is missing important explanatory variables. Also, based off a rudimentary comparison, we see that the F-statistic in Model 2 is much larger than Model 1, suggesting that Model 2 is statistically more significant than Model 1 in our regression analysis.

Table 2 also displays interesting results about the relationship between both healthcare and socioeconomic variables on pediatric asthma ED rates. As shown in the table above, we see predictor variables that were once significant are no longer statistically significant in Model 3 (i.e. the number of primary care physicians or the unemployment rate). However, we do see that the number of physician assistants and the poverty rate are statistically significant at the 1 percent level. Specifically, Table 3 demonstrates that the coefficient on the physician assistants is negative and significant. The interpretation is that the estimated reduction in the number of pediatric asthma ED rates due to an increase of 1 physician assistant (per 10K) is -4.6443 per 10,000 people. This aligns with my hypothesis that increasing primary care provider supply should decrease the downstream ED rates for pediatric asthma. In addition, the coefficient on the poverty rate is positive and significant. The interpretation is that the estimated increase in the number of pediatric asthma ED rates due to an increase in 1 impoverished individual is 0.0127 per 10,000 people. We can observe that the R-squared value has increased to 0.528, which is given since we are adding more predictor variables to our model.

TABLE II: Pooled OLS Results

| | *Dependent variable: Pediatric ED Asthma Rates, $Y_{i,t}$* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | $-2.8822^{***}$<br>(1.079) | | $-0.2260$<br>(0.818) |
| *PAs10K* | 1.0154<br>(2.092) | | $-4.6443^{***}$<br>(1.619) |
| *FNPs10K* | $-2.5742$<br>(1.518) | | $-1.8442$<br>(1.136) |
| *Encounter10K* | $0.0125^{***}$<br>(0.003) | | 0.0039<br>(0.002) |
| *NoHI10K* | | 0.0012<br>(0.002) | 0.0012<br>(0.002) |
| *Unemployed10K* | | $0.0128^{**}$<br>(0.005) | 0.0074<br>(0.005) |
| *Poverty10K* | | $0.0119^{***}$<br>(0.001) | $0.0127^{***}$<br>(0.002) |
| *MedHouseValue* | | $6.06e{-}06^{**}$<br>($2.67e{-}06$) | $2.686e{-}06$<br>($2.72e{-}06$) |
| Constant | $16.3350^{***}$<br>(0.918) | $-5.2865^{***}$<br>(2.513) | $-0.9726$<br>(2.595) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | No | No | No |
| Observations | 250 | 250 | 250 |
| $R^2$ | 0.107 | 0.488 | 0.543 |
| Adjusted $R^2$ | 0.092 | 0.480 | 0.528 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Regressions (1) - (3) are linear regressions without county or year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

### *B. Time Fixed Effect*

According to Table 3, we find that the parameters estimating the number of physicians and the number of primary care encounters are both statistically significant at the 1 percent level. In addition, the sign and magnitude of the coefficient on these explanatory variables are similar. According to our primary care physician coefficient, the estimated reduction in the number of pediatric asthma ED rates due to an increase of 1 physician (per 10K) is 2.97 per 10,000 people. Likewise, the coefficient on pediatric encounters demonstrates that the estimated increase in the number of pediatric asthma ED rates due to an increase of 1 pediatric/preventive medicine appointment (per 10K) is 0.0136 per 10,000 people. This may suggest frequent visits to the physician are associated with more acute episodes of asthma exacerbation or just a simple increase in preventive check-ups since this parameter is unable to discriminate against asthma-specific episodes. In a basic sense, the relationship between primary care physician supply and the volume of pediatric appointments/primary care service contacts are not influenced by time-invariant omitted variable bias.

Unlike Model 2 in our Pooled OLS model, we have removed the poverty parameter and replaced that with the per capita income variable due to the improvement of fit in the fixed effects model. I am unable to perform a complete comparison in terms of which regression approach is more sound since I have utilized different parameters in this case. This could suggest that there is some time-variant poverty-level factor that diminishes the explanatory power of the model, which is why I have removed it. Table 5 demonstrates that all socioeconomic variables, besides median house value, are statistically significant at the 1 percent level, in explaining the effect on pediatric asthma ED rates. A clear (while marginal) trend in the socioeconomic profile of counties (across the 5-year time period) shows that low-income communities (with the decreased provision of resources and social services as measured by unemployment and health insurance) have higher rates of ED visits for pediatric asthma. For example, the coefficient on *Unemployed_10K* is positive and significant. The interpretation is that the estimated increase in the number of pediatric asthma ED rates due to an increase of 1 unemployed individual (per 10K) is 0.0342 per 10,000 people. It is important to keep in mind that there are other factors in counties that compose the socioeconomic profile that we have not controlled for in our model that may be affecting these trends and ED outcomes.

When assessing the healthcare features with socioeconomic variables in the time-fixed effects model, it was interesting to discover that the coefficient for family nurse practitioners, rather than primary care physicians, was more significant at the 1 percent level in the effect on ED outcomes. For example, I find that the estimated reduction in the number of pediatric asthma ED rates due to an increase of 1 family nurse practitioner (per 10K) is 3.2377 per 10,000 people. In addition, the socioeconomic variables for health insurance and unemployment were also significant at the 1 percent level. Although I am unable to isolate the rurality and socioeconomic profiles of the counties in a nuanced fashion from the regression results above, I speculate that the significant effect of NPs can be explained by the fact that NPs are concentrated in smaller cities and rural settings, where they may see Medicaid populations. Of the additional healthcare variables, it was interesting to see that the coefficient for physician assistants was significant at the 10 percent level. In Model 1, physician assistants were hardly close to having a significant effect on the outcome and had a positive correlation with the response variable, which did not align with my hypothesis. However, this Model 3 shows that physician assistants do, indeed, have a negative correlation with pediatric asthma ED incidence rates.

TABLE III: Time Fixed Effects Results

| | *Dependent variable: Pediatric ED Asthma Rates,* $Y_{i,t}$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | $-2.9761^{***}$<br>(0.9964) | | $-0.8317$<br>(0.6887) |
| *PAs10K* | 0.7021<br>(1.3599) | | $-2.8567^{*}$<br>(1.5276) |
| *FNPs10K* | $-2.2652$<br>(1.5874) | | $-3.2377^{***}$<br>(0.9362) |
| *Encounter10K* | $0.0136^{***}$<br>(0.0014) | | $0.0090^{***}$<br>(0.002) |
| *NoHI10K* | | $0.0036^{***}$<br>(0.001) | $0.0041^{***}$<br>(0.0015) |
| *Unemployed10K* | | $0.0342^{***}$<br>(0.0028) | $0.0279^{***}$<br>(0.0024) |
| *PerCapIncome* | | $-0.0003^{***}$<br>(0.0001) | $-0.0002$<br>(0.0001) |
| *MedHouseValue* | | 7.09e−06<br>(4.486e−06) | 9.77e−07<br>(4.055e−06) |
| Constant | $15.974^{***}$<br>(0.2013) | $6.8785^{**}$<br>(2.7046) | $9.6179^{**}$<br>(4.1686) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | Yes | Yes | Yes |
| Observations | 250 | 250 | 250 |
| $R^2$ | 0.1202 | 0.3711 | 0.4321 |
| Adjusted $R^2$ | 0.1056 | 0.3256 | 0.4010 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Regressions (1) - (3) are linear regressions without county but with year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

## VI. Robustness Checks

### A. Cross-Sectional Year-Year Regression

My previous analyses operated under the assumption that my data preserved a constant variance term and was unrelated to the predictor variables. It was my initial assumption that the error term might have been heteroskedastic since there is more variability among counties that have higher physician supply or higher levels of outpatient pediatric appointments, for example. I used a Breusch-pagan test to confirm if there was heteroskedasticity in my linear regression model: I discovered that heteroskedasticity was ,indeed, present. As a result, I performed cross-sectional year-by-year regressions with heteroskedasticity-robust standard errors. I completed this test on all three models: (1) healthcare features, (2) socioeconomic variables, and (3) healthcare and socioeconomic variables.

In the following regression shown in Tables 4-8, I find that the poverty predictor variable is consistently significant at a 1% or 5% significance level throughout 2015-2019. In addition, I find that the number of pediatric appointments is statistically significant on a yearly basis. In 2017 and 2019, an increase in one pediatric appointment visit to a primary care provider increased the rate of admission by 0.0150 and 0.0140 individuals per 10,000 people, respectively. It is difficult to make a claim as to why the number of physician assistants and nurse practitioners is periodically significant across the spectrum as there is no one specific policy intervention that has directed my findings. In addition, it is interesting to find that 2018 yields no significant results in Models 2 and 3. It is important to note that relative to the time-fixed effects and pooled OLS models, the cross-sectional analyses contained 53 observations, except for 51 observations in 2015 and 52 observations in 2019. This increase in the number of observations per year may have had a marginal or insignificant effect on the assumptions used in our models. For the most part, I find that the objective correlation in the data supports my initial hypothesis that an increasing number of primary care clinicians and professionals should be associated with a downward trend in downstream pediatric asthma ED rates.

TABLE IV: Cross-Sectional Regression Year 2015

| | *Dependent variable: Pediatric ED Asthma Rates,* $Y_{i,t}$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | −0.5783 | | −1.4458 |
| | (2.818) | | (2.694) |
| *PAs10K* | −0.8811 | | −0.8408 |
| | (4.575) | | (3.489) |
| *FNPs10K* | −8.2808* | | −4.6022 |
| | (4.319) | | (4.033) |
| *Encounter10K* | 0.0168** | | 0.0070 |
| | (0.007) | | (0.006) |
| *NoHI10K* | | 0.0054 | 0.0048 |
| | | (0.004) | (0.005) |
| *Unemployed10K* | | 0.0077 | 0.0075 |
| | | (0.011) | (0.011) |
| *PerCapIncome* | | 0.0007 | 0.0006 |
| | | (0.001) | (0.001) |
| *Poverty10K* | | 0.0160*** | 0.0153*** |
| | | (0.005) | (0.005) |
| *MedHouseValue* | | −1.424e−05 | −1.277e−05 |
| | | (1.88e−05) | (2.04e−05) |
| Constant | 18.7065*** | −27.7759* | −21.6248 |
| | (1.986) | (14.763) | (17.812) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | No | No | No |
| Observations | 51 | 51 | 51 |
| $R^2$ | 0.135 | 0.488 | 0.551 |
| Adjusted $R^2$ | 0.060 | 0.431 | 0.453 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Regressions (1) - (3) are linear regressions for year 2015 without county and year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

TABLE V: Cross-Sectional Regression Year 2016

| | *Dependent variable: Pediatric ED Asthma Rates,* $Y_{i,t}$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | −0.3479 | | 1.9958 |
| | (1.864) | | (1.308) |
| *PAs10K* | −3.2253 | | −11.1112*** |
| | (3.979) | | (3.747) |
| *FNPs10K* | −3.8793 | | −4.9373*** |
| | (2.531) | | (1.731) |
| *Encounter10K* | 0.0155*** | | 0.0063 |
| | (0.006) | | (0.005) |
| *NoHI10K* | | −0.0020 | −0.0065 |
| | | (0.004) | (0.004) |
| *Unemployed10K* | | 0.0065 | −0.0054 |
| | | (0.016) | (0.017) |
| *PerCapIncome* | | −0.0003 | −0.0005** |
| | | (0.000) | (0.000) |
| *Poverty10K* | | 0.0116** | 0.0130*** |
| | | (0.005) | (0.005) |
| *MedHouseValue* | | 1.681e−05 | 1.582e−05 |
| | | (1.21e−05) | (1.14e−05) |
| Constant | 15.9164*** | 6.0660 | 23.9868** |
| | (1.573) | (10.802) | (10.228) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | No | No | No |
| Observations | 53 | 53 | 53 |
| $R^2$ | 0.151 | 0.454 | 0.588 |
| Adjusted $R^2$ | 0.081 | 0.396 | 0.502 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

Regressions (1) - (3) are linear regressions for year 2016 without county and year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

### B. County Fixed Effect Regression

A potential problem that I predicted when performing my original time-fixed effects regression analysis was the possibility of including dummy variables at a county level. However, I realized that introducing several dummy variables would only increase the noise in the model and diminish the explanatory power of my predictors. Other than implementing a time-fixed effects model that controlled for unobserved time-invariant confounders, I employed a unit-fixed effect model that adjusted for unobserved heterogeneity across counties. This robustness check test demonstrates similar but less significant results relative to the time-fixed effects approach.

Table 9 displays consistent yet less significant results across Models 1-3. The sample size employed in the county fixed effects approach is similar to the one used in the time-fixed effects analysis. Interestingly, in Model 1, when using the county-fixed effects, we find that the coefficient for the number of physician assistants is significant at the 1% level whereas the time fixed effects model demonstrates that it is not statistically significant and has a positive correlation with our response variable, which is the opposite of what I expected. In addition, we do see that the coefficients for the unemployment rate and per capita income in Model 2 are not what I had expected when comparing with our original time fixed effects model. In Model 3, we find that the coefficient on health insurance coverage is consistent with our original model; however, we yield less significant results.

TABLE VI: Cross-Sectional Regression Year 2017

| | Dependent variable: Pediatric ED Asthma Rates, $Y_{i,t}$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | −5.9470*** | | −2.9861 |
| | (2.181) | | (2.002) |
| *PAs10K* | 5.1570 | | −3.0040 |
| | (3.648) | | (4.428) |
| *FNPs10K* | −0.1139 | | 1.6211 |
| | (2.957) | | (3.126) |
| *Encounter10K* | 0.0150*** | | 0.0013 |
| | (0.005) | | (0.006) |
| *NoHI10K* | | 0.0017 | 0.0006 |
| | | (0.004) | (0.004) |
| *Unemployed10K* | | 0.0175 | 0.0074 |
| | | (0.013) | (0.014) |
| *PerCapIncome* | | −0.0001 | −0.0002 |
| | | (0.000) | (0.000) |
| *Poverty10K* | | 0.0101** | 0.0126** |
| | | (0.004) | (0.006) |
| *MedHouseValue* | | 8.803e−06 | 1.304e−05 |
| | | (1.02e−05) | (9.6e−06) |
| Constant | 16.1444*** | −1.4418 | 4.7117 |
| | (1.541) | (8.185) | (7.798) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | No | No | No |
| Observations | 53 | 53 | 53 |
| $R^2$ | 0.198 | 0.511 | 0.592 |
| Adjusted $R^2$ | 0.131 | 0.459 | 0.507 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Regressions (1) - (3) are linear regressions for year 2017 without county and year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

TABLE VII: Cross-Sectional Regression Year 2018

| | Dependent variable: Pediatric ED Asthma Rates, $Y_{i,t}$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | −4.5545* | | −1.0239 |
| | (2.333) | | (2.133) |
| *PAs10K* | 4.6714 | | −1.5186 |
| | (3.535) | | (4.055) |
| *FNPs10K* | 2.9499 | | 0.7758 |
| | (3.598) | | (3.167) |
| *Encounter10K* | 0.0083** | | 0.0018 |
| | (0.004) | | (0.004) |
| *NoHI10K* | | 0.0073 | 0.0048 |
| | | (0.009) | (0.009) |
| *Unemployed10K* | | 0.0169 | 0.0074 |
| | | (0.014) | (0.018) |
| *PerCapIncome* | | 9.781e−05 | −2.468e−05 |
| | | (0.000) | (0.000) |
| *Poverty10K* | | 0.0076 | 0.0088 |
| | | (0.005) | (0.006) |
| *MedHouseValue* | | −3.527e−06 | 7.942e−09 |
| | | (1.49e−05) | (1.52e−05) |
| Constant | 13.4454*** | −6.0134 | 1.0157 |
| | (1.396) | (9.442) | (12.104) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | No | No | No |
| Observations | 53 | 53 | 53 |
| $R^2$ | 0.133 | 0.433 | 0.455 |
| Adjusted $R^2$ | 0.062 | 0.374 | 0.343 |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Regressions (1) - (3) are linear regressions for year 2018 without county and year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

### C. Time Fixed Effect (without *Encounter10K* Variable

Another potential problem that I observed was the use of the *Encounter10K* variable in my healthcare model features. This variable was used in my primary analysis since I viewed it as conditional on the number of providers that individuals can access. However, I was worried that it could have issues of multicollinearity that may distort the coefficient estimates of the healthcare supply variables. Therefore, I employed a time-fixed effects model that removed the encounter variable to adjust for any correlation between it and other variables. This was done for Models 1 and 3 of the time fixed effect model[2].

Table 10 displays similar significant results for Model 1. However, we see that the variable measuring the number of physician assistants is now statistically significant at the 1% level whereas in the primary time fixed effects model, it was hardly significant. In Model 3, there is no major change to the model except for the statistical significance of the per capita income. There are various reasons that can be attributed to the narrative from Model 1. It could be that PAs are situated in areas with higher access, i.e. urban regions. Also, it could be that physician assistants are correlated with being in an area with more primary care encounters or that more encounters, in the aggregate, are coming from physician assistants. Although the regression model is conceptually supposed to parcel this out, the identification of the purported factors in this model is not great which makes it difficult to pinpoint the exact nature.

[2]I employed this robustness check only on the time fixed effect model since it was my primary tool of statistical analysis for reference in my results section.

TABLE VIII: Cross-Sectional Regression Year 2019

| | *Dependent variable: Pediatric ED Asthma Rates, $Y_{i,t}$* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | $-4.0559^{***}$<br>(1.497) | | $-0.4740$<br>(1.135) |
| *PAs10K* | 0.7055<br>(2.476) | | $-4.6801^{**}$<br>(1.938) |
| *FNPs10K* | $-1.2731$<br>(1.627) | | $-2.6663^{**}$<br>(1.226) |
| *Encounter10K* | $0.0140^{***}$<br>(0.004) | | 0.0035<br>(0.004) |
| *NoHI10K* | | $-0.0006$<br>(0.004) | $-0.0024$<br>(0.003) |
| *Unemployed10K* | | 0.0212<br>(0.014) | 0.0063<br>(0.013) |
| *PerCapIncome* | | $-0.0001$<br>(0.000) | $-0.0002$<br>(0.000) |
| *Poverty10K* | | $0.0094^{**}$<br>(0.004) | $0.0121^{***}$<br>(0.004) |
| *MedHouseValue* | | 7.704e−06<br>(6.94e−06) | 5.964e−06<br>(6.27e−06) |
| Constant | $13.979^{***}$<br>(1.226) | $-0.4998$<br>(7.800) | 7.7772<br>(5.992) |
| County Fixed Effects | No | No | No |
| Year Fixed Effects | No | No | No |
| Observations | 52 | 52 | 52 |
| $R^2$ | 0.183 | 0.561 | 0.690 |
| Adjusted $R^2$ | 0.114 | 0.514 | 0.624 |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Regressions (1) - (3) are linear regressions for year 2019 without county and year fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3.

TABLE IX: County Fixed Effect Results

| | *Dependent variable: Pediatric ED Asthma Rates, $Y_{i,t}$* | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| *PCPs10K* | 1.1603<br>(1.0466) | | 0.0810<br>(1.0266) |
| *PAs10K* | $-5.6002^{***}$<br>(1.6542) | | $-3.4745^{*}$<br>(1.7817) |
| *FNPs10K* | $-3.0001^{*}$<br>(1.6003) | | $-0.7913$<br>(1.2775) |
| *Encounter10K* | $-0.0042$<br>(0.0026) | | $-0.0008$<br>(0.0022) |
| *NoHI10K* | | $0.0090^{***}$<br>(0.0018) | $0.0080^{***}$<br>(0.0018) |
| *Unemployed10K* | | $-0.0116^{**}$<br>(0.0048) | $-0.0099^{*}$<br>(0.0051) |
| *PerCapIncome* | | $0.0004^{**}$<br>(0.0002) | $0.0004^{*}$<br>(0.0002) |
| *MedHouseValue* | | $-2.323e{-}05^{**}$<br>(1.138e−05) | $-2.202e{-}05^{*}$<br>(1.176e−05) |
| Constant | $20.932^{***}$<br>(1.2978) | $9.1435^{***}$<br>(2.0849) | $11.968^{***}$<br>(1.9914) |
| County Fixed Effects | Yes | Yes | Yes |
| Year Fixed Effects | No | No | No |
| Observations | 250 | 250 | 250 |
| $R^2$ | 0.1191 | 0.31 | 0.3407 |
| Adjusted $R^2$ | – | – | – |

*Note:* $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Regressions (1) - (3) are linear regressions without year but with county fixed effects. (1) denotes Model 1, (2) denotes Model 2, and (3) denotes Model 3. Also, this model was unable to produce reliable adjusted $R^2$ values and hence have been removed.

## VII. Expert Interviews

As an addendum to the quantitative findings presented above, I use this section to provide a qualitative dimension to the investigation between clinician supply and ED visit rates for asthma. This provides me the opportunity to unpack California's healthcare workforce supply and pediatric disparities through the lens of both a researcher and policy maker. Each interview focuses on how the expert's work in the field informed their understanding of the problem in addition to policy recommendations. It is important to understand that these interviews are not grounded in any specific or robust quantitative analysis but rather are meant to understand the nuances of the research question through the expert's experience and their own review of the literature.

### *A. Dr. Sunita Mutha*

Dr. Mutha is the director of the Healthforce Center at UCSF.

In my interview with Dr. Mutha, she unequivocally explained that primary care has been severely underinvested for several decades. Specifically, she identifies the perverse financial structure of reimbursement rates as being one of the primary contributors to the deficit of primary care physicians in California: primary care physicians are severely underpaid compared to their specialist counterparts. She states that most care is delivered in ambulatory care centers (outside the hospital); however, there has been an overemphasis on payments for specialist procedures and services. Dr. Mutha responded saying that other states (i.e. Oregon) have experimented with creating a price floor for paying an "X" proportion of services to primary care. Even then, this intervention has not solved the issue. She mentioned that in the absence of any intervention through state or national policies, the gap in primary care providers is only going to widen as a

TABLE X: Time Fixed Effect (without *Encounter10K*) Results

| | *Dependent variable: Pediatric ED Asthma Rates,* $Y_{i,t}$ | |
|---|---|---|
| | (1) | (2) |
| *PCPs10K* | $-2.1613^{**}$ | $-0.1995$ |
| | (0.9295) | (0.6347) |
| *PAs10K* | $4.1493^{***}$ | $-0.8983$ |
| | (1.0343) | (1.1707) |
| *FNPs10K* | $-0.7874$ | $-2.2139^{**}$ |
| | (1.6003) | (0.9662) |
| *NoHI10K* | | $0.0030^{**}$ |
| | | (0.0013) |
| *Unemployed10K* | | $0.0325^{***}$ |
| | | (0.0027) |
| *PerCapIncome* | | $-0.0004^{***}$ |
| | | (0.0001) |
| *MedHouseValue* | | 6.679e−06 |
| | | (4.63e−06) |
| Constant | $17.763^{***}$ | $11.554^{***}$ |
| | (0.5535) | (3.9297) |
| County Fixed Effects | No | No |
| Year Fixed Effects | Yes | Yes |
| Observations | 250 | 250 |
| $R^2$ | 0.0256 | 0.3955 |
| Adjusted $R^2$ | 0.0275 | 0.3504 |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Regressions (1) - (2) are linear regressions with time fixed effects. (1) denotes Model 1, (2) denotes Model 3.

TABLE XI: Expert Credentials

| **Expert Name** | **Qualifications** | **Years of Experience** |
|---|---|---|
| Dr. Sunita Mutha | MD, FACP | 15+ |
| Dr. Winston Wong | MD, MS, FAAFP | 35+ |
| Dr. Paige Bhansali | MD | 5+ |
| Dr. Janet Coffman | M.A.,M.P.P.,Ph.D | 20+ |
| Dr. José Alberto Arévalo | M.D., FAAFP | 35+ |

significant portion of the MD workforce, particularly, will enter retirement age. The number of PAs and NPs will still not be even enough to fill this gap since similar dynamics of incentives for specialty care also has an influence on these provider roles. In addition, Dr. Mutha addressed the necessity for the expansion of primary care residencies and increased compensation for primary care physicians working in rural and underserved regions of California.

### *B. Dr. Wong*

Dr. Wong is a Scholar in Residence at the UCLA Kaiser Permanente Center for Health Equity at the Fielding School of Public Health, former Medical Director for Community Benefit at Kaiser Permanente, and current board member for the California Endowment

In my interview with Dr. Wong, he explained that the economics of specialty care decisions is a big factor contributing to the deficit of primary care physicians in California. He explains that medical school students have to address the opposing nature of pursuing a career in primary care with the need to pay off their loans over the period of their adult life.

In addition, as explained by Dr. Mutha, the stigma of primary care not being a competitive trade has become embedded in medical culture which skews students to engage in other specialty areas. Dr. Wong traces the evolution of medical care from a cottage industry model to modern corporate medicine that has changed the structure of physician practice, making primary care much more difficult. He discusses the financial and economic independence of physician owned clinics to the current model where the insurance industry has embedded itself in a medical structure where hospitals now compete for doctors. He explained how highly intensive capitalism has reconfigured medicine into a predictable system of care, where patients and treatments are distilled into units of production. Through such a system, the highest-paid medical professions are protected and preventive care, i.e. primary care providers, is largely bureaucratic and standardized with little to no generation of value. This nature of trying to provide more units of service is penetrating throughout all specialties, primarily affecting primary care.

In order to improve both clinical performance and population based health outcomes, Dr. Wong advocated building an interdisciplinary medical care model where care is coordinated across a group of healthcare professionals. The hierarchy of treatment understanding is communicated between nurses, medical assistants, physicians, and other ancillary staff. Instead of having increased and intense physician intervention, health care organizations and systems can invest in a new set of different healthcare professionals that do much better in supporting patients without high costs and less non ambulatory care.

Dr. Wong also offered another solution proposing that treatments should be better tailored to individual patients (instead of cookie cutter approach) by recognizing their needs in a way that is population-based. Dr. Wong concluded his interview by positing that this is not only an issue of intensifying the number of providers around patients and families but more importantly, the quality of the engagement and care provided.

### *C. Dr. Bhansali*

Dr. Bhansali is currently a resident physician at the Boston Children's Hospital and a consultant for the Boston Consulting Group.

In my interview with Dr. Bhansali, she framed her understanding of the under-investment in primary care and pediatric health disparities through an examination of both endogenous and exogenous healthcare system factors. In terms of factors within the medical system, she found the traditional fee-for-service system used in the Medicaid program ( a federal health insurance program for low-income populations) forces many physician practices that serve this patient demographic to operate on negative financial margins. Most often, practices, clinics, or community-based hospitals that have a heavy Medicaid population load do not have the adequate resources to serve the community appropriately, on top of the fact that people are suffering from structural issues that put them at higher risk for having complex comorbidities or health outcomes. In another dimension, she explains how social determinants of health (the environmental conditions that impact an individual's quality of life outcomes) play a key role in contributing to health disparities. Dr. Bhansali remarks on the difference between discharging an English-fluent and high-income family from the hospital with adequate transportation means versus an unemployed immigrant with no vehicle to pick up their inhaler prescriptions. Unfortunately, these conditions of economic stability and community context significantly influence the health risks of populations. Dr. Bhansali bluntly replies saying that the system is set up in such a way that easily preventable procedures for an asthma attack (i.e. intubation) are compensated at a higher rate than a clinic visit for educating a patient on how to use an inhaler properly at home. Measures for establishing continuous and quality care are trumped by procedure-based cash cows.

Dr. Bhansali provided alternative payment models that are currently being experimented with to improve the quality and cost of care. She specifically mentioned the full capitation model, where physicians or healthcare provider organizations receive a risk-adjusted payment in order to support health services covered under the organization. This payment mechanism is adjusted on a timely basis depending on the clinical performance metrics of an organization, hospital, or independent medical practice. In such a model, providers are provided fixed yearly payments which enable physicians to implement strategies for robust health management in order to make it financially sustainable. For example, if a physician or group of physicians is able to generate savings from significant waste elimination and pass on some of these sayings downstream to patients, this may be a more optimal system to adopt.

### *D. Dr. Coffman*

Dr. Janet Coffman is the associate professor at UCSF and co-director of the Master of Science in Health Policy and Law Program.

In my interview with Dr. Coffman, she unequivocally stated that the demographic characteristics of California's primary care clinicians are important in understanding the deficit in the state. She explains the maldistribution of physicians is more paramount than the raw supply of providers. Although Dr. Coffman acknowledged the general difficulties of obtaining a primary care provider, there is more than a robust supply of physicians in metropolitan areas, i.e. Bay Area. Areas such as the San Joaquin Valley are burdened with chronic diseases since the ratio of primary care physicians to the general population is low; some of these areas do not meet the threshold clinician size to properly address the health challenges of the respective regions. One of the primary factors contributing to this is the aging of our clinicians, as more and more nurse practitioners are inching into retirement. This is a function of demographics, notably that most of these providers were born during the Baby Boom of the mid-twentieth century and there has not been such a demographic phenomenon to support the workforce since.

Dr. Coffman also tackled the rising cost of medical school education and the preference for specialty care. Dr. Coffman endorsed that more physicians are coming out of medical school and going into specialty or subspecialty care due to the increased return on investments in their education. She remarked that opening up new medical schools throughout the state would not solve the primary care deficit: it may get the state more practicing physicians but not necessarily primary care physicians. It is a given that procedural-based specialties earn more money. Also, there are hidden curriculum messages in medical school, including notions that intelligent students are better suited to go into specialty care rather than primary or family medicine. She even stated that these messages are conveyed between students, as well.

Dr. Coffman mentioned that the lack of team-based care is overburdening physicians with tasks that can be delegated to other healthcare professionals. She provided the historical context in the cottage industry model of medicine, mentioning that there was no "real" sense of a team or division of tasks between staff members. Physicians, then, did not have protected time to only focus on the care of the patient but also had to be concerned with the business operations of their practice, for example. She stated that moving forward, the creation of a team where the primary care provider does not have to do everything will help triage critical patients and better manage a diverse patient population. Such team members can include a health educator, pharmacist, nutritionist, physical therapist, etc.

Dr. Coffman explained the recent and growing investment in graduate medical education through the Song-Brown Program, a statewide program that provides grants to fund primary care residency programs in California. The objective is to train and place graduates in underserved regions of the state. Also, scholarship and loan repayment programs have been used to incentivize people to stay in underserved areas. She states that this investment process is recent and is contingent on the California state budget, therefore, it is hard to currently predict how it will pay off.

Other than the measures that can be taken to improve clinical delivery and institutional organization, Dr. Coffman concluded that building relationships with patients is important in managing illnesses, especially chronic diseases. How trust is built and how to maintain continuity are questions that Dr. Coffman believes to be important additions to the way we train our workforce as well. Taking the measures to create a primary care workforce that looks like and speaks like Californians is critical to establishing trusting relationships and long-term favorable health outcomes.

### *E. Dr. Arevalo*

Dr. Arevalo is the Chief Medical Officer for Sutter Independent Physicians, former Senior Vice-President for Medical Services for the Health Plan of the Redwoods, and current chair of the Latinx Physicians of California.

In my interview with Dr. Arevalo, he explained the primary issue contributing to the current deficit of primary care providers is access to care which is mediated by coverage. In the state of California, Dr. Arevalo mentioned that coverage has increased dramatically for one segment of the population: those who are eligible for Medi-Cal. Approximately, 1 out of 3 Californians are covered by Medical; however, it is not the premier coverage in the state of California and individuals still cannot find a primary care provider despite having medical coverage.

Another issue that he noted is that building provider linguistic competence is important in the patient experience, especially in the management of chronic conditions. He stated that, for example, there is a dearth of Latinx physicians practicing in California which makes it difficult to establish ethnic concordance for producing optimal health outcomes and interpersonal care. For chronic conditions, being able to create a robust communication program that addresses the nuances of a patient's condition is critical. It is then apparent that California is lacking the type of providers that can communicate concordantly in languages in underserved regions. Dr. Arevalo acknowledged that patients who identify with the physician across language or culture are more adherent to the care plan being provided.

In terms of solutions that are being implemented, he finds that the transition from the traditional fee-for-service model to global capitation or value-based care is proving fruitful. The focus from adding more procedures to now adding to the measurements of health improvement for reimbursement is important in linking the quality of clinical care directly to outcomes. Although there are problems in the current infrastructure, Dr. Arevalo is confident that organizations and medical practices are creating more sophistication in linking payment systems to value that impacts the care of the patient.

## VIII. CONCLUSION

In this study, I investigated the effect of the supply of primary care providers (i.e. PCPs) on ED visit rates for pediatric asthma in California's counties from 2015 to 2019. Although I did not look at any specific legislation or compute an event study, I examined whether there was an effect for these different healthcare provider characteristics over time. I conducted two approaches for my primary analysis: (1) pooled OLS regression and (2) time fixed-effects model.

In my Pooled OLS regression, I find a statistically significant decrease in the downstream ED visit rates for asthma amongst an increase in physician assistants by 4.6443 per 10,000 people in my combined healthcare and socioeconomic variables model. When I assessed the Californian counties only examining the healthcare features, I found there to be a significant decrease in downstream ED rates for asthma by 2.88 people (per 10,000 people) but a significant increase in ED visit rates by 0.0125 (per 10,000 people) when there was an increase in the number of counties primary care encounters. The effect of adding the socioeconomic determinants of asthma decreased the significance of physicians but enhanced the explanatory power of physician assistants. This may suggest that in a raw sense there may be a disparity in the number of physicians in primary care but when accounting for socioeconomic variables, the real issue is the maldistribution of physicians. However, although the observation regarding a reduction in ED rates by 3 people per 10,000 may be statistically significant, it is important to question whether this is significant in the aggregate. Since we are scaled per 10,000, if a county sees that hiring 1 full-time paid physician only reduces the number of asthma ED visits by 3 individuals, it would seem that this would be financially inefficient.

In the employment of the time fixed effects model, I found there to be a statistically significant decrease in ED rates by 2.97 individuals per 10,000 people with an increase in the number of primary care physicians. When accounting for the socioeconomic profiles of the different counties, the number of family nurse practitioners, and not physicians, was significant in reducing the downstream asthma ED visits. Over time, the feature of the counties that might elevate the explanatory nature of nurse practitioners is that they are more concentrated in rural and underserved areas, highlighting their important role in combating pediatric disparities in California.

Overall, I find these results corroborating many of the findings of experts in the field. Overall, the experts discussed how primary care has been underinvested in and the current ”rewarded“ system of specialty-based care has removed the financial lifeline for those practicing in primary care. Solutions such as expanding primary care residencies, increasing opportunities for telehealth, implementing value-based payment models, and establishing team-based care have been proposed and are currently being experimented with. It will take time for healthcare providers and patients to see the impact of these different policy measures.

Despite the findings above, it is important to acknowledge that this study has limitations and can be further improved. Firstly, the sample size of my data was sparse since I was limited to the individual county as the smallest geographic area that captured information on asthma ED rates. Even then, I had to remove some counties due to poor data quality which compromised the explanatory power of my regression models. As more public data on asthma ED visit rates are made available, one can use zip codes or census tracts to develop a granular frame of reference for investigation and subsequently, see the increased explanatory effect on the coefficients and features of the regression analyses. In addition, one could research the simultaneous impact of state-level dependent coverage provisions via the Affordable Care Act (2010) with the supply of primary care workforce on pediatric asthma ED rates. This would require more data on California's ED visit profile for pediatric asthma before 2015.

## References


[1] Alcala, Emanuel, Ricardo Cisneros, and John A. Capitman. Health care access, concentrated poverty, and pediatric asthma hospital care use in California's San Joaquin Valley: A multilevel approach. *Journal of Asthma* 55.11 (2018): 1253-1261.

[2] Coffman, Janet, Igoe Geyn, and Kristine Himmerick. *California's Primary Care Workforce: Current Supply, Characteristics, and Pipeline of Trainees.* Healthforce Center at UCSF, 2017, https://healthforce.ucsf.edu/sites/healthforce.ucsf.edu/files/publication-pdf/Research-Report_CA-Primary-Care-Workforce.pdf

[3] Dieu, Holly, et al. Asthma-related emergency department visits in North Carolina, 2010–2014. *North Carolina Medical Journal* 79.2 (2018): 81-87.

[4] Eum, Youngseob, EunHye Yoo, and Elizabeth Bowen. Socioeconomic determinants of pediatric asthma emergency department visits under regional economic development in western New York. *Social Science & Medicine* 222 (2019): 133-144.

[5] Halterman, Jill S., et al. The impact of health insurance gaps on access to care among children with asthma in the United States. *Ambulatory Pediatrics* 8.1 (2008): 43-49.

[6] Hasegawa, Kohei, et al. Association of insurance status with severity and management in ED patients with asthma exacerbation. *Western Journal of Emergency Medicine* 17.1 (2016): 22.

[7] Hasegawa, Kohei, et al. A population-based study of adults who frequently visit the emergency department for acute asthma. California and Florida, 2009–2010. *Annals of the American Thoracic Society* 11.2 (2014): 158-166.

[8] Largent, J., et al. Paediatric asthma hospital utilization varies by demographic factors and area socio-economic status. *Public health* 126.11 (2012): 928-936.

[9] McGovern, Colleen Marie, et al. A missed primary care appointment correlates with a subsequent emergency department visit among children with asthma. *Journal of Asthma* 54.9 (2017): 977-982.

[10] Morgan, Perri. Predicted shortages of physicians might even disappear if we fully account for PAs and NPs. *Journal of the American Academy of PAs* 32.10 (2019): 51-53.

[11] Shi, Leiyu. The impact of primary care: a focused review. *Scientifica* 2012 (2012).

## IX. Appendix

TABLE XII: California Counties Listed by Regions (Source: Census)

| Region | Counties |
|---|---|
| 1 | Butte, Colusa, El Dorado, Glenn, Lassen, Modoc, Nevada, Placer, Plumas, Sacramento, Shasta, Sierra, Siksiyou, Sutter, Tehama, Yolo, Yuba |
| 2 | Del Norte, Humboldt, Lake, Mendocino, Napa, Sonoma, Trinity |
| 3 | Alameda, Contra Costa, Marin, San Francisco, San Mateo, Santa Clara, Solano |
| 4 | Alpine, Amador, Calaveras, Madera, Mariposa, Merced, Mono, San Joaquin, Stanislaus, Tuolumne |
| 5 | Monterey, San Benito, San Luis Obispo, Santa Barbara, Santa Cruz, Ventura |
| 6 | Fresno, Inyo, Kern, Kings, Tulare |
| 7 | Riverside, San Bernardino |
| 8 | Los Angeles |
| 9 | Orange |
| 10 | Imperial, San Diego |

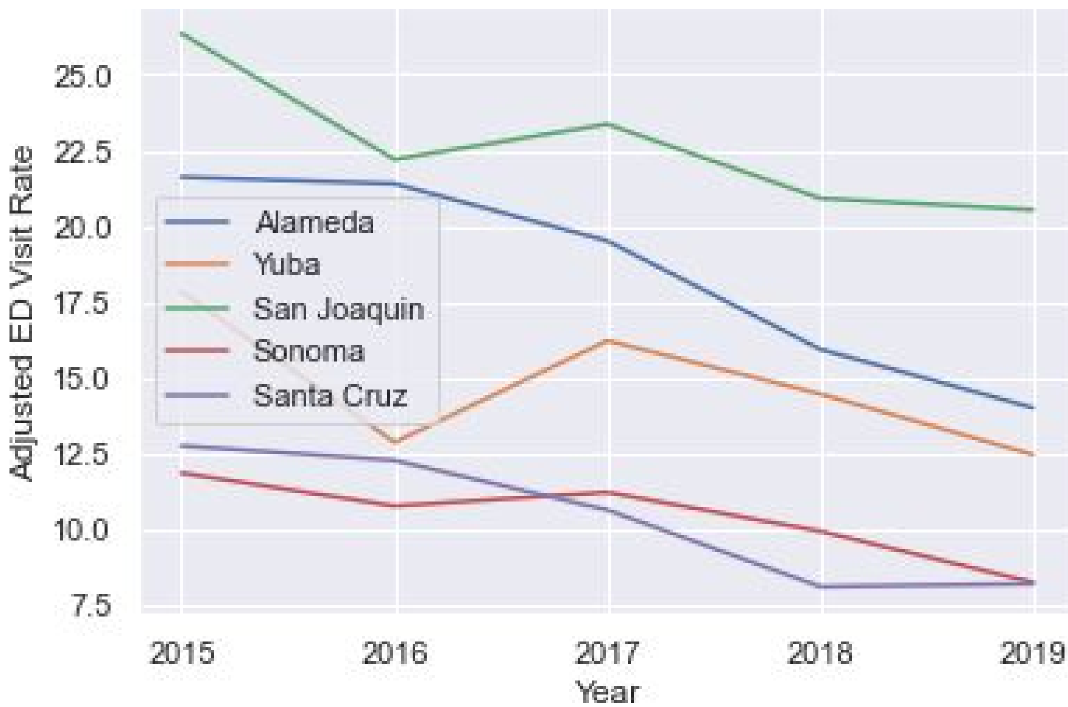



Fig. 2: Change in ED Rate per 10K in Counties from Regions 1-5 from 2015-2019

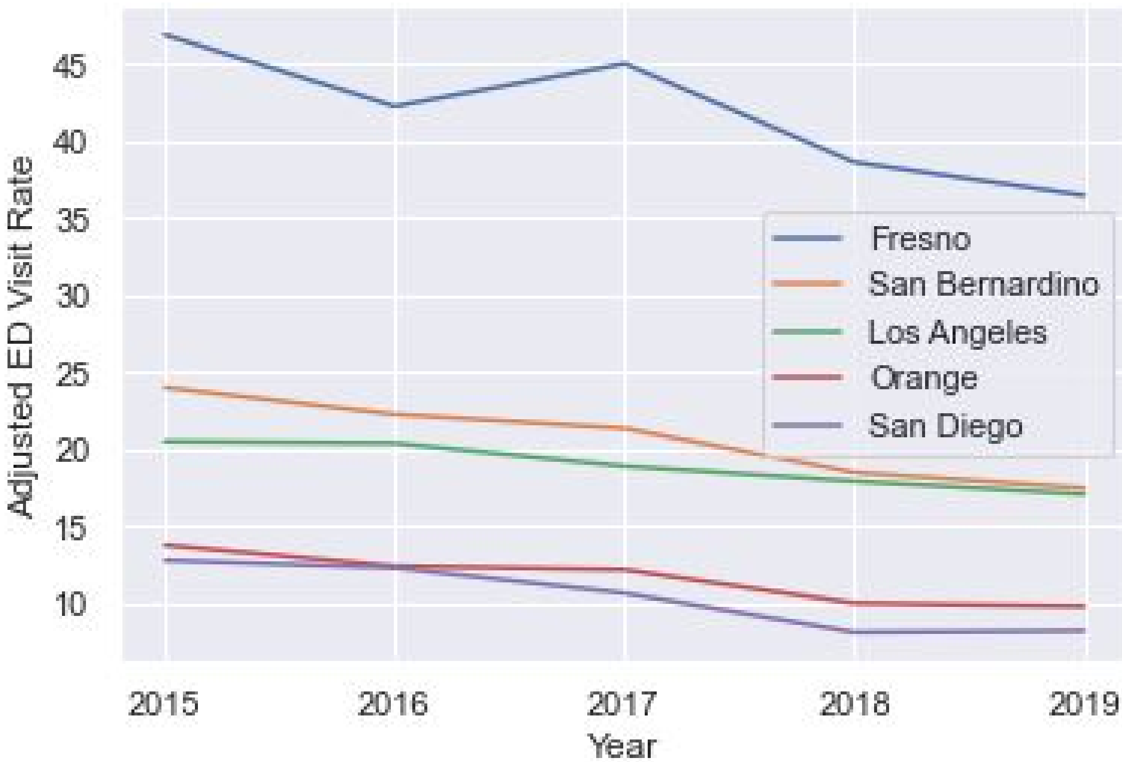



Fig. 3: Change in ED Rate per 10K in Counties from Regions 6-10 from 2015-2019

TABLE XIII: Descriptive Statistics for Socieconomic Determinants [ Mean ±SD]

| | *Year* | |
|---|---|---|
| | 2015 | 2019 |
| *Median House Value* | 327,980 ± 177,053 | 428,907.7 ± 240,980.1 |
| *Per Capita Income* | 28,201.03 ± 9057.39 | 33,821.87 ± 11,792.86 |
| *Health Uninsurance Rate* | 1,288.49 ± 339.89 | 669.06 ± 202.34 |
| *Unemployment Rate* | 488.27 ± 117.83 | 303.99 ± 80.58 |
| *Poverty Rate* | 1,206.66 ± 476.82 | 1,024.03 ± 429.20 |

*Note:* The median house value is expressed in terms of dollars ($). The per capita income is expressed in terms of per capita dollars ($).The rest of the socioeconomic characteristics are expressed as rates per 10,000 capita. SD: Standard Deviation

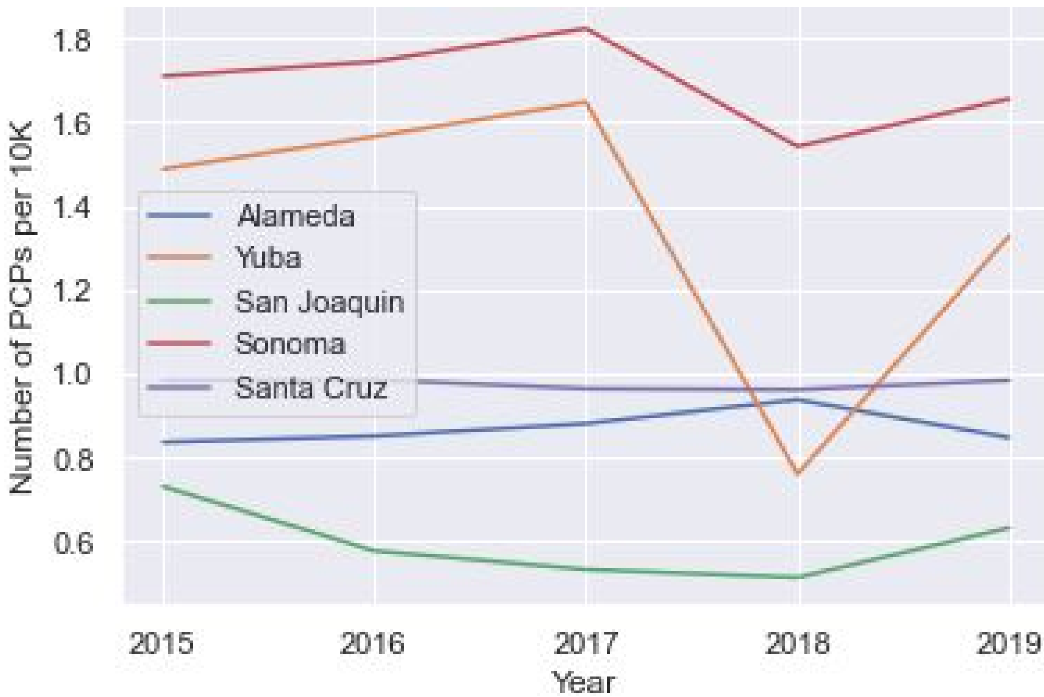



Fig. 4: Change in PCPs per 10K in Counties from Regions 1-5 from 2015-2019


Fig. 5: Change in PCPs per 10K in Counties from Regions 6-10 from 2015-2019

# Efficient Matchings on 7 Cups

Eric Gao

*Stanford University*

***Abstract*— We analyze two-sided asymmetric matching markets on 7 Cups, a site for social-emotional support where users in need of help can request to be matched with volunteer listeners who have the sole power to accept requests. The aim of this paper is to analyze user incentives to characterize what their dominant strategies are when deciding what to reveal when requesting a conversation. Listeners are treated as myopic in our model, with their only actions being to accept matches that work and terminate conversations that become undesirable for them. We find truth-telling to be a dominant strategy up to sufficiently small misrepresentations. Finally, we propose implementable suggestions to improve match outcomes.**

## I. Introduction

Mental health became an increasingly prevalent issue as a result of the COVID-19 pandemic, with rates of anxiety, stress, and depression-related symptoms increasing several-fold compared to pre-pandemic levels. Direct COVID-related stress, such as knowing infected individuals or caring for at-risk adults, as well as increased isolation due to social distancing, are both major contributors to the decline of mental-emotional well-being. Overall, around 40% of Americans suffer from some sort of mental or behavioral health condition, with over 30% being related to anxiety and depression. Furthermore, 26% of Americans have symptoms of trauma as a result from the pandemic, totaling to every other American having some sort of adverse symptom (Czeisler et al. 2020).

Fortunately, there are online resources for individuals to get social and emotional support. One such site is the focus of this paper: 7 Cups connects individuals with mental health concerns (“users”) with volunteers to confide in (“listeners”). Anyone can register to be a user and immediately start using the system. On the other hand, listeners are required to pass a short introductory course on acceptable listener conduct. Most listeners do not have professional training as mental health experts, resulting in the site mainly being used as a place for users to talk about what they are going through, opposed to being a source of advice or treatment (in fact, the listener training course explicitly instructs listeners to not give advice). While not a replacement for professional treatment, studies such as Baumel (2015) have found that 7 Cups does have an overall positive contribution to individual mental health.

As a user, requesting a conversation is as simple as clicking a button that indicates the user wants to be placed in the match-making queue. Users can also select a particular area of help from a pre-made list of topics. From the listener’s end, there is a continuously-updated list of users requesting conversations along with their desired topics of conversation. First and foremost, listeners are told to only accept conversation requests from users who choose topics the listener is comfortable talking about. Secondarily, listeners are recommended to accept requests which have been waiting in queue for longer. Once matched, a chat room is created for the pair with the conversation ending when either the listener or user wants to leave. Anonymity is preserved for both parties; any personal information is up to the user or listener to voluntarily reveal (which is strongly discouraged).

While the ease of entering a conversation helps users obtain prompt support, the low “time cost” of requesting a listener inadvertently leads to misuse of the system, which brings us to the problem this paper seeks to address. In fact, users can even request a new conversation *even while already in one*. Listeners have reported that some users misuse 7 Cups by having inappropriate conversations or asking for personal information such as social media accounts.[1] Although listeners have the option to ban specific users and give users a rating, the ease of making new accounts renders both defenses useless. Furthermore, being unable to view detailed information about a user’s conversation request before accepting makes it extremely difficult for listeners to gain information about users’ authenticity. In the following sections, we shall explore possible alterations to the current matching process to improve matching efficiency.

### A. Related Literature

In their seminal work, Gale & Shapley (1962) pioneered analysis of matching markets, demonstrating that two-sided matching with preferences leads to stable results with truth-telling as a dominant strategy. However, in the case of choosing roommates where there is one group being matched among themselves, there is no guarantee of stability. Our setting is a mix of both: while there are two groups, only users can request matches and listeners have unilateral control over which requests to accept. For a broad overview of the deferred acceptance algorithm developed by Gale and Shapely, see Roth (2007). Adachi (2003) also analyzes marriage markets but in a richer context where individuals search for matches outside of a central planner. Similar to our setting, utility is nontransferable between individuals

I would like to thank Professor Ilya Segal, Dr. Bruce Brege, Daniel Luo, and Dr. Dana Paquin for their helpful feedback and insightful comments.

[1] While there is no aggregated data on the rate of misuse, see Lin (2020) and Patterson (2020) for some colloquial reviews of the site.

who matched, with any gains being purely due to the match itself. While the pairing process on 7 Cups is not carried out by a centralized algorithm per se,[2] much of the reasoning as to why truth-telling is a dominant strategy is similar, albeit more nuanced for our specific needs here. Roth (2002) provides results and application of mechanism design in general, including applications of matching markets outside of marriage.

While most work on matching markets has been done assuming two groups which both have "veto power" over whether a match is made, there is some literature on one-sided markets. Bhalgat et al. (2011) analyzes matching mechanisms where the side not offering matches are taken to be goods (who have zero say, opposed to listeners in our setting who still have the power to accept requests). Instead of focusing on incentive compatibility, the authors analyze several different proposed mechanisms and derive expectations and bounds on efficiency given assumptions about participant utilities. Zhou (1990) provides an example as to why it is difficult to achieve stability in one-sided matching markets. Broadly speaking, one-sided matches are often unstable because less people need to have improvements for a blocking group to form. For our approach, stability is not a primary concern. Since users have no way of knowing if a "better listener" for them exists, if a matching is sufficiently good then it will automatically be stable.

A somewhat compatible setting is that of call centers or customer support departments. Similar to 7 Cups, there is one group seeking help (customers with a problem to report) and one group providing assistance (customer support workers). Furthermore, information is asymmetric yet manipulable: customers have no exposure to employees' private values, yet there is also no "rule" mandating that their customer support ticket is truthful. Finally, utility is non-transferable between agents; the only utility customers can get is from the call itself and not directly from the person who answers their request. Koole & Mandelbaum (2002) provides an overall description of how call centers operate as well as a survey of research done concerning different components of their operation. Opposed to formulating the problem as one of mechanism design and incentives, most of the literature analyzing call center operations try to optimize between worker efficiency and wait times. Borst, Mandelbaum, & Reiman (2004) analyzes the optimal number of staff needed to manage a call center. In their model, the only variation between users is the time it takes to process their request. Akşin & Harker (2003) looks at optimal server capacity distributions while assuming that there are a finite number of "topics" (in the form of a company's different products that could potentially require servicing) a user could request a conversation for. Our research deviates from this literature in two main ways. First, opposed to assuming discrete topics, users are supported over a richer type-space which allows for more nuanced revelation, making some deviations potentially beneficial. Second, we abstract away from optimizing wait times or number of people serviced and focus on inducing truth-telling. Future extensions of our model could be developed to take those two performance measures into account, but we leave that to further research.

More recently, Che & Tercieux (2001) takes a mechanism design approach to queuing theory. Instead of focusing on analyzing users' incentives when they decide what to reveal, the mechanism designer optimizes over the queuing process itself. The designer has the option to choose the entry and exit rules for joining the queue, the queuing rule which dictates how individuals in the queue are served, and what information someone in the queue has access to. Agents wanting a serve receive utility if served, but waiting imposes a time-cost on them. In this case, incentive-compatibility means people who are already in the queue do not exit before being served. Ultimately, the paper finds that a first-come first-serve mechanism where individuals waiting in line have access to zero information is optimal. Coincidentally, the 7 Cups platform satisfies those conditions: users are served first come first serve (up to existence of a listener comfortable with the longest-waiting user's desired topic of conversation) and users have no knowledge of how many other users or listeners there are.

Another setting that involves online matching between two individuals where each side has the ability to deviate from truthfully revealing their type is in online dating. There has been empirical work done analyzing deception in online dating profiles: Hancock, Toma, & Ellison finds that "the magnitude of the deceptions is usually small", which might be empirical support of our theoretical results in Section 3.2. In Adachi (2003), whether a match is successful or not is up to chance, and their work focuses more on when someone should keep a match or reject it to get a better match in the future. Hitsch (2010), estimates the utilities of different matches and compares empirical results to what would have happened if a Deferred Acceptance algorithm is used. However, neither of those two papers focus on analyzing what information participants submit to the matching mechanism and instead take match utilities or probabilities as exogenous. The focus on this paper is to endogenize the information revealed and analyze user strategies.

The remainder of the paper will be as follows. Section two formalizes the model and introduces the assumptions that will be used. Section three analyzes user strategies and derives a weakly dominant strategy for all users. While section three is only able to establish weak dominance, section four derives the extent to which users can deviate without decreasing their expected utility. Finally, section five concludes.

[2]This is another point on which this paper deviates from the majority of the mechanism design literature. In most mechanisms, agents submit a revelation and the mechanism returns some outcome or recommendation. For example, agents in auctions submit bids and get goods or their money back; agents waiting for kidney transplants submit body/kidney characteristics and the mechanism decides how to match donors with recipients. In our case, agents themselves still make the final choice on who to match with, and the mechanism itself is solely a platform for transactions.

## II. Model

### A. Topic Space

Permissible actions depend on the platform users are submitting conversation requests to. In particular, there are two parameters that matter: what sort of topics a user can input, and how specific they can be.

Let $\mathfrak{T}$ denote the set of all possible topics of conversation, and suppose there is a distance function $d$ between different topics such that $(\mathfrak{T}, d)$ forms a metric space. For an informal example of what a possible metric could be, define $d(a, b)$ to be the shortest time it takes to progress from topic $a$ to topic $b$ in a "natural" conversation (no randomly jumping from $a$ to $b$). Given any starting topic, you're already at that topic at time 0, so $d(a, a) = 0$. Similarly, all time measurements are positive by definition. If there exists some shortest conversation between topics $a$ and $b$, then it should be possible to hold that conversation "in reverse", so $d(a, b) = d(b, a)$.[3] Finally, given three topics $a, b, c$, it holds that $d(a, c) \leq d(a, b) + d(b, c)$ as the conversation which progresses from topic $a$ to $b$ and then to $c$ is still a conversation starting at topic $a$ and ending at $c$, and hence the shortest conversation from $a$ to $c$ must be at most as long as the conversation which touches upon $b$. While there may be many better distance functions, this example should capture the idea of "similarity" between topics.

As mechanism designers, we have control over what topics users are allowed to request as well as how specific their request is. The set of permissible topics is just a subset of $\mathfrak{T}$. As for specificity, let $\mu^* \in (0, \infty)$ denote a system-side constraint of how specific a user can be where the smaller $\mu^*$ is, the more specific a request can get. For example, if the site only offers multiple choice questions then $\mu^*$ will be quite large; allowing for open-ended responses would lead to a small value of $\mu^*$. Unlike users, listeners never submit any information about their preferences.

### B. Agents and Actions

There are two different agents involved: users and listeners. Going forward, let sets of users be denoted by $U$ while individual users are represented with $u$. Similarly, let sets of listeners be denoted by $L$ and individual listeners by $l$.

**Users** are characterized by their most-preferred topic combined with a measure of how tolerant they are of conversations "near" but not exactly that topic. Let $\theta_u \in \mathfrak{T}$ denote a user $u$'s true preferences on what topic they want to have a conversation about. Also, let $\epsilon_u$ denote user $u$'s tolerance value so the set of topics a user wants to talk about can be expressed as the $\epsilon_u$ neighborhood of $\theta_u$. While users have no way to choose who their assigned listener is, their role in the matchmaking process happens earlier when deciding what information to reveal.

When requesting a conversation, users submit the ordered pair $(\widehat{\theta}_u, \mu)$ where $\widehat{\theta}_u \in \mathfrak{T}$ is the user's (possibly untruthful) revealed type and $\mu \in [\mu^*, \infty)$ is the level of specificity of the user's request. We can think of the set of topics a user reveals they would be happy to talk about to be the $\mu$ neighborhood of $\widehat{\theta}_u$;[4] thus, a smaller $\mu$ implies a more specific request. Of course, 7 Cups won't literally ask users for the ordered pair, but the user has internal knowledge of that pair which is then translated into words when their request is submitted.

**Listeners** have their own preferences on what topics they are comfortable discussing. Let $\theta_l \in \mathfrak{T}$ denote a listener $l$'s most preferred topic and $\delta_l$ represent how open listener $l$ is to have different conversations outside of their most preferred topic. Each listener $l$ can be hence be represented by the ordered pair $l = (\theta_l, \delta_l) \in \mathfrak{T} \times \mathbb{R}_+$. Then, listener $l$'s range of acceptable topics is just the $\delta_l$-neighborhood of $\theta_l$. After looking at a user's request $(\widehat{\theta}_u, \mu)$, the listener accepts if[5]

$$N_\mu(\widehat{\theta}_u) \cap N_\delta(\theta_l) \neq \varnothing$$

since there then exists a mutually desirable topic and ignores the request otherwise. In other words, listeners would accept a request if there is overlap between what the listener would like to talk about and what the user reveals they would like to talk about.

### C. Timing and Payoffs

As soon as a user joins, they submit a conversation request. Then, all listeners who are currently in the system view the user's request and decide whether or not to accept. In the case of multiple acceptances, we break ties randomly. If the user's request is still unaccepted, then any listener that arrives while the user's request is active also decides whether or not to accept. However, users do not know when a listener arrives so we can assume without loss of generality that all listeners are already present when the request is made and ignore the temporal aspect of the environment.

After a match is made, the user and listener pair go through a series of topics until either they arrive at a mutually desirable topic (a topic $\theta$ is mutually desirable if $d(\theta_u, \theta) < \epsilon_u$ and $d(\theta_l, \theta) < \delta_l$) or realize that no such topic exists. In the former case, the user receives a normalized utility of one; otherwise they receive a utility of zero. While most of the literature imposes some reservation utility on individuals being matched, our formulation of the user's problem provides a good reason to abstract away from considering this. Suppose users had some reservation utility. If the (pre-normalization) utility a user receives when $d(\theta_t, \theta_u) < \epsilon_u$ is greater than their reservation utility, then

[3]For the skeptical reader, we could define a new metric by $d'(a, b) = d(a, b) + d(b, a)$ to address this possible problem. It satisfies the triangle inequality, as

$$\begin{aligned} d'(a, c) &= d(a, c) + d(c, a) \\ &\leq [d(a, b) + d(b, c)] + [d(c, b) + d(b, a)] \\ &= [d(a, b) + d(b, a)] + [d(b, c) + d(c, b)] \\ &= d'(a, b) + d'(b, c). \end{aligned}$$

Clearly, this distance function is reflexive and positive-definite.

[4]It is possible for users to not actually want to talk about topics that they reveal to be desirable. We will address the topics users actually want to discuss shortly.

[5]Notationally, let $N_a(b) = \{x \in \mathfrak{T} : d(x, b) < a\}$ denote the $a$-neighborhood of $b$.

any good match works. Otherwise, if the utility gained from a good conversation is not more than their reservation, the user would not have requested a conversation in the first place. Alternatively, we can formulate users with a higher reservation utility as just having a smaller value of $\epsilon_u$; a conversation needs to be more tailored to them to be good enough. Overall, the user's problem is either unchanged, as users have no way or knowing a conversation is good enough ex-ante, or trivial (users would never request if a good conversation is not good enough). Thus, we can define the following:

**Definition 1** (Good Listeners)**.** Suppose there is a user $u$ with parameters $(\theta_u, \epsilon_u)$. Define the set of good listeners for that user to be the set of listeners that produce a good match when paired with the user. Denote the set of good listeners by

$$L_G = \{l \in L : d(\theta_l, \theta_u) < \epsilon_u + \delta_l\}.$$

**Definition 2** (Accepting Listeners)**.** Suppose there is a user $u$ who revels $(\widehat{\theta}_u, \mu)$. Define the set of feasible matches for that user to be the set of listeners who would accept the user's request. Thus, we can denote the set of feasible matches by

$$L_A = \{l \in L : d(\theta_l, \widehat{\theta_u}) < \mu + \delta_l\}.$$

Observe that good listeners are defined with respect to the user's true parameters while accepting listeners are defined with respect to the user's revealed preferences.

### D. Assumptions

Before introducing our results, we first make two assumptions about listeners.

*Assumption* II.1 (Distribution of Listeners). Listeners' true preferences are distributed according to some measure $\nu$ such that $\nu(U) > 0$ for every open set $U \in \mathfrak{T}$.[6]

Our first assumption guarantees that as long as users want to have a conversation, then there will be potential listeners regardless of what their specific topic is. Generally, this is true on the current 7Cups site: even though some topics are more popular than others, there usually is a steady flow of listeners that clears the market for matches decently well. In particular, if there is a finite set of topics and we use the discrete metric, this assumption just states that there is a positive probability for there to be a listener willing to talk about every topic. Equivalently, we could also prune $\mathfrak{T}$ to only include topics that are covered with positive probability.

*Assumption* II.2 (Nondiscerning Listeners). Suppose there is a listener with conversation preferences $(\theta_l, \delta_l)$ who is choosing between accepting conversations from two users who reveal $(\widehat{\theta}_u, \mu_u)$ and $(\widehat{\theta}'_u, \mu'_u)$ such that the listener is a feasible match for both. Then, the listener is indifferent between accepting conversation requests from the two users, even if $\min_\theta d(\theta_l, \theta \in N_{\mu_u}(\widehat{\theta}_u)) \neq \min_\theta d(\theta_l, \theta \in N_{\mu'_u}(\widehat{\theta}'_u))$. In other words, listeners are indifferent between feasible matches.

[6]Note that $\nu$ does not necessarily have to be a probability distribution. Instead, think of $\nu$ as a measure of the flow rate of listeners into the system that are willing to discuss a particular topic.

A direct consequence of our second assumption is that listeners won't deliberately wait for "better" matches. Once again, this assumption is also typically true, as any listener who opens themselves up to have a conversation goes to help others instead of seeking the best discussion for themselves. Furthermore, if there were certain topics that a listener would prefer to indefinitely defer instead of accepting, that would be captured in the listener simply having a lesser value of $\delta_l$.

With these two assumptions in place, we present our first result that allows us to fully characterize the behavior of users.

## III. Characterization of User Strategies

We start with a technical Lemma.

**Lemma 1.** *Suppose $\nu$ is a measure satisfying $\nu(U) > 0$ for all open sets $U$. Let $A = N_\epsilon(a)$ and $B = N_\delta(b)$ for any $a, b \in \mathfrak{T}$ and $\delta, \epsilon > 0$. If $A \setminus B$ is nonempty then either $\nu(A \setminus B) > 0$ or $A \subseteq \bar{B}$ where $\bar{B}$ is the closure of $B$.*

*Proof:* Suppose $A \setminus B$ is nonempty and $A \not\subseteq \bar{B}$. Then, there exists $x \in A$ such that $x \notin B$, which implies that $d(x, b) > \delta$. As such, we can write $d(x, b) = \delta + \gamma$ for some $\gamma > 0$. As $x \in A$, we also have $d(x, a) < \epsilon$ so we can write $d(x, a) = \epsilon - \rho$ for some $\rho > 0$. Let $\xi = \min\{\rho/2, \gamma/2\}$. We now show that $N_\xi(x) \subset A \setminus B$.

Let $y \in N_\xi(x)$. First, $y \in A$: we have

$$d(y, a) < d(x, y) + d(x, a) < \rho/2 + \epsilon - \rho = \epsilon - \rho/2 < \epsilon.$$

Next, $y \notin B$: we have

$$d(y, b) > |d(x, y) - d(x, b)| > |\delta + \gamma - \gamma/2| = \delta + \gamma/2 > \delta.$$

Then,

$$\nu(A \setminus B) \geq \nu(N_\xi(x)) > 0$$

as desired.

In particular, the logic of the Lemma means that it suffice to show that unless two revelations are essentially identical, if one topic is covered by one request but not by another then we can always find a set of positive measure around that topic. Using this, we will now work towards the following full characterization of user behavior:

**Theorem 1** (Truthful and Specific Revelation)**.** *For users with true preferences $(\theta_u, \epsilon_u) \in \mathfrak{T} \times \mathbb{R}^+$, it is their (weakly) optimal strategy to reveal*

$$(\widehat{\theta}_u, \mu) = \begin{cases} (\theta_u, \epsilon_u) & \text{if } \mu^* \leq \epsilon_u \\ (\theta_u, \mu^*) & \text{if } \mu^* > \epsilon_u. \end{cases}$$

*Furthermore, it is beneficial to users for $\mu^*$ to be as small as possible.*[7]

[7]In much of the queuing theory literature, it is assumed that there is a discrete set of topics. If that's the case, we can assume a discrete metric of

$$d(a, b) = \begin{cases} 0 & \text{if } a = b \\ 1 & \text{otherwise} \end{cases}$$

While we could have formulated the revelation as

$$(\widehat{\theta}_u, \mu) = (\theta_u, \max\{\epsilon_u, \mu^*\}),$$

using the max function takes away from the intuition that users want to submit the smallest possible level of ambiguity. We break our proof into three main components: first, we'll analyze what $\widehat{\theta}$ users want to reveal, then consider users' optimal $\mu$, and finally put the two results together.

**Lemma 2** (Topic Reporting)**.** *For any fixed level of specificity $\mu \in \mathbb{R}^+$, a user $u$ has $(\theta_u, \mu)$ as their (weakly) dominant revelation. In other words, truthful reporting of desired conversation topic is efficient regardless of specificity.*

*Proof:* By assumptions II.1 and II.2, we only need to look at the size of the sets of possible good matches without worrying about any other effects, as all possible matches are equivalent to one another. Thus, it will suffice to show that for any $\widehat{\theta} \neq \theta_u$, revealing $(\widehat{\theta}, \mu)$ will only open the user up to more bad matches while not increasing the possibilities for good matches. We break things down into two cases: either $\mu \geq \epsilon_u$ or $\mu < \epsilon_u$.

**Case one:** $\mu \geq \epsilon_u$

As $\mu \geq \epsilon_u$ by assumption, $d(\theta_l, \theta_u) < \epsilon_u + \delta_l$ implies $d(\theta_l, \theta_u) < \mu + \delta_l$. Thus, $L_G \subseteq L_A$ and all listeners that would constitute good matches are willing to accept the user when they reveal $(\theta_u, \mu)$. As all good matches are already possible, it is straightforward to see that any other $\widehat{\theta}$ cannot produce any new good matches.

Next, consider the possible acceptances a user gets when revealing $(\widehat{\theta}, \mu)$ but not when revealing $(\theta_u, \mu)$. Listeners that would accept a user revealing $(\widehat{\theta}, \mu)$ come from the set

$$\widehat{L_A} = \{l \in L : d(\theta_l, \widehat{\theta}) < \mu + \delta_l\}$$

so we are looking at the listeners that are in $\widehat{L_A}$ but not in $L_A$. Let $\widehat{l} = (\widehat{\theta_l}, \widehat{\delta_l})$ be a listener in $\widehat{L_A} \setminus L_A$. As $\widehat{l} \notin L_A$, we know that $d(\widehat{\theta_l}, \theta_u) \not< \mu_u + \widehat{\delta_l}$ which implies $d(\widehat{\theta_l}, \theta_u) \geq \mu_u + \widehat{\delta_l}$. As noted before, $\mu \geq \epsilon_u$ so $d(\widehat{\theta_l}, \theta_u) \geq \epsilon_u + \widehat{\delta_l}$. Thus, $\widehat{l} \notin L_G$ and hence listener $\widehat{l}$ is not a good match. As we chose $\widehat{l}$ arbitrarily, every possible additional match gained by revealing $(\widehat{\theta_u}, \mu_u)$ instead of $(\theta_u, \mu_u)$ will be a bad match. Since $\theta_u$ and $\mu$ were chosen arbitrarily as well, the result holds for all users and all fixed levels of specificity.

**Case two:** $\mu < \epsilon_u$

Going by the same definitions of $L_G$, $L_A$, and $\widehat{L_A}$ as before, $L_A \subset L_G$ as $\mu < \epsilon_u$ implies $\mu + \delta_l < \epsilon_u + \delta_l$ while any $\widehat{L_A}$ may or may not be a subset of $L_G$. If $\widehat{L_A} \subset L_G$, then both $L_A, \widehat{L_A}$ have elements that are all good matches. Both also have the same "radius" so the number of topic areas covered are the same. Thus, the likelihood to match with a user is equivalent between the two revelations, so the user is indifferent between the two strategies. On the other hand, if $L'_A$ is not a subset of $L_G$, then there must be elements of $L'_A$ that are bad matches. As the size of $L_A$ and $L'_A$ are the same, then this implies that $L'_A$ must have less good matches, which means that revealing $(\widehat{\theta}, \mu)$ is strictly dominated by revealing $(\theta_u, \mu)$.

In particular, the dominance is weak, as it is possible for small deviations in most preferred topics to not lead to losses in efficiency. If the fixed level of specificity is less than the user's tolerance value, then deviations of magnitude less than the difference between specificity and tolerance will not reduce efficiency. We will address these questions later. For now, we move on to analyzing how users choose how specific to be.

**Lemma 3** (Specificity)**.** *Suppose $\mu, \mu' \in [\epsilon_u, \infty)$ with $\mu < \mu'$. For a user $u = (\theta_u, \epsilon_u)$, reporting $(\theta_u, \mu)$ is (strictly) preferred to reporting $(\theta_u, \mu')$. Thus, specificity is preferred by users who report their true desired topic of conversation.*

*Proof:* We'll use a similar approach to the last lemma. The set of listeners that produce good matches does not depend on the level of reported specificity, so $L_G$ is as previously defined. Then, the set of accepting listeners when the user reveals $(\theta_u, \mu)$ is still $L_A$. Finally, the set of accepting listeners when the user reveals $(\theta_u, \mu')$ is

$$L'_A = \{l \in L : d(\theta_l, \theta_u) < \mu' + \delta_l\}.$$

By assumption, $\epsilon_u \leq \mu < \mu'$ so $l \in L_G$ implies

$$d(\theta_l, \theta_u) < \epsilon_u + \delta_l \leq \mu + \delta_l < \mu' + \delta_l.$$

Thus, $L_G \subseteq L_A \subsetneq L'_A$ and revealing $\mu'$ opposed to $\mu$ does not give the user any more chances to get a good match.

To see why the preference is strict, consider listeners $l \in L'_A \setminus L_A$. We know that such a listener must exist with nonzero probability as the two sets are not equal and listeners are distributed uniformly. As $L_G \subseteq L_A$, we know $L_G \cap (L'_A \setminus L_A) = \varnothing$ so $l$ must be a bad match. Thus, revealing $\mu'$ opposed to $\mu$ exposes the users to additional bad matches, which is clearly undesirable.

Now, we prove Theorem 1 by showing that given any deviant revelation

$$(\theta', \mu') \neq \begin{cases} (\theta_u, \epsilon_u) & \text{if } \mu^* \leq \epsilon_u \\ (\theta_u, \mu^*) & \text{if } \mu^* > \epsilon_u, \end{cases}$$

it will perform worse than $(\theta, \mu)$ as described in the Theorem and in the RHS above.

*Proof:* First, given $(\theta', \mu')$, suppose $\theta' \neq \theta_u$. Holding $\mu'$ fixed, we know that $(\theta_u, \mu')$ is (weakly) preferred compared to $(\theta, \mu')$ by Lemma 2. Direct application of our first lemma already takes care of half the Theorem.

All we need to do now is to show that the user would prefer revealing specificity of $\max\{\epsilon_u, \mu^*\}$.[8] We will have to do a bit of casework here, both in terms of how $\epsilon_u$ and $\mu^*$ are related as well as how $\max\{\epsilon_u, \mu^*\}$ and $\mu'$ are related. Luckily, most of the cases are directly dealt with by Lemma 3 so things do not get too messy.

**Case one:** Suppose $\epsilon_u \leq \mu^*$. Then, $\max\{\epsilon_u, \mu^*\} = \mu^*$ and any permissible $\mu'$ must be in the interval $[\mu^*, \infty)$. Thus, if $\mu' \neq \mu^*$ then $\mu^* < \mu'$ so we can directly apply Lemma 3

and set $\mu^*, \epsilon_u < 1$. Then, this theorem states that the only topic in the user's revelation will be the single topic they want help with.

[8] Recall that this is equivalent to the way we formulated $\mu$ in the theorem.

to conclude that revealing $(\theta_u, \mu^*) = (\theta_u, \max\{\epsilon_u, \mu^*\})$ is strictly preferred to revealing $(\theta', \mu')$. This finishes case one.

**Case two:** Suppose $\mu^* < \epsilon_u$. In this case, $\max\{\epsilon_u, \mu^*\} = \epsilon_u$. Here, the relative positions of $\mu'$ and $\epsilon_u$ matter. Assuming $\mu' \neq \epsilon_u$ (as otherwise, we already arrive at the desired result), we either have $\mu^* < \epsilon_u < \mu'$ or $\mu^* \leq \mu' < \epsilon_u$. We consider each of these sub-cases.[9]

**Case 2a:** Suppose $\mu^* < \epsilon_u < \mu'$. Here, $\epsilon_u < \mu'$ so direct application of Lemma 3 gives the desired result, similar to Case one.

**Case 2b:** Suppose $\mu^* < \mu' < \epsilon_u$. Unfortunately, we now need to do a bit of work. Here, $\max\{\epsilon_u, \mu^*\} = \epsilon_u$ so if the user reveals $(\theta_u, \max\{\epsilon_u, \mu^*\}) = (\theta_u, \epsilon_u)$, the set of feasible matches is

$$L_A = \{l \in L : d(\theta_l, \theta_u) < \epsilon_u + \delta_l\}.$$

This is equivalent to the set of listeners that constitute good matches. Thus, the user has access to all possible good matches while exposing themselves to no bad matches so we know that truthful and specific revelation is at least as good as any other revelation.

Now, consider the set of listeners that would accept a user's request if they revealed $(\theta_u, \mu')$ denoted by

$$L'_A = \{l \in L : d(\theta_l, \theta_u) < \mu' + \delta_l\}.$$

As $\mu' < \epsilon_u$ by assumption, we know that $L'_A \subsetneq L_G$. Thus, $L_G \setminus L'_A$ is non-empty and consists of listeners who produce good matches that would have accepted the user's request if they revealed $(\theta_u, \epsilon_u) = (\theta_u, \max\{\epsilon_u, \mu^*\})$, but not if the user reveals $(\theta_u, \mu')$. Thus, the user does lose out if they do not accurately reveal specificity, which implies the preference for revealing specificity truthfully is strong.

To summarize, we have shown that given any user $u = (\theta_u, \epsilon_u)$ and hypothetical revelation $(\theta'_u, \mu')$, revealing $(\theta'_u, \mu')$ is (weakly) dominated by revealing $(\theta_u, \mu')$ which is in turn (strongly) dominated by revealing $(\theta_u, \max\{\epsilon_u, \mu^*\})$. This is equivalent to saying that revealing

$$(\widehat{\theta}_u, \mu) = \begin{cases} (\theta_u, \epsilon_u) & \text{if } \mu^* \leq \epsilon_u \\ (\theta_u, \mu^*) & \text{if } \mu^* > \epsilon_u \end{cases}$$

is the optimal strategy. In other words, untruthful topic and untruthful specificity reporting is weakly dominated by truthful topic but untruthful specificity reporting, which is in turn strongly dominated by truthful topic and truthful specificity reporting.

Moving onto the final statement of the proof: we'll show that if the system-side lower bound on specificity is $\mu^\star < \mu^*$, then users will either be unaffected or benefited. For users with $\mu^* \leq \epsilon_u$, we have $\mu^\star < \epsilon_u$ and thus $\max\{\epsilon_u, \mu^\star\} = \epsilon_u$ so those users have their situation remain unchanged. On the other hand, for users with $\epsilon_u < \mu^*$, we know that $\max\{\epsilon_u, \mu^\star\} < \mu^*$ so revealing $(\theta_u, \max\{\epsilon_u, \mu^\star\})$ is strictly preferred to revealing $(\theta_u, \max\{\epsilon_u, \mu^*\})$ for them by Lemma 3.

[9]The case of $\mu' < \mu^* < \epsilon_u$ cannot happen as $\mu^*$ is a system-side limitation and thus playing $\mu' < \mu^*$ for their level of specificity is not allowed in the first place.

Users will also not have to worry about other users taking up potential listeners when the system-side constraint on specificity decreases, as every user's neighborhood of acceptable conversations either stays the same or shrinks. In fact, allowing for increased specificity actually decreases competition between users, a desirable result regardless of increasing efficiency.

Theorem 1 is a very nice result, as not only does it show that truth-telling is best for users, but also says that being as specific as possible is optimal as well. Furthermore, it presents a fairly straightforward change for matching mechanisms to implement: allowing for more specificity when submitting requests can only help. As briefly touched upon earlier, this could take the form of making free-response questions available to the user. In fact, making this change would not substantially change the user experience apart from inducing better matches: currently, users who have been waiting in queue for a while sometimes start typing out what they are going through before a listener gets to the chat room. For users with urgent need of someone to talk to, this could mean going through less listeners before finding someone who works.

An interesting question for further investigation is the possibility of a "hybrid/dynamic" revelation mechanism where users can submit any baseline level of specificity, but update it afterwards. Users could initially submit a general multiple-choice style request in some topic area to immediately be put in the pool and gradually update it while waiting for a listener to accept. Intuitively, such a mechanism would not lead to any additional wait times (as the process is equivalent to the current process from the user's perspective) while improving matching outcomes based on the analysis done previously. The conclusion involves a more in-depth analysis of such a mechanism.

## IV. Analysis of Deviations

While we were only able to show that revealing exact truthfulness is a weakly dominant strategy, we can put bounds on how much a user can deviate from truth-telling without losing out on efficiency. We'll first analyze a sufficient condition for a deviation to not be sub-optimal before moving on to finding a necessary condition. Finally, we consider which deviations make it impossible for a user to get a good match. Going forward, let $\Delta\theta_u = d(\theta_u, \widehat{\theta_u})$ denote the magnitude of the user's deviation.

**Lemma 4** (Sufficient Condition for Efficient Deviations)**.** *For a user $u = (\theta_u, \epsilon_u)$, deviation from revealing $(\theta_u, \max\{\epsilon_u, \mu^*\})$ to $(\widehat{\theta_u}, \max\{\epsilon_u, \mu^*\})$ does not sacrifice efficiency for the user if*

$$\Delta\theta_u < \max\{\epsilon_u, \mu^*\} - \epsilon_u.$$

Intuitively, as long as the magnitude of the deviation is less than the distortion caused by the system itself not allowing you to be fully specific, then efficiency is not sacrificed.

*Proof:* To avoid casework, let $\max\{\epsilon_u, \mu^*\} = s \in [\epsilon_u, \infty)$ be the user's optimal level of specificity (we know this is optimal by Lemma 3). As before, let

$$\widehat{L_A} = \{l \in L : d(\theta_l, \widehat{\theta_u}) < s + \delta_l\}.$$

represent the set of listeners that accepts when the user deviates to $(\widehat{\theta_u}, s)$. As $s \in [\epsilon_u, \infty)$ implies $s \geq \epsilon_u$, we know $L_G \subseteq L_A$ so any efficient deviation must satisfy $L_G \subseteq \widehat{L_A}$. On the other hand, for a deviation to be inefficient, there must exist some listener $l$ such that $l \in L_G$ but $l \notin \widehat{L_A}$. A deviation could be inefficient if $|\widehat{L_A}| > |L_A|$ but $|\widehat{L_A} \cap L_G| = |L_A \cap L_G|$ because the user would open themselves up to more potentially bad matches due to the uniform distribution of listeners. However, we ignore this case because the size of our neighborhood is constant between the two as $s$ is invariant with respect to changes in revealed $\theta$.

Suppose $\Delta\theta_u < s - \epsilon_u$. We will show $L_G \subseteq \widehat{L_A}$ by demonstrating that $l \in L_G$ implies $l \in \widehat{L_A}$. Let $l = (\theta_l, \delta_l)$ be some listener in $L_A$. By definition, $d(\theta_l, \theta_u) < \delta_l + \epsilon_u$. Applying the triangle inequality,

$$\begin{aligned} d(\theta_l, \widehat{\theta_u}) &\leq d(\theta_l, \theta_u) + d(\theta_u, \widehat{\theta_u}) \\ &< (\delta_l + \epsilon_u) + (s - \epsilon_u) \\ &= \delta_l + s \end{aligned}$$

so $l \in \widehat{L_A}$ as desired.

The converse is a lot more tricky to show. We will prove the result taking $\mathfrak{T} \subset \mathbb{R}^n$ with the usual Euclidean metric, but the result should generalize to sufficiently nice metric spaces.

**Lemma 5** (Necessary Condition for Efficient Deviations)**.** *For a user $u = (\theta_u, \epsilon_u)$, deviation from revealing $(\theta_u, \max\{\epsilon_u, \mu^*\})$ to $(\widehat{\theta_u}, \max\{\epsilon_u, \mu^*\})$ does not sacrifice efficiency only if*

$$\Delta\theta_u < \max\{\epsilon_u, \mu^*\} - \epsilon_u.$$

*Proof:* Fix some listener tolerance value $\delta_l$; if the Lemma holds for all listeners with that tolerance value then it must also hold for all listeners. Also, let $s = \max\{\epsilon_u, \mu^*\}$ similar to the last proof. Suppose $\Delta\theta_u > s - \epsilon_u$. We will show that there exists a listener who is a good match for the user but does not accept the user's request if they deviate and reveal $(\widehat{\theta_u}, \max\{\epsilon_u, \mu^*\})$. Going forward, let

$$N_{s+\delta_l}(\widehat{\theta_u}) = \{\theta \in \mathfrak{T} : d(\theta, \widehat{\theta_u}) < s + \delta_l\}$$

denote the set of topics $\theta$ such that a listener with topic $\theta$ as their truly preferred topic and tolerance level $\delta_l$ will accept a user revealing $(\widehat{\theta_u}, s)$.[10] Observe that for each fixed $\delta_l$, the set of ordered pairs $(\theta, \delta_l)$ such that $\theta \in N_{s+\delta_l}(\widehat{\theta_u})$ forms a subset of

$$\widehat{L_A} = \{l \in L : d(\theta_l, \widehat{\theta_u}) < s + \delta_l\}.$$

Thus, the problem of finding $l \in L_G$ such that $l \notin \widehat{L_A}$ for fixed $\delta_l$ is equivalent to finding $\theta \in \mathfrak{T}$ such that $d(\theta, \theta_u) < \delta_l + \epsilon_u$ but $\theta \notin N_{s+\delta_l}(\widehat{\theta_u})$.

For our first possible point, consider $\theta_u$, which is clearly in $L_G$ given any fixed $\delta_l$ (your exact topic always works regardless of the listener's tolerance: $d(\theta_u, \theta_u) = 0 < \epsilon_u + \delta_l$for all $\delta_l$). If $\theta_u$ is not in $N_{s+\delta_l}(\widehat{\theta_u})$, then that produces a point with the desired properties, and we are done. Thus, we can assume $\theta_u \in N_{s+\delta_l}(\widehat{\theta_u})$.

Now, define the topic $\theta^* \in \mathfrak{T}$ satisfying $d(\theta^*, \widehat{\theta_u}) = s + \delta_l$ to be the topic that minimizes $d(\theta^*, \theta_u)$:

$$\begin{aligned} \theta^* = \text{argmin } & d(\theta^*, \theta_u) \\ \text{s.t. } & \theta^* \in \mathfrak{T}; \\ & d(\theta^*, \hat{\theta_u}) = s + \delta_l. \end{aligned}$$

In other words, $\theta^*$ is the boundary point of $N_{s+\delta_l}(\widehat{\theta_u})$ that is closest to $\theta_u$. We claim that $d(\theta^*, \theta_u) < \delta_l + \epsilon_u$ as follows.

Consider the line formed by the set of topics $\{\theta : d(\theta_u, \theta) + d(\theta, \widehat{\theta_u}) = d(\theta_u, \hat{\theta_u})\}$. We can extend this line past $\theta_u$ to include all $\theta$ that satisfies $d(\theta, \theta_u) + d(\theta_u, \widehat{\theta_u}) = d(\theta, \widehat{\theta_u})$. Now, we consider the intersection of this line and the boundary of $N_{s+\delta_l}(\widehat{\theta_u})$. Let $\theta'$ be the topic such that

$$\theta' \in \{\theta : d(\theta, \theta_u) + d(\theta_u, \widehat{\theta_u}) = d(\theta, \hat{\theta_u})\} \cap \{\theta : d(\widehat{\theta_u}, \theta) = s + \delta_l\}.$$

We know such $\theta'$ exists, as the line can be extended arbitrarily far away from $\widehat{\theta_u}$ while the boundary is bounded. Thus, $\theta'$ is a point on the boundary of $N_{s+\delta_l}(\widehat{\theta_u})$ satisfying $d(\theta', \theta_u) + d(\theta_u, \widehat{\theta_u}) = s + \delta_l$ which implies

$$\begin{aligned} d(\theta_u, \theta') &= (s + \delta_l) - d(\theta_u, \widehat{\theta_u}) \\ &< (s + \delta_l) - (s - \epsilon_u) \\ &= \delta_l + \epsilon_u. \end{aligned}$$

where the second line holds because $d(\widehat{\theta_u}, \theta_u) > s - \epsilon_u$ by assumption. Then, as $\theta^*$ minimizes $d(\theta_u, \theta^*)$ subject to $d(\theta^*, \widehat{\theta_u}) = s + \delta_l$, combined with the fact that $\theta'$ satisfies the latter requirement, we know that

$$d(\theta_u, \theta^*) \leq d(\theta_u, \theta') < \epsilon_u + \delta_l$$

as desired. Going forward, let $d(\theta_u, \theta^*) = \lambda < \epsilon_u + \delta_l$.

We finish by showing the existence of a topic $\tilde{\theta}$ such that a listener with preferences $(\tilde{\theta}, \delta_l)$ forms good match but does not accept the user's un-truthful request. Let $\gamma = \frac{\epsilon_u + \delta_l - \lambda}{2}$ and consider the neighborhood

$$N_\gamma(\theta^*) = \{\theta : d(\theta, \theta^*) < \gamma\}.$$

This neighborhood is a subset of $N_{\epsilon_u+\delta_l}(\theta_u)$ as if $\theta \in$

[10]Using previous terminology, this would be almost the same as $\widehat{L_A}$. However, the key difference is that $\widehat{L_A}$ is a set of ordered pairs $(\theta_l, \delta_l)$ that allows for varied values of $\delta_l$, but $N_{s+\delta_l}(\widehat{\theta_u})$ is a set of solely topics given a fixed value of $\delta_l$. This distinction becomes important later on, as we need the stronger property of this set being a neighborhood of our metric space. For an interesting connection, note that

$$\widehat{L_A} = \bigcup_{\delta_l \in \mathbb{R}} \left(\theta \in N_{s+\delta_l}(\widehat{\theta_u}), \delta_l\right).$$

$N_\gamma(\theta^*)$ then

$$\begin{aligned} d(\theta,\theta_u) &\le d(\theta,\theta^*) + d(\theta^*,\theta_u) \\ &\le \gamma + \lambda \\ &= \frac{\epsilon_u + \delta_l - \lambda}{2} + \frac{2\lambda}{2} \\ &= \frac{\epsilon_u + \delta_l}{2} + \frac{\lambda}{2} \\ &< \frac{\epsilon_u + \delta_l}{2} + \frac{\epsilon_u + \delta_l}{2} \\ &= \epsilon_u + \delta_l \end{aligned}$$

so $\theta \in N_\gamma(\theta^*)$ implies $\theta \in N_{\epsilon_u+\delta_l}(\theta_u)$. Then, $\theta^*$ is a boundary point of $N_{S+\delta_l}(\widehat{\theta_u})$ so every neighborhood of $\theta^*$ has a point in and a point not in $N_{S+\delta_l}(\widehat{\theta_u})$. As $N_\gamma(\theta^*)$ is a neighborhood of $\theta^*$, we can conclude that there must be a point in $N_\gamma(\theta^*)$ that is not in $N_{S+\delta_l}(\widehat{\theta_u})$. Let $\tilde{\theta}$ be that point (if there are many, choose one arbitrarily). Combining this with the fact that $N_\gamma(\theta^*) \subset N_{\epsilon_u+\delta_l}(\theta_u)$, we can conclude that a listener with topic preference of $\tilde{\theta}$ lies in the neighborhood which constitutes good matches, but is outside of the neighborhood that defines the set of listeners which accepts the user's deviated revelation. This completes our proof.

For a visualization of each of the components of this proof, the following figure illustrates the proof in $\mathbb{R}^2$.

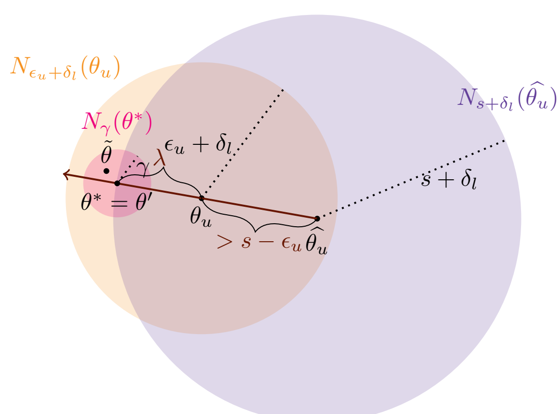



Fig. 1: Lemma 5 Illustrated in $\mathbb{R}^2$

So far, we've defined a revelation to be efficient if it does not decrease the probability of being matched over all possible listeners. Thus, even if efficiency is lost, it does not mean that there is a zero chance for a match. We now consider the conditions for deviations in which no match is possible.

**Lemma 6** (Incompatible Revelations)**.** *Fix some $\delta_l \in \mathbb{R}^+$. For a user $u = (\theta_u, \epsilon_u)$, deviation from revealing $(\theta_u, \max\{\epsilon_u, \mu^*\})$ to $(\widehat{\theta_u}, \max\{\epsilon_u, \mu^*\})$ results in the user being unable to have a good match with any listener with tolerance value less than or equal to $\delta^*$ if and only if*

$$\Delta\theta_u \ge \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^*.$$

*Proof:* Suppose the user deviates and reveals

$$(\widehat{\theta_u}, \max\{\epsilon_u, \mu^*\})$$

such that

$$d(\theta_u, \widehat{\theta_u}) > \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^*.$$

We will show that for every listener $l$ with parameters $(\theta_l, \delta_l)$ where $\delta_l \le \delta^*$, the listener cannot both be a feasible match and a good match at the same time.

Towards a contradiction, suppose the listener was is both willing to accept the user's request and produces a good match. For the listener to be a good match, it must hold that

$$d(\theta_l, \theta_u) < \epsilon_u + \delta_l$$

while the listener accepting the request implies that

$$d(\widehat{\theta_u}, \theta_l) < \max\{\epsilon_u, \mu^*\} + \delta_l.$$

Applying the Triangle Inequality, we know

$$\begin{aligned} d(\theta_u, \widehat{\theta_u}) &\le d(\theta_u, \theta_l) + d(\theta_l, \widehat{\theta_u}) \\ &< (\epsilon_u + \delta_l) + (\max\{\epsilon_u, \mu^*\} + \delta_l) \\ &= \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta_l \\ &\le \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^*. \end{aligned}$$

Putting the first and last lines together, we get $d(\theta_u, \widehat{\theta_u}) \le \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^*$ which contradicts our assumption that $d(\theta_u, \widehat{\theta_u}) = \Delta\theta_u \ge \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^*$. Thus, the user cannot acquire any good matches with listeners with listeners having a tolerance value less than or equal to $\delta^*$ if

$$\Delta\theta_u \ge \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\max_{l\in L}\delta^*$$

as desired.

Conversely, we'll show that the bound cannot be weakened. If we only required

$$\Delta\theta_u \ge (\epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^*) - \xi$$

for some $\xi > 0$, then it could be the case that there exists a listener $l$ with parameters $(\theta_l, \delta_l)$ which satisfies

$$\epsilon_u + \delta^* - \frac{\xi}{4} < d(\theta_l, \theta_u) < \epsilon_u + \delta^*$$

and

$$\max\{\epsilon_u, \mu^*\} + \delta^* - \frac{\xi}{4} < d(\theta_l, \widehat{\theta_u}) < \max\{\epsilon_u, \mu^*\} + \delta^*.$$

Thus, the listener is both a good match and a feasible match. To see why this does not violate the Triangle Inequality, observe that

$$\begin{aligned} &\left(\epsilon_u + \delta^* - \frac{1\xi}{4}\right) + \left(\max\{\epsilon_u, \mu^*\} + \delta^* - \frac{\xi}{4}\right) \\ =&\epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^* - \frac{\xi}{2} \\ >&\epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\delta^* - \xi \end{aligned}$$

which implies that

$$d(\theta_u, \theta_l) + d(\theta_l, \widehat{\theta_u}) \ge d(\theta_u, \widehat{\theta_u})$$

as required.

Now, we are ready to fully characterize a user's range of possible deviations. As before, we hold the user's revealed level of specificity to be fixed at $\max\{\epsilon_u, \mu^*\}$ as being less specific will never help the user by previous analysis.

**Theorem 2** (Characterization of User Deviations)**.** *Suppose a user with parameters $(\theta_u, \epsilon_u)$ is a user and assume the system-side limitation on specificity is $\mu^*$. Then,*

- *if $0 \leq \Delta\theta_u < \max\{\epsilon_u, \mu^*\} - \epsilon_u$, no efficiency is lost;*
- *on the interval $\max\{\epsilon_u, \mu^*\} - \epsilon_u < \Delta\theta_u < \max\{\epsilon_u, \mu^*\} + 2\min_{l\in L}\delta_l$ the user's revelation is sub-optimal;*
- *on the interval $\max\{\epsilon_u, \mu^*\} + 2\min_{l\in L}\delta_l < \Delta\theta_u < \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\max_{l\in L}\delta_l$, efficiency is lost as $\Delta\theta_u$ increases;*
- *the user will not get any matches if $\Delta\theta_u \geq \epsilon_u + \max\{\epsilon_u, \mu^*\} + 2\max\limits_{l\in L}\delta_l$.*

*Proof:* Points one and two comes from Lemma 4 and Lemma 5. To see why the third point is true, apply the converse direction of Lemma 6 to increasing values of $\delta_l$ within the range $[\min\delta_l, \max\delta_l]$. The fourth point follows directly from the forward direction of Lemma 6.

In the region $\max\{\epsilon_u, \mu^*\} - \epsilon_u < \Delta\theta_u < \max\{\epsilon_u, \mu^*\} + 2\min_{l\in L}\delta_l$, deviating more cannot produce more matches. Any listener who would accept the user's request if the user increased their deviation must have topic preferences (weakly) further from the user's true topic preference than listeners who would have accepted the user's initial deviation. Thus, they cannot be a good match. However, it is an open question as to how exactly efficiency changes; whether it remains constant (at some sub-optimal level) or decreases as $\Delta\theta_u$ increases. Either way, the overall characterization that after a certain threshold, deviating further results in a monotonic decreases in efficiency until no matches are possible still holds. The following graph gives a rough display of relative efficiency as a function of deviation magnitude. Clearly, there is no evidence that efficiency decreases linearly; the graph is not to scale either.

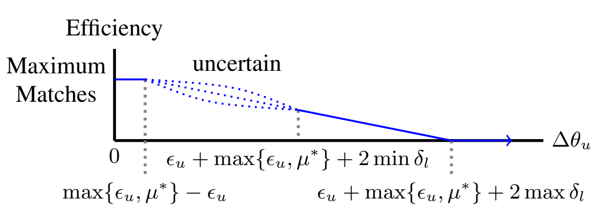



Fig. 2: Relative Efficiency Versus Deviation

Even though our current model is unable to induce perfect truth-telling as a strictly optimal strategy, when it comes to real-world implementation, having range of efficient deviations as dictated in Theorem 2 might be better than requiring users to be spot-on. When requesting a conversation, users might not be completely certain as to what exactly they want, or even how to put what they want to talk about into writing. Thus, having a region of efficient deviations gives users some leeway to not perfectly describe their situation.[11]

## V. Conclusion and Further Research

We analyze user incentives on 7 Cups and show that truthful revelation of topic preferences is a weakly dominant strategy, while best reporting of tolerance values is strictly dominant. Thus, we can conclude that the current matching mechanism is incentive compatible with truthful reporting. Furthermore, we are also able to put bounds on how much of a deviation users can have before losing out on efficiency. In terms of immediately implementable improvements, allowing users to submit more detailed requests, thus lowering the lower bound on specificity, would improve matching outcomes while also decreasing competition between users. Allowing users to type out their desired topics of conversation in a free response text box would address this issues.

Empirically verifying these results would be very interesting. However, it would be difficult to gather data, and even more so to convert that data into results concerning how users' preferences relate to what they reveal. Perhaps using a methodology similar to Hitsch, Hortaçsu, and Ariely (2010), which empirically compares the results of dating app matches to what would have happened in a centralized deferred acceptance algorithm, could be fruitful. Interestingly enough, real-life matches were comparable to optimized deferred acceptance matches in terms of efficiency. Empirical data might also shed light on behavioral questions: if there are certain trends for a particular type of deviation, we might be able to expand our model to have more explanatory or predictive power instead of purely being theoretical.

There are two main open questions for further research. First is expanding the model to encompass different types of users. In particular, there could be disingenuine users, who use the site for conversation but not for mental health related purposes. These users can be modeled by having topic preferences outside of $\mathfrak{T}$ and hope to come across a listener who also happens to be open to talking about other topics. Second, there are some malicious users who use the site with nefarious intent. For instance, some users mess with volunteer listeners or seek personal information. How can the platform be further improved to account for these two types of users?

Another interesting question would be to investigate the feasibility of a dynamic revelation system. Currently, users who are waiting in queue have the option to start typing their thoughts into an empty chat room. Furthermore, when listeners accept a request the conversation usually starts in a very similar fashion each time ("Thanks for visiting 7 Cups! What would you like to talk about today?"). Instead of leaving listeners to ask for this baseline information,

[11] An interesting parallel is the result found in Roth (1984): Nash equilibria in the marriage problem are stable even if there are some non-cooperative deviations. While there is no obvious underlying similarity between the marriage problem and the situation studied in this paper, further work can be done to identify a broad class of games for which deviations do not sacrifice efficiency and/or stability.

the interface could instead prompt users to share while waiting. In turn, listeners could see what users have typed to better inform their decisions. This would increase request specificity, which we have shown to be beneficial to all users and listeners.

## REFERENCES


[1] Czeisler, M., Lane, R., Petrosky, E., Wiley, J., Christensen, A., Njai, R., Rajaratnam, S. (08 2020). Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic - United States, June 24-30, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69, 1049–1057. doi:10.15585/mmwr.mm6932a1

[2] Baumel, A. (10 2015). Online emotional support delivered by trained volunteers: Users satisfaction and their perception of the service compared to psychotherapy. *Journal of Mental Health* (Abingdon, England), 24, 313–320. doi:10.3109/09638237.2015.1079308

[3] Lin, J. (02 2020). *My Experience as a 7 Cups Volunteer Listener.* Medium.

[4] Patterson, E. (07 2020). *7 Cups Review: What's Offered, Cost, and Who It's Right For* (B. Troy, Ed.). Retrieved from https://www.choosingtherapy.com/7-cups-review/

[5] Gale, D., & Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1), 9–15. Retrieved from http://www.jstor.org/stable/2312726

[6] Roth, A. E. (7 2007). *Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions*. doi:10.3386/w13225

[7] Adachi, H. (2003). A search model of two-sided matching under nontransferable utility. *Journal of Economic Theory*, 113(2), 182–198. doi:10.1016/S0022-0531(03)00085-1

[8] Roth, A. E. (2002). The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica*, 70(4), 1341–1378. doi:10.1111/1468-0262.00335

[9] Bhalgat, A., Chakrabarty, D., Khanna, S., Goldberg, L. A., Jansen, K., Ravi, R., & Rolim, J. D. P. (2011). Social Welfare in One-Sided Matching Markets without Money. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 87–98. Springer Berlin Heidelberg.

[10] Zhou, L. (1990). On a conjecture by gale about one-sided matching problems. *Journal of Economic Theory*, 52(1), 123–135. doi:10.1016/0022-0531(90)90070-Z

[11] Koole, G., & Mandelbaum, A. (07 2002). Queueing Models of Call Centers: An Introduction. *Annals of Operations Research*, 113, 41–59. doi:10.1023/A:1020949626017

[12] Borst, S., Mandelbaum, A., & Reiman, M. I. (2004). Dimensioning Large Call Centers. Operations Research, 52(1), 17–34. doi:10.1287/opre.1030.0081

[13] Akşin, O. Z., & Harker, P. T. (2003). Capacity sizing in the presence of a common shared resource: Dimensioning an inbound call center. European Journal of Operational Research, 147(3), 464–483. doi:10.1016/S0377-2217(02)00274-6

[14] Che, Y.-K., & Tercieux, O. (01 2021). *Optimal Queue Design*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3743663

[15] Hancock, J. T., Toma, C., & Ellison, N. (2007). The Truth about Lying in Online Dating Profiles. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 449–452. Presented at the San Jose, California, USA. doi:10.1145/1240624.1240697

[16] Hitsch, G. J., Hortaçsu, A., & Ariely, D. (3 2010). Matching and Sorting in Online Dating. *American Economic Review*, 100(1), 130–163. doi:10.1257/aer.100.1.130

[17] Roth, A. (1984). Misrepresentation and stability in the marriage problem. *Journal of Economic Theory*, 34(2), 383-387. doi:https://doi.org/10.1016/0022-0531(84)90152-2

# My way or the riot way: (Markov) Equilibrium in almost-Rubinstein Bargaining with Costly Deferral

Daniel Luo

*Northwestern University*

*Abstract*— **This paper studies "two-stage" perfect information dynamic bargaining. In the first stage, players cannot split surplus, but only agree whether or not to proceed to the second stage, where canonical Rubinstein bargaining occurs. Bargaining power is realized both through an exogenously evolving state variable and an endogenous choice of one player (the activist) to destroy some share of the other player's (the government) surplus in a costly fashion. This second-order framework formalizes the intuition offered by activists during interviews that rioting is a justified response to repeated state ignorance of their movements and demands as a way to force engagement and secure a "seat at the table." I prove existence of and characterize Markov perfect equilibrium in this generalized setting under the assumption the first stage must end in $T$ periods. Under some technical assumptions on the opportunity cost of rioting, the equilibrium is unique, and moreover, under certain boundary conditions, the *appearance* of rioting is volatile and discontinuous even as its *intensity* decays monotonically in support, reflecting the volatile nature of rioting. Finally, I characterize the welfare effects of endogenous surplus destruction and show the option to riot *strictly* improves activist welfare, even if it is almost never exercised.**

## I. Introduction

Riots have long been used as a political tool to extract concession from the state, with its motivation and consequences a subject of extensive empirical study by both political scientists and economists alike. Several empirical studies have found government concession to activist pressure follow riots from those demands; notably, the 1992 Los Angeles riots were succeeded by government reallocation of public goods to areas with high minority populations, in line with activist demands (Enos, Kaufman, and Sands, 2019). Similarly, the south Sudanese riots in the early 21st century led to increased public support for the south Sudanese demands for *secession* (even among the stringently opposed North), though this came at the cost of decreasing support for naturalization of southern population by the dominant Northern government (Beber, Roessler, and Scacco, 2014). This nonmonotonicity in support for riots mirrors changes in support for Palestinian demands following pro-Palestinian terrorist violence, which forced right-wing moderation by Israeli politicians on the issue of independence *at first*, though eventually increased the popularity of violent reprisal against Palestinians instead (Gould and Klor, 2010). Together, the literature on previous riots combined with the nonmonotone nature of the consequences of rioting suggest a fundamentally strategic component to the act on a macropolitical scale worth investigation and which cannot be captured by existing literature on bargaining.

Towards this end, I develop a theoretical model which captures the strategic impetus for rioting and which yields nuanced comparative statics capable of explaining the behavioral logic of riots from a theoretical perspective. In particular, by conditioning activist and government strategies on *public support* (which is exogenous and evolves stochastically), I am able to capture the dynamic effects of both the *appearance* of rioting, explaining its apparent volatility, and the *intensity* of those riots, which allows for multidimensional analysis to arise from a simple finite-horizon framework.

## II. Basic Model

### A. Preliminaries

The model is inspired by recent political violence in the United States, particularly that accompanying the racial justice movements in summer 2020 following the murder of George Floyd, Jr. (Kishi and Jones, 2020). Specifically, several prominent authors and activists defended the riots, claiming it was a necessary tool to force a government response to a problem it had costlessly ignored for too long (see Dastagir, 2020; Escobar and Klor, 2010; and Lossin, 2020). Eventually, after a spate of both peaceful protest and violent riots, municipalities, states, and even the federal government engaged in a variety of reforms to policing in their jurisdictions (see Subramanian and Arzy, 2021), though only after after discussions (modeled here as bargaining a la Rubinstein, 1982) with local activists.

Heuristically, first consider a "two-stage" game of perfect information with two players, a government (G) and activist (A). In the first (finite-time) stage of the game, G and A decide *whether* or not to proceed into the second stage of the game at each period t. Should they agree, they then proceed to the latter stage, where they engage in standard Rubinstein bargaining over unit surplus with discounts $(\delta_G, \delta_A)$ respectively, yielding the unique subgame perfect allocation of surplus for the government and the agent respectively as

$$R_G = \frac{1-\delta_A}{1-\delta_G\delta_A} \text{ and } R_A = \delta_A\left(\frac{1-\delta_A}{1-\delta_G\delta_A}\right)$$

I would like to thank Harry Pei for invaluable discussions, feedback, and guidance, and Ansuman Bardalai, Eric Gao, Joshua Miller, and Sebastian Sardon, and two referees for helpful comments. This work was financially supported in part by the Northwestern Office of Undergraduate Research. All errors are, of course, my own.

respectively [18], noting in this equilibrium allocation the government makes the first offer in the second stage, partially capturing the power differential between the government and the activist. However, should the government never agree to proceed to the latter stage, then at the final period of the first stage the game ends, the activist loses, and the government keeps all the surplus (less some flow costs incurred for ignoring the activist in the intervening periods).

The solution to the problem without rioting is formalized and solved in section (3.3) of this paper, but the key economics problem is as follows: in the first stage, the government considers their net expected surplus at the last period of the first stage of the game, and concedes if and only if that payoff is smaller than the one obtained in Rubinstein bargaining; since the activist as of now has no action set in the first stage, this problem is endemic only to the government, and independent of the activist's preferences.

From here, I add the ability of the activist to decrease the former payoff by rioting (though this comes at the risk of hurting themselves as well), which will allow them to force government concession in some instances where the government may otherwise prefer simply to ignore an impotent activist, formalizing the intuition offered by Martin Luther King, Jr. that rioting is the "language of the unheard." In particular, at any time in the first stage of the game, the rioter can choose to destroy surplus (both physical, e.g. storefronts, and reputational, e.g. government legitimacy) to decrease the government's payoff, but at some recoil cost to themselves (e.g. the opportunity cost of time, physical danger, the risk of being arrested, etc.). Thus, rioting may not be a dominant strategy, and the activist faces an economic tradeoff at each point in the state space.

The overall method of analysis in this paper follows a robust tradition of utilizing game-theoretic methods to formally investigate the strategic impetus of rioting (see [1], [2], [3], [6], and [8]), though I diverge from these models as they endogenize *regime change* and often have finite (in particular, binary) state spaces. Since my model does not endogenize changes in the government's regime, I am better able to model activists who riot (instead of rebel) to enact *policy change* as opposed to *regime change*.

In a similar way, I do not broadly consider a government's ability to repress nascent protest[1] While this is not particularly useful for modelling long struggles between insurgents and governments (for example, the IRA), this simplification allows for the model to focus primarily on the rioter's strategy while still accurately modelling situations where the activist's demand is "one-shot" (for example, surplus bargaining).

*B. Formalization*

In the second stage, the Rubinstein equilibrium is uniquely pinned down by the discount values $(\delta_A, \delta_G)$ and independent of other model primitives; thus, the infinite-horizon model is *equilibrium and outcome equivalent* to a finite-horizon game (only the "first stage") where the Rubinstein payoffs are exogenously fixed. Without loss, consider a finite-time game with total periods $T$ (the first stage from above). At each time, the government and activist face the same time-invariant actions, $a^i \in \mathcal{A}^i$ for $i \in \{A, G\}$. The government can choose to concede (ending the game), or wait, so $\mathcal{A}^G = \{C, W\}$. Meanwhile, the activist picks a nonnegative value $\mathcal{A}^A = [0, \infty)$, which represents the cost of rioting that they inflict on the government. If $a^A = 0$, the activist does not riot.

Actions are conditioned after observing the state, $\theta \in \Theta$, representing the level of public support for the activist's demand on the government, and $\theta$ evolves according to some Markov transition function $f(\cdot|\theta)$, which is continuous in the second argument and a probability distribution over $\Theta$ in the first. The histories of the game are natural:

**Histories**: A time $t$, a history is a sequence of previous actions by both players in each $0 \leq k \leq t-1$, a state for each of those time periods, and the state currently faced by the actors, $\theta_t$. Thus, the time-$t$ histories are given by $\mathcal{H}_t = \{\Theta \times (\mathcal{A}^A \times \mathcal{A}^G \times \Theta)^t\}$, and so the set of all histories is $\mathcal{H} = \bigcup_{t=1}^T \mathcal{H}_t$. As government concession ends the game, I restrict to considering only the *pruned game tree*: any node on the path of play cannot include $C$ at any time. Thus, the on-path histories are induced by the *pruned* time-$t$ histories $\mathcal{H}_t^p = \Theta \times (\mathcal{A}^A \times \Theta)^t$, with the pruned game tree $\mathcal{H}^p = \bigcup_{t=1}^T \mathcal{H}_t^p$. From here, set $\mathcal{H} = \mathcal{H}^p$ and $\mathcal{H}_t = \mathcal{H}_t^p$ so that off-path histories are never considered. Then the terminal nodes are

$$\mathcal{Z} = (\mathcal{H}_T \times \mathcal{A}^A \times W) \cup \left( \bigcup_{j=0}^T \mathcal{H}_j \times \mathcal{A}^A \times \{C\} \right)$$

which consists of a history at each possible time of concession, the activist's action at time $t$ (which is not in $h_t$), and time $T$ cases where the government never concedes. On this path, $\mathcal{H}$ is endowed with a partial order $\succ$ where $h_t \succ h_s$ if $h_t$ is still on the path after $h_s$ is realized; this can be read as $h_t$ being *on the feasible path given* $h_s$. As long as $f(\cdot|\theta)$ has full support for all $\theta$, all $h_t \in \mathcal{H}_s$ are on the feasible path for all $h_s \in \mathcal{H}_s$

To define payoffs, we formalize the incentives facing the government and activist outlined above. First, recall the state space $\theta$ signals popular support. Thus, at each stage $t \leq T$, the government faces a continuous and time-homogenous cost $c(\theta)$ to ignoring the will of $\theta$ of the population, where ignorance is costless if no one supports the demand ($\theta = 0$), and is very costly if everyone does ($\theta = 1$). Moreover, the activist may decide to riot, i.e. pick an (additively separable) per-period flow penalty into the governments payoff in each periods, the *rioting cost*, though this results in them incurring

[1]For some literature which endogenizes government repression under uncertainty, see Pierskalla (2016) [17], though their class of payoffs is significantly less rich than those which I consider and thus they do not consider statics or dynamics at equilibrium.

a"recoil cost" for having decided to riot, denoted by $\eta$.[2] The more the activist chooses to riot, the higher the recoil cost of rioting is for themselves. These, along with some technical assumptions, are assumed throughout the paper and enumerated below.

**Assumption 1** The following assumptions are satisfied.

1) $\Theta$ is a closed interval in $\mathbb{R}$ with left endpoint 0; wlog, $\Theta = [0, 1]$.
2) The function $c : \Theta \to \mathbb{R}_+$ is nondecreasing and continuous. It satisfies the *zero-movement* and *unanimity* boundary conditions $c(0) = 0$ and $c(1) = 1$.
3) The function $\eta : \mathcal{A}^A \to \mathbb{R}$ is strictly increasing and continuous. It satisfies the *zero-movement* boundary condition $\eta(0) = 0$.
4) The markov transition function $f : \Theta^2 \to \mathbb{R}$ satisfies the following:
   a) Stochastic Dominance in $\Theta$: If $\theta_1 \geq \theta_2$, then $f(\theta|\theta_1) \succsim f(\theta|\theta_2)$, where $(\succsim)$ is the partial order induced by first order stochastic dominance.[3]
   b) Continuity in Conditioning: if $\{\theta_n\} \subset \Theta$ with $\lim_{n\to\infty} \theta_n = \tilde{\theta}$, then $f(\theta|\theta_n) \to f(\theta|\tilde{\theta})$ for each $\theta \in \Theta$ in the first argument pointwise.

If $f(\cdot|\theta)$ satisfies the fourth assumption, and also is sufficiently volatile, that is, for all $\theta'$, $f(\theta|\theta')$ is supported in $[0, 1]$, then $f$ is regular. Volatility ensures the state-space is fully supported in each time-period, and is necessary for characterize welfare.

Fix a period $t$. Within each period, the players move *sequentially* and not simultaneously in each stage, with the government first (importantly) able to observe the level of rioting the activist commits to before acting, which will be important in understanding the formal structure of payoffs. Each stage proceeds as follows.

1) At the beginning of time $t$, $\theta_t$ is drawn from $\Theta$ according to the conditional law of motion $f(\theta_t|\theta_{t-1})$ and is made common knowledge.
2) The activist commits to a level of rioting, $a_t^A$, which is observed by the government.
3) The government observes $(\theta, a_t^A)$ and picks some element of $\mathcal{A}^G$.
4) If $t < T$, the following occurs: If the government concedes, the game ends and utility is collected at the corressponding terminal node. If the government waits, the history updates to include $(\theta_t, a_t^A)$, and the game proceeds to time $(t + 1)$ and repeats again at time (1).
5) If $t = T$, the game ends and payoffs are collected according to their corressponding terminal nodes. The game does not proceed to stage $(T + 1)$.

[2]This construction may appear odd, but is a simplification that may be taken without loss. In particular, consider instead that the activist picks an action $r$ and then decreases the government's payoff by $g(r)$, invoking some recoil cost $h(r)$. If $g(r)$ is strictly increasing in $r$, (the more the activist riots, the more it hurts the government), then it is sufficient to let the activist pick $g(r)$ and incur cost $h(g^{-1}(r))$. We choose to work in this payoff structure instead for the sake of clarity.

[3]This condition is analogous to Milgrom's monotone likelihood ratio property; see [?].

Payoffs are given as follows at each history.

**Payoffs:** Payoffs are functions $p^i : \mathcal{Z} \to \mathbb{R}$ for $i \in \{A, G\}$. There are two distinct cases. First, fix $z \in \mathcal{Z}$ such that $z = \{h_t, a_t^a, C\}$; that is, the government concedes. Then

$$p^G(z) = \delta_G^t R_G - \left( \sum_{j=0}^{t-1} \delta_G^j [c(\theta_j) + a_j^a] \right)$$

and

$$p^A(z) = \delta_A^t R_A - \sum_{j=0}^{t} \delta_A^j \eta(a_j^a)$$

Second, fix $z = \{h_T, a^a, W\}$. Then the payoffs are

$$p^G(z) = \delta_G^T - \sum_{j=0}^{T} \delta_G^j [c(\theta_j) + a_j^a]$$

and

$$p^A(z) = -\sum_{j=0}^{T} \delta_A^j \eta(a_j^a).$$

Note the structure of the payoffs is slightly unconventional: if the government concedes at time $t$, then they do not internalize the cost of ignoring the individual $c(\theta_j)$ and the cost of rioting, $a_t^j$ (though the activist still feels the costs of the riot, $\eta(a_t^j)$). This can be interpreted as saying that if the government concedes, the myriad of costs borne to rioting (reputational, political, etc.) are no longer borne by the government as it has conceded, while the costs to rioting (arrests, time, organizational costs) for the activist are borne regardless of the government's actions.

**Strategies** A strategy takes any history and assigns an action for each player. Formally, a strategy is a mapping $\sigma^i : \mathcal{H} \to \mathcal{A}^i$. Let $\mathcal{S}^i$ be the set of all strategies for player $i$. A stage-game strategy is a strategy at any time $t$; that is, a mapping $\sigma_t^i : \mathcal{H}_t \to \mathcal{A}^i$. Each strategy in $\mathcal{S}^i$ is *equivalent* to a sequence of stage game strategies, $\{\sigma_t^i\}_{t=0}^T$. The set of all stage-game strategies for player $i$ at time $t$ is denoted $\mathcal{S}_t^i$.

**Expected Utility** The notion of utility is *ex-ante* expected utility with risk neutral agents. In particular, given a pair of strategies $(\sigma^A, \sigma^G)$, expected utility can defined at any node on the path of play in the game as the expected discounted stream of payoffs from future terminal nodes. That is, the expected utility functions $u^i : \Theta \times \sigma_t^A \times \sigma_t^G \to \mathbb{R}$ take a pair of stage-game strategies and give the payoff to the respective player if that pair of strategies is played, and that state of the world realized. Concretely, utility is

$$u_t^i(\theta, \sigma^A, \sigma^G) = \int_{\mathcal{Z}} p^i(z) d\mathbb{P}_t(\theta, \sigma^A, \sigma^G)$$

where $\mathbb{P}_t(\theta, \sigma^A, \sigma^G)$ is the probability distribution over the terminal nodes induced by $\theta$ given that $\sigma^A, \sigma^G$ are played. Note that the support of $\mathbb{P}_t$ is exactly

$$\left\{ h_{t+j} \in \mathcal{Z} \;\middle|\; \begin{array}{c} \sigma_{t+j}^G(h_{t+j}) = C \text{ and } \sigma_{t+k}^G(h_{t+k}) = W \\ \text{for all } h_{t+j} \succ h_{t+k} \end{array} \right\}$$

where $k < j$ and $k, j \geq 0$.

Relevant notation introduced in the previous section is collected below in Table (1) for both clarity and ease of exposition.

TABLE I: Game Notation

| Notation | Meaning |
|---|---|
| $(R_G, R_A)$ | Rubinstein payoffs determined by discount rates $(\delta_G, \delta_A) \in (0,1)^2$. |
| $(\mathcal{A}^G, \mathcal{A}^A)$ | Feasible stage-game actions, $\{C, W\} \times [0, \infty)$. |
| $\Theta$ | The state space. A bounded subset of $\mathbb{R}$. Interpreted as activist support. |
| $f(\theta\|\theta')$ | A conditional time-homogenous Markov distribution for $\Theta$. |
| $\mathcal{H}_t$, $\mathcal{H}$ | The time $t$ histories and total histories, respectively. |
| $\mathcal{Z}$ | The terminal nodes. If given at $t < T$, must include government concession. |
| $c(\theta)$ | A *cost* function $\theta \to \mathbb{R}$ representing the cost of waiting to the government. |
| $\eta(a_t^a)$ | A *recoil* function $a_t^a \to \mathbb{R}$ representing the (economic, criminal, temporal) cost to the activist of rioting. |
| $(\sigma_t^A, \sigma_t^G)$ | Stage-game strategies, $\sigma_t^i : \mathcal{H}_t \to \mathcal{A}^i$ dictating game action. |

### C. Markov Perfection

The solution concept used throughout is Markov perfect equilibrium. First introduced by Maskin and Tirole, 2001 (see also: Fudenberg and Tirole, 1991[4]) and explicitly motivated as appearing in a wide-variety of applied theory, Markov perfection requires that the strategies are contingent only on the state, and not the history more broadly. This natural restriction has appeared in observing (broadly understood) principle-agent stochastic games (see, for example, Liu and Roth 2017), and has also found particularly potent use in analyzing conciliatory bargaining between the state and dissident groups, as utilized by Acemoglu and Robinson, 2000 and Acemoglu and Robinson, 2001. In this model, a Markov perfect equilibrium is a pair of strategies s.t. both players are best responding to (and only to) every state at every time. This is formalized below.

**Definition 2** A <u>markov perfect equilibrium</u> is a tuple $(\sigma^G, \sigma^A) \in \mathcal{S}^G \times \mathcal{S}^A$ such that for all $0 \leq t \leq T$ and for each $\theta \in \Theta$, the associated stage-game strategies satisfy

$$\sigma_t^i = \underset{\{\sigma_t \in \mathcal{S}_t^i\}}{\operatorname{argmax}}\, u_t^i(\theta, \sigma_t, \sigma_t^j) \text{ for } i, j \in \{A, G\}, i \neq j$$

that is, for each state $\theta$, both individuals are best responding.

The requirement that equilibrium be Markov perfect is natural given the fact that the payoffs are constructed in such a way that they are *not* contingent at each time periods on the history, but instead only the current state. Inspecting the utility functions and noting the independence of the probability distributions to the histories (i.e. $f(\cdot|\theta)$ is Markov) makes it clear *future* utility is independent of past realizations of states before time $t$. Thus, the restriction to Markov perfection does not substantively alter the analysis. Formally, this is given by the following proposition.

**Proposition 3** Every subgame perfect equilibrium is Markov perfect.

## III. Equilibrium Characterization

### A. Existence

Consider first the special case when $T = 1$; that is, the game consists only of two periods. In this case, an equilibrium can be constructed of the following form:

$$\sigma_t^G(\theta) = \begin{cases} C \text{ if } \theta \in B_t \\ W \text{ if } \theta \in B_t^C \end{cases} \quad \text{and } \sigma_t^A(\theta) = \begin{cases} r^t(\theta) \text{ if } \theta \in B_t \\ 0 \text{ if } \theta \in B_t^c \end{cases}$$

To see why this is the case, consider backwards induction. At the last stage, the government can observe the activist's rioting, and thus given this information simply makes a decision based on whether or not their payoff is greater in playing the Rubinstein equilibrium and obviating the costs of ignorance and rioting or in incurring the costs but not foresaking surplus. The activist, aware of this, commits to rioting if and only if rioting forces government concession *and* the recoil cost is less than the surplus obtained from bargaining. The set of states satisfying both the these conditions is $B_1$, and the activist picks rioting $r^1(\theta)$, exactly to make the government indifferent between conceding and waiting (as if they play any other value they can strictly increase their payoff since $\eta$ is increasing), inducing $(B_1, r^1)$.

In the former stage, the activist and government understand the equilibrium in the latter stage. Thus, given some realization $\theta_0$, the government and activist both have access to $f(\theta|\theta_0)$, consider the expected utility of proceeding to the next stage, and repeat the logic as in the last stage but with the expected payoffs instead of the Rubinstein ones. This again induces the appropriate $(B_0, r^1)$.

There is, of course, nothing particularly special about the case $T = 1$, and the characterization applies to an arbitrarily long finite repeated game with time-inhomogenous stage game. The following proposition summarizes this argument. A proof can be found in Appendix (1.1).

**Proposition 4** Fix $T \in \mathbb{N}$. In the $T$-period version of the game, there exists a set of functions $\{r^t : \Theta \to \mathbb{R}\}_{t=0}^T$ and subsets of $\Theta$, $\{B_t\}_{t=0}^T$ such that the stage game strategies

$$\sigma_t^G(\theta) = \begin{cases} C \text{ if } \theta \in B_t \\ W \text{ if } \theta \in B_t^c \end{cases} \quad \text{and } \sigma_t^A(\theta) = \begin{cases} r^t(\theta) \text{ if } \theta \in B_t \\ 0 \text{ if } \theta \in B_t^c \end{cases}$$

induce a strategy profile $(\sigma^G, \sigma^A)$ that composes a Markov perfect equilibrium.

It is necessary in formalizing this equilibrium that we make a *choice* of equilibrium: in particular, it is assumed that at indifference for the government, they choose to concede. It is easy to see that if the government does not concede, then *no* equilibrium exists, as for every possible level of rioting played that induces concession, the activist can (locally) decrease rioting while still inducing concession, and thus

[4]We rely heavily on existence, various characterizations of theorems, and explicit algorithms for computing Markov perfect equilibrium isolated in this book.

no best reply for the activist exists, and so this choice at indifference is necessary, and in some sense, "unique."

Moreover. inspecting the above proof implies that $(B_t, r^t)$ are themselves insufficient in completely characterizing equilibrium. In particular, it is necessary to define stage game strategies $\sigma_t$ from which payoff can be inductively constructed, as well as a formal notion of the government's expected utility from waiting at time $t$ , $\omega_t : \Theta \to \mathbb{R}$ and the activist's continuation value $\lambda_t : \Theta \to \mathbb{R}$. Together, the tuple $\{\omega_t, \lambda_t, r^t, B_t\}_{t=0}^T$ are sufficient to completely pin down equilibrium behavior, and are explicitly constructed in the appendix. We define some ancillary notions below which are useful in defining the "big picture" tuples characterizing equilibrium that were given above.

**Definition 5** Fix the equilibrium constructed in proposition (4) and call utility *costless* if $c(\theta_t) + r^t(\theta_t) = 0$. Then:

1) Let the function $\psi_t(\theta)$ be the costless *ex-ante* time $t+1$ utility of waiting:

$$\psi(\theta_t) =$$
$$\delta_G \left( \int_\Theta R_G[\mathbf{1}_{B_t} + \omega_{t+1}(\theta_{t+1})\mathbf{1}_{B^c_{t+1}}] f(\theta_{t+1}|\theta_t) d\theta_{t+1} \right)$$

such that the expected utility from waiting is

$$\omega_t(\theta) = \psi_t(\theta) - r^t(\theta) - c(\theta_t).$$

2) Let $\xi_t(\theta)$ be the costless *ex post* time $t$ equilibrium utility at each individual $\theta$ for the government, given by

$$\xi_t(\theta) = R_G\mathbf{1}_{B_t} + \omega_t(\theta)\mathbf{1}_{B^c_t}.$$

3) Let $\varphi_t(\theta)$ be the equilibrium *ex post* utility at each individual $\theta$ for the activist, given by

$$\varphi_t(\theta) = [R_A - \eta(r^t(\theta))]\mathbf{1}_{B_t} + \lambda_t(\theta)\mathbf{1}_{B^c_t}.$$

Each of these objects are defined for every $t$, $0 \le t < T$, with time $T$ terminal objects $\psi_T(\theta) = 1$ and $\lambda_T(\theta) = 0$. Obviously, at any point, the logic induced by the two-period game implies that

$$\omega_t(\theta) = \psi_t(\theta) - c(\theta) - \sigma^A_t(\theta)$$

that is, the expected utility from waiting is the utility in expectation if rioting and ignorance costs in this period is zero, less the actual cost of ignorance and the cost of rioting (endogenously) chosen by the activist. Similarly,

$$\lambda_t(\theta) = \delta_A \int_\Theta \varphi_{t+1}(\theta') f(\theta'|\theta) d\theta'$$

where the activist's continuation value is the (discounted) expected value of future surplus in the next period. Accordingly, $r^t(\theta)$, the (nonnegative) value of rioting which makes the government indifferent between rioting and concession, is $\max\{\psi_t(\theta) - c(\theta) - R_G, 0\}$. These definitions are collected below in Table (2) for ease of exposition, and will be referred to liberally in the proceeding analysis.

TABLE II: Equilibrium Notation

| Notation | Value |
|---|---|
| $\xi_t(\theta_t)$ | $R_G\mathbf{1}_{B_t} + \omega_t(\theta)\mathbf{1}_{B^c_t}$ |
| $\psi_t(\theta)$ | $\delta_G \int_\Theta \xi_{t+1}(\theta') f(\theta'|\theta) d\theta'$ |
| $r^t(\theta)$ | $\max\{\psi_t(\theta) - c(\theta) - R_G, 0\}$ |
| $\omega_t(\theta)$ | $\psi_t(\theta) - c(\theta) - \sigma^A_t(\theta)$ |
| $\varphi_t(\theta)$ | $[R_A - \eta(r^t(\theta))]\mathbf{1}_{B_t} + \lambda_t(\theta)\mathbf{1}_{B^c_t}$ |
| $\lambda_t(\theta)$ | $\delta_A \int_\Theta \varphi_{t+1}(\theta') f(\theta'|\theta) d\theta'$ |
| $B_t$ | $\{\theta \in \Theta : \eta(r^t(\theta)) \le R_A - \lambda_t(\theta)\}$ |

### *B. Properties*

For the entirety of this section, fix a $T$-period equilibrium as constructed in proposition (4) and enumerated in Table (2). Our characterization of equilibrium proceeds in three stages. First, the smoothness of the equilibrium objects is established. From here, changes in optimal rioting as $\theta$ and $t$ vary are both considered, though with additional restrictions on behavior (an alternative interpretation of these restrictions in the language of differentials is discussed in Appendix (II)). Proofs for the propositions here can be found in Appendix (I.2). We first have the following.

**Proposition 6** The equilibrium objects $\{r^t, \lambda_t\}_{t=0}^T$ are continuous in $\Theta$. $\{\omega_t, \sigma^t_A\}$ are continuous on $(B_t, B^c_t)$ respectively.

The intuition for this is as follows. Note that since equilibrium objects are forward looking, they in general are additively composed of parts of form

$$e(\theta) = \delta_i \left( \int_\Theta b(\theta') f(\theta'|\theta) d\theta' \right)$$

where $b(\theta')$ is some bounded function over $\Theta$. Thus, by the dominated convergence theorem, continuity of the left hand side is exactly equivalent to continuity in conditioning of the Markov transition function (Assumption (1.4.b)); since the probability mass that the forward- equilibrium objects varies smoothly, so will the expected future values, and these uniquely pin down equilibrium. In fact, the equilibrium objects are $\mathcal{C}^n$ if and only if $f(\theta'|\theta) \in \mathcal{C}^n$ in the second argument; this is implied by a slight modification of the proof of Proposition (6) and Lemma (6.1), both given in the appendix. However, we do not need differentability properties in the succeeding analysis, so we need to assume differentiability of the Markov transition function. However, the assumption that $f(\cdot|\theta)$ is continuous is worthy of discussion. In particular, note that this assumption says that as $\theta$ is locally perturbed by a small amount, then the probability density function for $\theta_{t+1}$ varies by a small amount pointwise as well. Sequentially, if $\{\theta_n\} \to \theta_0$, then continuity requires that $f(\theta|\theta_n) \to f(\theta|\theta_0)$, which simply says that probability of achieving $\theta$ giving realization $\theta_0$ can be well-approximated by looking at the probability of achieving $\theta$ at values of support $\theta_n$ close to $\theta_0$, which is quite intuitive.

Finally, note that it is not guaranteed that the objects $\{\omega_t, \sigma^t_A\}$ are continuous over all of $\Theta$. This discontinuity implies a certain degree of volatility over the rioter's strategy in the long run; even if optimal rioting itself varies continuously, it is not necessary that the strategies where rioting is *adopted* result in *observed* behavior that is smooth. This is due to the fact that $B_t$ is almost entirely characterized by the activist's decisions and exogenous objects, allowing for it to take on a particularly pathological properties. In particular, the *opportunity cost* to forcing a concession is the foregone ex-ante surplus of waiting, $\lambda_t(\theta)$ summed with the self-imposed recoil cost from rioting, $\eta(r^t(\theta))$. When this is nonunique, the set of points $\{\theta : \lambda_t(\theta)+\eta_t(\theta) = R_A\}$, which characterize indifference, can take on many shapes (e.g. there are multiple points where the decisionmaker's problem is identical). Our key assumption in pinning down the shape of equilibrium will be that this opportunity cost is unique; in particular,

**Assumption 7** For all $0 < t < T$, the function $\lambda_t(\theta) + \eta(r^t(\theta))$ is injective.

Note that this assumption implies that $\lambda_t + \eta \circ r^t$ *crosses* the value $R_A$ exactly once, and thus this assumption packs significant bite by partitioning the state space neatly into connected intervals where states induce concession and where states do not (since $R_A$ is the surplus gained from the activist by forcing concession). Thus, this assumption *is* with loss, and not particularly intuitive; however, Appendix (II) gives an alternative interpretation in the language of differentials and may be interpreted more intuitively, though showing it implies the necessary argument is somewhat more technical. For now, though, Assumption (7) is sufficient to force regularity; as a consequence of single-crossing, the intervals for concession are uniquely determined. This is shown in the following propositions. First, a topological interpretation:

**Proposition 8** Fix an equilibrium. Then for all $t \leq T$, $B_t$ is compact while $B_t^c$ is open. Moreover, the sets $B_t$ and $B_t^c$ are connected intervals.

The behavior of proposition (8) is particularly regular and highlights further the role Assumption (7) in eliminating pathological behavior. In particular, connectedness of $B_t$ is what one would expect from equilibrium: restated over $\mathbb{R}$, it implies that if rioting occurs at any two distinct levels of support, then it occurs for each intermediate value: in particular, if I find it optimal to riot at $\theta_1$ and $\theta_2$, then over any convex combination of $\theta_1$ and $\theta_2$, I should find it optimal to riot as well. There are two economic interpretations of this property. First is a monotone characterization: if the government concedes at some level of support $\theta$, then it is intuitive that they would feel *even more* pressure when support is higher, and thus concede (in particular, government action in response to public support is monotone; the more popular a policy is, they more likely they are to enact it). The second characterization favors the activists: as the only connected subsets of $\mathbb{R}$ are intervals, any Markov perfect equilibrium features a decision rule with a *cutoff point* $\theta^*$ above which, the activist feels emboldened enough to riot (if necessary) and below which, support is never high enough that rioting is feasible. Given the potential computational complexity of strategies perhaps not featuring a cutoff point (for example, it may be possible in full generality to construct an equilibrium where the activist riots only on the fat Cantor set), it is behaviorally reasonable to impose the structure implied by proposition (8) which follows from Assumption (7).

The interval structure of proposition (8) has one further consequence: in particular, that there are exactly two stage-game equilibria: when $0 \in B_t$, and when $0 \in B_t^c$, with $\theta^*$ the single point lying in $\overline{B_t} \cap \overline{B_t^c}$ and is the inflection point in the *cutoff strategy* enumerated above. However, of these two possible equilibria, only one is ever realized nontrivially.

**Lemma 9** At equilibrium, if $B_t^c$ is nonempty, then $0 \in B_t^c$.
*Proof:* Note that $\psi_t(\theta)$ is bounded by 1, as that is the total possible surplus to divide, and all flow costs are nonnegative. Now assume not. If $B_t^c$ is nonempty, by proposition (8), it is open, and so it is an interval and in particular contains 1 as $B_t^c$ and $B_t$ form an interval partition of $[0, 1]$. But if the government waits at 1, they collect utility $\psi_t(1) - c(1) = \psi_t(\theta) - 1 \leq 0$, contradicting the assumption that this is a best response, as concession gives $R_G > 0$.

Lemma (9) pins down equilibrium so that the government's decision is monotone, which spills over into equilibrium objects themselves being monotone, as shown below.

**Proposition 10** Fix an equilibrium. Then for all $0 \leq t \leq T$, $r^t(\theta)$ and $\omega_t(\theta)$ are decreasing on $\Theta$, while ex-ante activist surplus $\lambda_t(\theta)$ is increasing over $\Theta$.

These results are intuitive: $\omega_t$ is declining in support since a higher level of support *today* implies less surplus is realized tomorrow. Likewise, ex-ante expected surplus from waiting for the activist is increasing in $\theta$; both of these follow (just as the formal argument does) from stochastic dominance in conditioning of the Markov transition function, $f(\cdot|\theta)$, as this term is the only term where realizations of the state today affect expected realizations of states affecting *future* utility tomorrow.

The vital nature of the stochastic dominance assumption in delivering this result merits it a little more discussion. Intuitively, stochastic dominance says that support is "sticky:" that is, if $\theta_1 > \theta_2$, then *in expectation*, the level of support *tomorrow* will be higher; in particular, $\int f(\theta|\theta_1) - f(\theta|\theta_2) d\theta > 0$, which implies that higher levels of support today generally engender higher expectations of support tomorrow. This is reasonable in two contexts. First, in political contexts, allegiances are likely to evolve very slowly due to cognitive dissonance, and second, activist movements are generally *more likely* to gain attention (and thus future support) if they are already larger, both because of network effects and information transmission effects.

Finally, $r^t(\theta)$ decreases as well immediately from the decreasing nature of $\omega_t$ (in particular, the decreasing nature of $\psi_t$), and thus $\eta(r^t(\theta))$ is decreasing as well, implying that when rioting occurs, movements with larger levels of support are less likely to riot. However, this does imply, since $\lambda_t$ is increasing, that the function $\lambda_t(\theta)+\eta(r^t(\theta))$ does not have any easily conceivable monotonicity properties, and thus absent Assumption (7), the behavior of $B_t$ could be quite pathological.

Luckily, with Assumption (7), equilibrium is uniquely characterized, and as a result $\sigma_t^A(\theta)$ displays an intuitive intermediate nonmonotonicity property. In particular, let Assumptions (1) and (7) hold and assume that fixed $t \in T$, $B_t^c$ is nonempty, and consider its stage-game equilibrium; since $c(1) = 1$, then there exists by continuity some $\hat{\theta} \leq 1$ s.t. over $[\hat{\theta}, 1]$, $\sigma_t^A(\theta) = 0$ but concession occurs, since the cost of ignorance is so high that concession is immediate. However, sine $0 \in B_t^c$, then there exists some $\theta^*$ s.t. $[0, \theta^*)$, rioting does not happen *and* concession does not occur, since the activist's policy support, and consequently their bargaining power, are insufficient to incite a sufficiently large riot to provoke concession (as any riot that woiuld incentivize government concession would be so large that the recoil effects would far outweigh the gain in surplus). Finally, though, in $[\theta^*, \theta)$, rioting occurs at nonnegative value, thus *strictly* increasing activist surplus.[5]. Together, these characterize the intermediately nonmonotinicity property characterized above, in which $\sigma_t^A(\theta)$ takes on a discontinuous "inverted U" shape: when support is very low, there is *no rioting*, while once a threshold is reached, rioting suddenly jumps to a relatively high nonzero value $r^t(\theta^*)$, and afterwards decreases monotonically until it vanishes. This characterization captures the important intratemporal aspects of rioting, and is illustrated below by Figure (1). Similarly, Figure (2) plots government surplus.

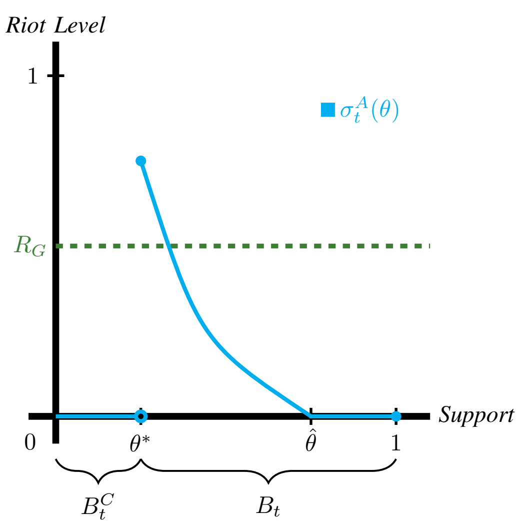



Fig. 1: Stage-$t$ Equilibrium Rioting and Ex-Ante Government Surplus

With the stage-game strategies completely pinned down, the next question is how equilibria *change* with respect to time. Generally, the finite nature of $T$ allows for relatively

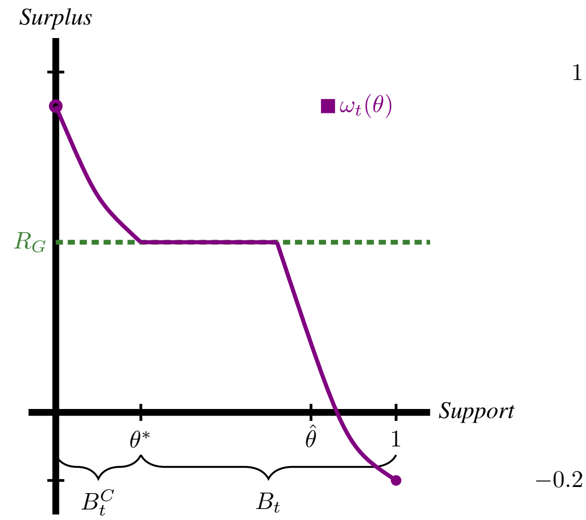



Fig. 2: Stage-$t$ Equilibrium Rioting and Ex-Ante Government Surplus

pathological behavior, though the additional parameter allows an additional measure of *exogenous* bargaining power absent in the rest of the paper. In particular, as the length of the game increases, then the (discounted) payoff from not conceding is worth *less* from the perspective of period (1). Formally, if $B_0(T)$ is the equilibrium set $B_0$ when the equilibrium is $T$-periods long, then asymptotically, government concession must occur *immediately*:

**Proposition 11** Let $\theta \in B_t^c$. Then $R_G < \omega_t(\theta) \leq \delta_G^{T-t}$. Thus, $\lim\limits_{T\to\infty} B_0(T) = \Theta$.

Proposition (11) is as far as one can go with a dynamic characterization absent additional assumptions. In particular, while the convergence to 0 at $\infty$ is guaranteed, the process of convergence need not be *monotone*, as, in a similar pathology to the problem underlying the need for Assumption (7), as time increases, $r^t(\theta)$ intuitively may increase, as there is more *government* surplus that the activist needs to offset to force indifference to concession, but the activist themselves face a deadline approaching at time $T$ where their surplus vanishes, implying that $\lambda_t(\theta)$ is decreasing in time, implying that $\{B_t\}$ may not be nested. However, given the backwards-inductive nature of equilibria, when $T > 2$, this problem occurs in the converse direction; as the $\{B_t\}$ are not nested, the statics for $\lambda_t$ and $\omega_t$ are not monotone in time, and data is lost in finite time, making the asymptotic argument the sharpest possible characterization.

To obviate this problem, it is necessary to fix an assumption; I choose to require that the sets best nested. In particular, this argument is a little more intuitive: if at some level of support $\theta$ during time $t$, the government would concede, then they should also concede at time $s < t$ at $\theta$, as they (weakly) may only have more to lose. Formally, this is Assumption (12):

**Assumption 12** Fix an equilibrium and let $s < t$. If $\theta \in B_t$,

[5]To formally show this, it is necessary to solve the government's problem when the activist cannot act. This is done in (3.3).

then $\theta \in B_s$.

This restriction to finite behavior is sufficient to completely characterize the equilibrium objects in time. In particular, the rough intuition given above can be formalized.

**Proposition 13** Fix an equilibrium satisfying Assumption (12). Then for all $t < T$, $\omega_t(\theta) \leq \omega_{t+1}(\theta)$ and $r^{t+1}(\theta) \geq r^t(\theta)$ for fixed $\theta$. For the activist, $\lambda_{t+1}(\theta) \leq \lambda_t(\theta)$.

Two remarks are in order regarding this proposition. First, as alluded to above, the backwards-inductive nature of the proof highlights the necessity of Assumption (12) in order to resolve both the nonmonotone nature of $\omega_{t+1}(\theta)$ and $\sigma_t^A(\theta)$ at each time and the pathology of $B_t$. Second, this is clearly an intertemporal version of Assumption (7), where injectivity was applied to infer that when $\lambda_t(\theta)+\eta(r^t(\theta)) \leq R_A$, in equilibrium, a *rightwards* shift in $\theta$ maintained the inequality. Analogously, Assumption (12) says that for fixed $\theta$ as opposed to $t$, a *downwards* shift in time maintains the inequality, giving the nested nature of the concession sets $B_t$ and thus the pointwise monotone nature of the equilibrium objects and thus equilibrium strategies. Note, finally, that an easy consequence of Assumption (12) and proposition (13) is that $\sigma_t^A(\theta) \leq \sigma_{t+1}^A(\theta)$, so the *strategy profile* of the activist themselves increases in time.

### C. Welfare

I conclude by sharpening the zero-sum intuition briefly mentioned above and contrasting the government's equilibrium strategy with (their) first best strategy. Clearly, welfare decreases when rioting is allowed, and moreover with $B_t^c$ is nonempty for all $t$, this decrease in realized welfare is strict. To do this, it is necessary to formalize the government's decision problem, which is analogous to the multi-agent environment above (and so the discussion is somewhat more terse).

The action space for the government is, as it was in Section (3.2), still $\{C, W\}$. A history at time $t$ is a sequence $h_t \in \{\Theta \times (\mathcal{A}^G \times \Theta)^t\} = \mathcal{H}_t$. The set of all histories is $\mathcal{H} = \bigcup_{j=0}^T \mathcal{H}_j$, as it was in the two-player game. the set of all *pruned* histories are those which do not include a terminal note (that is, $h_t = \{\theta_t\}_{t=0}^T \times \{W\}_{t=0}^T$, or concession has not occurred), and is denoted $\mathcal{H}^p$. Equivalently, $\mathcal{H}^p = \bigcup_{j=0}^T \mathcal{H}_t^p$, where $\mathcal{H}_t^p = \{\Theta \times (\{W\} \times \Theta)^T\}$. Note these are all of the strategies which are relevant in the decision-making process, and thus *strategies* may be defined: here, they are functions $s : \mathcal{H}^p \to \mathcal{A}^G$, taking each history on the path of play to an action. A stage-game strategy is a function $s_t : \mathcal{H}_t^p \to \mathcal{A}^G$. The set of all strategies is given by $\mathcal{S}$. All terminal nodes are either of form $\mathcal{H}_t^p \times \{C\}$ or $\mathcal{H}_T^p \times \{W\}$ at the last node, and so each one can be written as $(h_t, C)$ or $(h_T, W)$ for some history $h$. The set of all terminal nodes is $\mathcal{Z}$. The index $t$ is the *time of concession*. Payoffs are defined at each terminal node, and is thus a function $p : \mathcal{Z} \to \mathbb{R}$; there are two cases, for $z = (h_t, C)$ and $z = (h_T, W)$. Respectively, these payoffs are

$$p(h_t, C) = R_G - \sum_{j=0}^{t-1} \delta_G^j c(\theta_j) \text{ or } p(h_T, W) = 1 - \sum_{j=0}^{T} \delta_G^j c(\theta_j)$$

Utility is defined in the same way as in the game, with $u_t^i : \mathcal{H}_t^p \times \mathcal{S} \to \mathbb{R}$ given by

$$u_t^i(h_t^p, s) = \int_{\mathcal{Z}} p(z) d\mathbb{P}(h_t^p, s)$$

where $\mathbb{P}(h_t^p, s)$ is the probability measure over all pruned terminal nodes succeeding $h_t^p$ induced by the strategy $s$ and the realization at $t$ of $h_t^p$.

**Definition 14** A strategy is stochastically ex-ante optimal if it satisfies

1) State-Contingency: For all $h_{t-1}^p, \tilde{h}_{t-1}^p$ s.t. $s(h_{t-1}^p) = s(\tilde{h}_{t-1}^p) = W$, then $s(h_{t-1}^p, W, \theta) = s(\tilde{h}_{t-1}^p, W, \theta)$; that is, strategies are history-independent.
2) Optimality: For all $t$ and $h_t^p \in \mathcal{H}_t^p$, the strategy $s$ satisfies
$$s_t = \underset{\{s_t \in \mathcal{S}_t\}}{\operatorname{argmax}}\, u_t^i(h_t^p, s)$$

By state-contingency, it is sufficient to write utility as a function of $\theta$, noting that, as in proposition (3), this is general because of utility is structured. Existence is established next.

**Proposition 15** A stochastic ex-ante optimal strategy exists.

Note that proposition (15) does not show this is the unique ex-ante optimal strategy; in particular, mixed strategies have not been considered, and at one point a *choice* for equilibria was made, at the point of indifference between the two choices the actor always chooses to concede. This is to mirror the case in the game itself, where concession is chosen at indifference (in part to avoid the computationally and intuitively taxing problem of the activist picking utility at $R_G - \varepsilon$ for all $\varepsilon > 0$ to get concession but not getting concession at $R_G$). However, off this point of indifference, this induced equilibrium strategy is unique (optimality requires strictly one choice) and thus, when contrasted with the constructed equilibrium in proposition (4), is useful to compare the welfare implications of rioting.

There are two useful equilibrium objects constructed in the proof of proposition (15) that are worth reiterating explicitly. First is the set of concession, $C_t$ which is analogous to $B_t$ in the two-player game. In particular, the optimal strategy is concession on $C_t$ and nonconcession at $C_t^c$ for all $t$. Second is the "ex-ante" utility from waiting, $\kappa_t$, which is defined abstractly in the proof but more concretely can be thought as a (backwards) inductively defined object

$$\kappa_t(\theta^*) = \delta_G \int_{\Theta} [R_G \mathbf{1}_{C_{t+1}} + (\kappa_{t+1} - c(\theta)) 1_{C_{t+1}^c}] f(\theta|\theta^*) d\theta$$

where $\kappa_t$ is thus analogous to $\psi_t$ in the equilibrium game. These two objects together completely pin down the government's decision problem above, and so relating these objects to the objects in a Markov perfect equilibrium yields interesting comparisons on the injection of riots into the problem. The below proposition contrasts the markov-perfect equilibrium in proposition (4) with the ex-ante optimal strategy from proposition (15).

**Proposition 16** Fix equilibrium objects $\{c, \eta, f(\cdot|\theta)\}$ and consider the equilibria objects in Proposition (4) and the optimal strategy from Proposition (15). Then if $\{B_t, \omega_t, \lambda_t, r^t\}$ and $\{C_t, \kappa_t\}$ are the respective relevant induced objects,

1) For all $t \le T$, $\psi_t(\theta) \le \kappa_t(\theta)$ and $C_t \subset B_t$.
2) For all $t < T$ such that $B_t^c \ne \varnothing$, $\{\theta : \sigma_t^A(\theta) \ne 0\}$ is nonempty and $\psi_t < \kappa_t$.

Two remarks about this proposition are in order. Finally, the first part of (2) is not in itself a welfare comparison, bur rather simply an existence theorem: when the government does not immediately concede, the model is "nontrivial" in the sense that it predicts some rioting at levels of support, which gives the model some nontrivial predictive bite. Second, and more importantly, the welfare comparisons. First, (1) is a general statement; weakly, that the option to riot makes the government worse off in expectation, as $\psi_t$ and $\kappa_t$, respectively, track the expected utility of the government should they not concede at time $t$. (2) *sharpens* this: so long as the rioting problem is nontrivial, (e.g. there is a point of nonconcession), then the decrease of welfare experienced by the government at *every* time (less the last one) is *strict*: that is, the option to riot always makes the government worse off. This has a few political implications. First, it gives a general existence statement for the presence of rioting, regardless of how steep the cost of doing so is. In particular, so long as the government has enough state capacity to resist the activist's demand at some arbitrarily low level of support, (say, 0), then there will always exist a set of *positive* (Lebesgue) measure on which rioting is optimal for the activist, and the government concedes. This follows intuitively from two facts: first, that the activist can always gain more surplus from forcing a concession, and second, that the government *must* concede at one point. Thus, there always exists an interval where the government is "close enough" to concession that rioting forces a reticent government to concede: exactly the intuition offered by activists for the prevalence of rioting.

## IV. Final Remarks

### A. Historical Discussion

The nuanced statics that fall out of equilibrium in the model can help rationalize a wide variety of phenomena associated in riots. First and foremost, the discontinuity at $\theta^*$ implies that the *appearance* of rioting can be volatile, as a movement may appear peaceful one day but then suddenly erupt into violent protest even with just a small perturbation in policy support; this discontinuity lies at the core of the model. This sensitivity to small changes (though the point of sensitivity is unique) mirrors a variety of phenomena in riots, which are often described as unpredictable in *appearance*. However, once this first discontinuity threshold is crossed and existence is established, rioting *decays* continuously in support, which rationalizes why larger movements which hope to eventually parlay with the government self-police the viciousness of their destructive behavior and instead call for "peaceful protest" as opposed to the more radical, violent ideologies espoused by smaller political groups. Such phenomena is highlighted by two recent riots.

The riots which occurred on January 6, 2021 in Washington, D.C. encapsulate the phenomena isolated in this model; first, general support for their political demand was not sufficiently high to warrant a government concession by itself, and because their bargaining position was very weak, they had very little to lose and much to gain from rioting, which, at equilibrium in the model, would not only imply existence of rioting even with only low levels of support, but also particularly volatile rioting should it occur. The violence of those riots may also have been at least partially temporal; as predicted by the model, rioting becomes gradually more volatile as the finite-horizon deadline approaches, since activists have less to lose but more to gain, which may explain the timing of the attempted insurrection at January 6, the final day at which the rioter's (loosely articulated) demands could be met.

The racial justice protests which occurred in the summer of 2020 also provide an instructive example in a few ways. First, the model predicts that rioting is transient, in that it occurs if and only if concession can be credibly ensured (see Subramanian and Arzy, 2021 for examples of concessions in policing and criminal justice in 2019) and thus at most occurs in one period. This helps hint at why, perhaps, riots are often not sustained, though it also highlights the strength of the perfect-information rationality assumptions at the start of this model: rational activists with political demands sustain the cost of rioting if and only if it will help them, and as a result will not "riot earlier," as the model abstracts away from reputational questions and imperfect information.[6] However, this prediction is at least somewhat empirically valid; often, riots over the same *specific* political demands do not occur again (for example, while riots over racial justice have generally occurred in different times, the impetus and demands often change, as was the case in the 1960s civil rights movement, 1992 Los Angeles riots, and 2020 police reform protests). In addition, the *timing* of the racial equity riots is instructive, as they largely died down by July of 2020 at the latest, far before any police reform or racial justice legislation was passed, and almost a year before the trial of Derek Chauvin. This indicates that the political impetus behind the riots was not to extract material surplus from the government, but rather to force a government *response* and indication that it was open to negotiation. In particular, many of the protests died down after state and local governments signalled openness to reforming their policing systems (and prior to them actually enacting them), which is consistent with the intuition that rioting is simply a mechanism to force good-faith bargaining with a government.

Finally, the sensitivity of the existence of rioting, and thus,

[6]Note that miscalibrated beliefs, or imperfect information about the state, $\theta$, may also lead to *lack* of state concession despite rioting, and so miscalibration on the part of January 6 rioters may explain the difference in results between the two examples discussed here.

government concession, to the recoil cost of rioting to the government highlights that a government where rioting is more costly for the activist is likely to have to concede less, which strictly decreases the responsiveness of governments to semi-popular political demand while strictly increasing state utility. Thus, states with higher capacity may increase penalties to rioting to arbitrarily high quantities, which explains the laws against public protest among authoritarian governments. Conversely, explicit protections for protest conversely may *decrease* the cost of rioting, which serves as another mechanism for which democratization may force a government to commit to sharing surplus with poorer individuals who have the option to riot, which offers a non-redistributive mechanism of the payoffs in a state's decision to democratize, as assumed by Acemoglu and Robinson, 2000. However, government repression is note endogenized in this model, and is an extension that may be considered at a future time.[7]

### B. Conclusion

My model formalizes the notion that movements are the most volatile when support is neither particularly high nor low, and offers some parametric predictions for when political demands are likely to result in violent riots. The core intuition embedded in the model is that rioting can force an otherwise unresponsive government to listen to activist demands and engage in concession by increasing the cost of that ignorance, and thus allowing for good-faith surplus bargaining to occur.

This model, of course, has limitations, and several questions induced from this paper may be subjects of future research. In particular, the role of riots in building a reputational bargaining postures for the activist, and its endogenous effect on support, both cannot be addressed by this model and can help relate this analytically simple framework to more sophisticated methods in reputation, which may allow for more model complexity and differentiated insights. Other areas of research involve relaxing the exogenous nature of the recoil function, and endogenizing state capacity to study the differential in responses to rioting between authoritarian and democratic governments. Moreover, the Rubinstein split of surplus in the second stage is exogenous; endogenizing the surplus at the terminal node as a function of the level of support (i.e. extending the interpretation of support as bargaining strength to the second stage) and studying these effects at equilibrium is also a potentially fruitful direction of query and not something my paper addresses.[8]

Despite its analytical simplifications, this theory offers a potentially useful paradigm in understanding the rational impetus behind rioting, and can deliver several rationalizations both for its apparent volatility and its persistence as a modern tool of political bargaining. By departing from the standard bargaining literature by introducing a second-order of bargaining from which surplus payoffs are invariant to time and posture, I am able to obtain a *discontinuous* inverted-U intratemporal equilibrium which delivers nuanced statics and has several implications for government-activist interaction that had previously not been explored, and demonstrating that Dr. King's famous words that riots were the "language of the unheard" are, perhaps surprisingly, consistent within a microfounded game-theoretic economic framework.

[7]However, this issue has already been tackled by the political science literature repeatedly; Davenport, 2007 contains a particularly thorough review of the literature; further extensions of this model may help import significant mathematical and game-theoretic generality into the literature, though, something existing, more simplistic models lack.

[8]Note, however, that the Rubinstein position merely microfounds the exogenous second-stage-payoffs; the analysis in this paper is invariant under any exogenous tuple of payoffs $(\alpha, 1-\alpha)$ for any $\alpha \in (0,1)$.

## REFERENCES

[1] Acemoglu, D. and Robinson, J.A. (2001). A Theory of Political Transitions. *American Economics Review* 91.4, pp. 938–963. doi: 10.1257/aer.91.4.938.

[2] Acemoglu, D. and Robinson, J.A. (2000). Why Did the West Extend the Franchise? *The Quaterly Journal of Economics* 115.4, pp. 1167–1199. doi: 10.1162/003355300333042.

[3] Little, A. et. al. (2015). Elections, Protest, and Alternation of Power. *The Journal of Politics* 77.4, pp. 1142–1156. doi: 10.1086.682569.

[4] Beber, B. et. al. (2014). Intergroup Violence and Political Attitudes: Evidence from a Dividing Sudan. *The Journal of Politics* 76.3, pp. 649–665. doi: https://doi.org/10.1017/S0022381614000103.

[5] Enos, R. et. al. (2019). Can Violent Protest Change Local Policy Support? Evidence from the Aftermath of the 1992 Los Angeles Riots. *American Political Science Review* 113.4, pp. 1012–1028. doi: 10.1017/S0003055419000340.

[6] Ash, K. (2011). A game-theoretic model for protests in the context of postcommunism. *Communist and Post-Communist Studies 44.1*, pp. 1–15. doi: 10.1016/j.postcomstud.2011.01.007.

[7] Dastagir A.E. (2020). "'Riots,' 'violence,' 'looting': Words matter when talking about race and unrest, experts say." Yahoo News. https://news.yahoo.com/riots-violencelooting-words-matter-133619461.html.

[8] Davenport, C. (2007). State repression and political order. *Annual Review of Political Science* 10, pp. 1–23. doi: 10.101405.143216.

[9] Escobar, N. and Osterweil, V. (2020). "One Author's Controversial View: 'In Defense of Looting"'. NPR. https://www.npr.org/sections/codeswitch/2020/08/ 27/906642178/one-authors-argument-in-defense-of-looting.

[10] Fudenberg, D. and Tirole, J (1991). "Game Theory". Massachusetts Institute of Technology. isbn: 9780262061414.

[11] Gould, E. and Klor, E. (2010). Does Terrorism Work? *The Quarterly Journal of Economic*s 125.4, pp. 1459–1510. doi: https://doi.org/10.1162/qjec.2010.125.4.1459.

[12] Kishi, R. and Jones, S. (2020). Demonstrations & Political Violence in America. Armed Conflict Location & Event Data Project. https://acleddata.com/2020/09/03/demonstrations-politicalviolence-in-america-new-data-for-summer-2020/.

[13] Liu, E. and Roth, B (2017). Keeping the Little Guy Down: A Debt Trap for Lending with Limited Pledgeability. SSRN. doi: 10.2139/ssrn.3080682.

[14] Lossin, R.H. (2020). In Defense of Destroying Property. The Nation. https://www.thenation.com/article/activism/blm-looting-protest-vandalism/.

[15] Maskin, E. and Tirole, J. (2001). Markov Perfect Equilibrium. *Journal of Economic Theory*, pp. 191–219. doi: 10.1006.jeth.2000.2785.

[16] Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics* 12.2, pp. 380–391. doi: 10.2307/ 3003562.

[17] Pierskalla, J. H. (2010). Protest, Deterrence, and Escalation: The Strategic Calculus of Government Repression. *The Journal of Conflict Resolution* 54.1, pp. 117–145. doi: 10.1177/0022002709352462.

[18] Rubinstein, A. (1982). A Perfect Equilibrium in a Bargaining Model. *Econometrica* 50.1, p. 97. doi: 10.2307/1912531

[19] Subramanian, R. and Arzy, L. (2021). "State Policing Reforms Since George Floyd's Murder". Brennan Center for Justice. https://www.brennancenter.org/our- work/researchreports/state-policing-reforms-george-floyds-murder.


## APPENDIX

### A. Existence

**Proposition 4** Fix $T \in \mathbb{N}$. In the $T$-period version of the game, there exists a set of functions $\{r^t : \Theta \to \mathbb{R}\}_{t=0}^T$ and subsets of $\Theta$, $\{B_t\}_{t=0}^T$ such that the stage game strategies

$$\sigma_t^G(\theta) = \begin{cases} C \text{ if } \theta \in B_t \\ W \text{ if } \theta \in B_t^c \end{cases} \quad \text{and } \sigma_t^A(\theta) = \begin{cases} r^t(\theta) \text{ if } \theta \in B_t \\ 0 \text{ if } \theta \in B_t^c \end{cases}$$

induce a strategy profile $(\sigma^G, \sigma^A)$ that composes a Markov perfect equilibrium. *Proof:* The proof proceeds in three steps and utilizes backwards induction (by the same argument as in the proof of Proposition (3), this is sufficient to give a markov perfect equilibrium).

Step 1: The Base Case. This is similar to the case in Proposition (3); objects ought to be backwards-inducted (though in arbitrary time). The rioting function $r^t$ must be set to make the government indifferent between waiting and conceding while also minimizing $\eta(r^T(\theta))$; since $r^T(\theta)$ cannot be negative, then

$$r^T(\theta) = \begin{cases} 1 - c(\theta) - R_G \text{ if } c(\theta) \leq 1 - R_G \\ 0 \text{ else} \end{cases}$$

If $r^T(\theta)$ is played, then, it is always a best response by the government to concede; next, it is necessary to find the set $B_T \subset \Theta$ s.t. it is a best response by the activist to adopt this strategy. In particular, let $B_T = \{\theta : \eta(r^T(\theta)) \leq R_A\}$; that is, they get more surplus from forcing a concession than letting a government wait. Then the induced stage-game strategy profile $(\sigma_T^G, \sigma_T^A)$ are best responses to one another.

Step 2: Constructing $r^t$ inductively. The base case is now done; assume the equilibrium objects have been defined for times $\{j\}_{j=t+1}^T$. At time $t$, let the ex-ante utility of waiting by the government be

$$\begin{aligned}\omega_t(\theta) =& \delta_G \int_\Theta [R_G \mathbf{1}_{B_{t+1}} + \omega_{t+1}(\theta')\mathbf{1}_{B_{t+1}^c}] f(\theta'|\theta) d\theta' \\ &- c(\theta) - \sigma^t(\theta)\end{aligned}$$

For brevity's sake, let $\xi_t(\theta)$, $\varphi_t(\theta)$ be defined s.t.

$$\xi_{t+1}(\theta) = R_G \mathbf{1}_{B_{t+1}} + \omega_{t+1}(\theta)\mathbf{1}_{B_{t+1}^c}$$

and

$$\psi_t(\theta) = \delta_G \int_\Theta \xi_{t+1}(\theta') f(\theta'|\theta) d\theta'$$

Then the goal for rioting is to make the government indifferent between waiting and concession; since concession always gives $R_G$, then it is sufficient to set

$$r^t(\theta) = \max\{\psi_t(\theta) - c(\theta) - R_G, 0\}$$

Noting that this is equivalent to the piecewise definition given above. Then $r^t$ is a nonnegative value s.t. waiting is a best response by the government. Moreover, when $r^t(\theta) = \sigma_t^A(\theta)$, then it must also be the best response for the activist, as any lower value implies the government waits (the activist always weakly prefers concession; see Lemma (7.2)), while a higher value cannot be optimal as a lower value of rioting strictly increases the activist's utility while still inducing concession by the government.

Step 3: Time $t$-Equilibrium. The proof concludes by constructing $B_t$; that is, the set of all $\theta$ where $r^t(\theta)$ is indeed a best response. To do this, let $\lambda_t(\theta)$ be the expected utility if the game continues into the next period; in particular,

$$\lambda_t(\theta) = \delta_A \int_\Theta [R_A \mathbf{1}_{B_{t+1}} + \lambda_{t+1}(\theta')\mathbf{1}_{B_{t+1}^c}] f(\theta'|\theta)d\theta'$$

Then if $B_t = \{\theta : \eta(r^t(\theta)) \leq R_A - \lambda_t(\theta)\}$, this will exactly be the set s.t. $\lambda_t(\theta) \leq R_A - \eta(r^t(\theta))$; that is, the states where the activist prefers to induce concession. Thus, the above logic is sufficient to suggest the induced strategy profile $(\sigma_t^A(\theta), \sigma_t^G(\theta))$ is an equilibrium, and indeed is Markov perfect.

### *B. Properties*

**Proposition 6** The equilibrium objects $\{r^t, \lambda_t\}_{t=0}^T$ are continuous in $\Theta$. $\{\omega_t, \sigma_A^t\}$ are continuous on $(B_t, B_t^c)$ respectively. *Proof:* First, $\psi_t(\theta)$ and $\lambda_t(\theta)$ are continuous. If $t = T$ these are $1, 0$ respectively and the proof is done. Else, for $\psi_t(\theta)$, it is sufficient by Lemma (6.1) to show $\xi_{t+1}(\theta) = R_G \mathbf{1}_{B_{t+1}} + \omega_{t+1}(\theta)\mathbf{1}_{B_{t+1}^c}$ is bounded. This follows immediately by noting that $\theta \in B_{t+1}^c \iff \omega_{t+1} \geq R_G$ (see Lemma (11.1)), so the function is bounded below by $R_G$, and that $\omega_t$ is bounded above by 1 for all $\theta$ as all flows costs are weakly negative and total surplus cannot exceed 1. Similarly, $\lambda_t(\theta)$ is continuous if $\varphi_{t+1}(\theta)$ is bounded, which follows by noting that if $\theta \in B_{t+1}$, then by assumption $R_A - \eta(r^{t+1}(\theta)) \geq \lambda_{t+1}(\theta)$, and so the function is bounded from below by $\lambda_{t+1}(\theta)$, which is nonnegative as no flow costs are imposed on the activist, and above by $R_A$, as $\eta(r^t(\theta))$ has range $[0, \infty)$. From here, continuity of $r^t(\theta)$ follows by recognizing that $r^t(\theta) = \max\{\psi_t(\theta) - c(\theta) - R_G, 0\}$, the maximum of two continuous functions, which is continuous. Finally, $\sigma_t^A(\theta)$ and $\omega_t(\theta)$. On $B_t$, $\sigma_t^A(\theta) = r^t(\theta)$, while $\sigma_t^A = 0$ on $B_t^c$; both are continuous. Similarly, $\omega_t(\theta) = \psi_t(\theta) - c(\theta) - \sigma_t^A(\theta)$. The first two are continuous on $\Theta$, while the last is continuous componentwise, so $\omega_t$ must be continuous componentwise.

**Lemma 6.1** Let $f(y, x) : [a, b]^2 \to \mathbb{R}$ be a nonnegative totally bounded function. If $g(y) : [a, b]$ is pointwise bounded by an integrable function over $[a, b]$, then the function

$$q(x) = \int_a^b g(y)f(y, x)dy$$

is continuous if $f(y, x)$ is continuous in $x$. Moreover, if $f(y, x)$ is differentiable in $x$, then $q(x)$ is differentiable in $x$, with derivative given by

$$\frac{dq(x)}{dx} = \int_a^b g(y)\left(\frac{\partial f(y, x)}{\partial x}\right)dy$$

*Proof:* First, continuity. Fix a sequence $\{x_n\}$ in $[a, b]$ accumulating at $x$. Since $f(y, x)$ is totally bounded by an $M \in \mathbb{R}$ and $q(x)$ is pointwise bounded by an integrable function, say $p(x)$, then $|g(y)f(y, x)| \leq Mp(y)$ is integrable over $[a, b]$. Thus the functions in $y$, $\{g(y)f(y, x_n)\}_n$ are pointwise bounded by $Mp(y) \in \mathcal{L}^1$. Thus,

$$\begin{aligned}\lim_{n\to\infty} q(x_n) &= \lim_{n\to\infty} \int_a^b g(y)f(y, x_n)dy \\ &= \int_a^b \lim_{n\to\infty} g(y)f(y, x_n)dy \\ &= \int_a^b g(y)f(y, x)dy \\ &= q(x)\end{aligned}$$

where interchange is justified by the dominated convergence theorem, giving continuity. Differentiability follows similarly. Taking the Newton quotient gives

$$\begin{aligned}&\lim_{t\to x} \frac{q(t) - q(x)}{t - x} \\ &= \lim_{t\to x} \frac{1}{t - x}\int_a^b g(y)[f(y, t) - f(y, x)]dy \\ &= \int_a^b g(y)\left(\lim_{t\to x} \frac{f(y, t) - f(y, x)}{t - x}\right)dy \\ &= \int_a^b g(y)\left(\frac{\partial f(y, x)}{\partial x}\right)dy\end{aligned}$$

where integration in $y$ gives that $\frac{1}{t-x}$ is but a scalar, and interchange of limits is (locally) justified by dominated convergence as differentiability of $f$ in $x$ implies the sequence

$$\left\{\frac{f(y, t) - f(y, x)}{t - x}\right\}_{t\in B(x,\varepsilon)}$$

is bounded (locally) in the *second* argument and thus globally integrable in the *first* (restricting to a sufficiently small neighborhood of $x$). This finishes the argument.

**Proposition 8** Fix an equilibrium. Then for all $t \leq T$, $B_t$ is compact while $B_t^c$ is open. Moreover, the sets $B_t$ and $B_t^c$ are connected intervals. *Proof:* Let $g_t(\theta) = \eta(r^t(\theta)) + \lambda_t(\theta)$; $g_t(\theta)$ is continuous by Assumption (2). Note $B_t = \{\theta : g_t(\theta) \leq R_A\}$; equivalently, $B_t = g_t^1((-\infty, R_A]) \cap [0, 1]$. But $(-\infty, R_A]$ is closed, so $B_t$ is as well (noting this is in the subspace topology). Since $B_t$ is totally bounded, it is compact by Heine-Borel. Finally, since $B_t$ is closed, $B_t^c$ is open.

Next, connectedness. First, let $t < T$ so that Assumption (7) holds. Since $\Theta = [0, 1]$ is connected, $g_t(\Theta)$ is as well. Thus, $g_t(\Theta)$ must be an interval. Setting $A = g_t(\Theta) \cap (-\infty, R_A]$ and $A^c = g_t(\Theta) \cap (R_A, \infty)$, these must both be empty or intervals (in particular, they are both connected). Since $g$ is injective by Assumption (7), it is bijective onto $g_t(\Theta)$, and thus has continuous inverse $g_t^{-1} : g_t(\Theta) \to \mathbb{R}$. Thus, $g_t : [0, 1] \to g_t(\Theta)$ is a homeomorphism, which implies $B_t = g_t^{-1}(A)$ and $B_t^c = g_t^{-1}(A^c)$ are connected (and thus intervals). Finally, consider the not necessarily injective function $g_T(\theta)$. Since $\lambda_T(\theta) = 0$, this function is just $\eta(r^T(\theta))$, which is monotone decreasing. But then clearly $B_T$ is an interval, as if $x, y \in B_T$, with $x < y$, then $\eta(r^T(z)) \in [\eta(r^T(y)), \eta(r^T(x))] \subset B_T$ by the definition of $B_T$. A similar argument gives, $B_T^c$ must be an interval as well.

We will note that restricting the assumption to $t < T$ is actually necessary; if $\eta(r^T(\theta))$ is injective, then $r^T(\theta)$ is as well, implying that it may only reach 0 at one point (note it must be at 1). This implies that $\omega_T(\theta) = R_G$ on $[\hat{\theta}, 1)$ and $\omega_T(\theta) = 0$ on 1, a contradiction. So no equilibrium exists if Assumption (7) is extended to $t = T$. In other instances, since $\mu(B_T)$ is of positive measure by continuity of $c(\theta)$ and the boundary condition, $\lambda_T(\theta)$ is nonzero and monotone increasing, and so this problem may be obviated while still working in existing equilibria[9].

Next is the proof of proposition (10). Before supplying it, a key technical remark should be made here about the need for Assumption (7). The mathematical intuition is that injectivity forces sufficient regularity on the decisionmaker that their choice to riot is almost everywhere locally "smooth," which allows for systematic interrogation of the (almost) continuous equilibria object $\omega_{t+1}(\theta)$. Absent injectivity, for example, it may be possible to construct a stage-$t$ equilibrium where the activist riots on only the fat Cantor set, and nowhere else! Since the equilibrium objects are constructed via backward induction, pathology of $\omega_{t+1}$ has downstream effects to future objects, and must be controlled for (injectivity, however, is not the weakest condition that forces regularity: see Appendix (II) for a deeper discussion).

**Proposition 10** Fix an equilibrium. Then for all $0 \le t \le T$, $r^t(\theta)$ and $\omega_t(\theta)$ are decreasing on $\Theta$, while ex-ante activist surplus $\lambda_t(\theta)$ is increasing over $\Theta$. *Proof:* The proof proceeds in four independent steps.

Step 1: The Base Case. Fix $t = T$. Then $\lambda_T = 0$, which finishes this argument. Since $r^T(\theta) = \max\{1 - R_G - c(\theta), 0\}$ and $1 - R_G - c(\theta)$ is decreasing as $c(\theta)$ is increasing, then this function must be decreasing as well. Finally, $\omega_T(\theta) = 1 - c(\theta)$ on $B_T^c$, while on $B_T$, $\omega_T(\theta) = \min\{1 - c(\theta), R_G\}$ as if the activist riots, they set surplus of concession to $R_G$, and otherwise, $1 - c(\theta) < R_G$ is collected. These are both decreasing as well, and thus decreasing on $B_T$. Note moreover that $0 \in B_T^c$ then implies $\omega_T(\theta)$ is everywhere declining, as $1 - c(\theta) \ge \min\{1 - c(\theta), R_G\}$ for fixed $\theta$ and $1 - c(\theta)$ is decreasing over $[0, 1]$.

Step 2: $\psi_t(\theta)$ is decreasing. First, $\xi_{t+1}(\theta)$ is decreasing. There are two cases. First, if $B_t^c$ is empty, then $\Theta = B_t$, so $\xi_{t+1}(\theta) = R_G$ which is constant. Else, $B_t^c$ is nonempty and by Lemma (1) and the interval structure of the sets, $0 \in B_t^c$ and there exists some $\theta^*$ s.t. $[0, \theta^*) = B_t^c$. Recall that

$$\xi_{t+1}(\theta) = R_G \mathbf{1}_{B_{t+1}} + \omega_{t+1}(\theta)\mathbf{1}_{B_{t+1}^c}$$

which is decreasing on both pieces. Moreover, by Lemma (10.1), if $\theta \in B_{t+1}^c$, then $\omega_{t+1}(\theta) > R_G$, and so since $B_{t+1}^c$ is on the left, $\xi_{t+1}(\theta)$ is everywhere (weakly) decreasing. Thus, $-\xi_{t+1}(\theta)$ is increasing. Thus, by regularity of $f$, if $\theta_1 \le \theta_2$, then

$$\begin{aligned} -\psi_t(\theta_1) &= \delta_G \int_\Theta -\xi_{t+1}(\theta) f(\theta|\theta_1) d\theta \\ &\le \delta_G \int_\Theta -\xi_{t+1}(\theta) f(\theta|\theta_2) d\theta \\ &= -\psi_t(\theta_2) \end{aligned}$$

and thus $\psi_t(\theta_1) \ge \psi_t(\theta_2)$, so $\psi_t(\theta)$ is decreasing.

Step 3: The Inductive Step for $r^t, \omega_t$. Note at equilibrium that $r^t(\theta) = \max\{\psi_t(\theta) - c(\theta) - R_G, 0\}$. The first argument is decreasing by Step (2) an the fact $c(\theta)$ is increasing (so it is the sum of decreasing functions). Thus, the maximum of the continuous decreasing function and a constant is decreasing, so $r^t(\theta)$ is as well. For $\omega_t(\theta)$, it must be that $\omega_t(\theta) = \psi_t(\theta) - c(\theta) - \sigma_t(\theta)$. For it to be globally decreasing, note that $\omega_t(\theta) = \psi_t(\theta) - c(\theta)$ on $B_t^c$, an interval including 0, and $\omega_t(\theta) = \min\{\psi_t(\theta) - c(\theta), R_G\}$ on $B_t$. Since $B_t$ is globally to the right of $B_t^c$ and $\psi_t(\theta) - c(\theta)$ is globally decreasing, $\omega_t(\theta)$ is as well.

Step 4: The Inductive Step for $\lambda_t$. The key again is to show that $\varphi_t(\theta)$ is increasing and then use stochastic dominance. To do this, note that $\lambda_{t+1}(\theta)$ is increasing by the inductive hypothesis over $[0, 1]$, and $\eta \circ r^{t+1}$ is decreasing as $\eta$ is increasing and $r^{t+1}$ is decreasing. Thus, since $\theta \in B_{t+1}$ implies $R_A - \eta(r^t(\theta)) > \lambda_t(\theta)$ by definition, $\varphi_t(\theta)$ is globally increasing, and thus, an identical argument as that in Step (2) (without multiplying first by $-1$) gives $\lambda_t(\theta)$ is increasing as well.

**Lemma 10.1** At equilibrium, $\omega_t(\theta) \le R_G$ if and only if $\theta \in B_t$, and $\lambda_t(\theta) < R_A$ for all $\theta$. *Proof:* This follows from standard equilibrium reasoning. Since $\theta \in B_t$ implies concession, then concession follows if and only if it gives the government weakly more utility. Second, x $\lambda_t(\theta) \le R_A$. To do this, note $\lambda_T(\theta) = 0$. By a backwards induction argument,

$$\begin{aligned} \lambda_t(\theta) = \delta_A \int_\Theta \Big( &[R_A - \eta(r^{t+1}(\theta'))]\mathbf{1}_{B_{t+1}} \\ &+ \lambda_{t+1}(\theta')\mathbf{1}_{B_{t+1}^c} \Big) f(\theta'|\theta) d\theta' \end{aligned}$$

Noting that $\eta \cdot r^{t+1}$ is nonnegative and that $\lambda_{t+1}(\theta) \le R_A$ by the inductive hypothesis, this is bounded from above by

$$\delta_A \int_\Theta R_A f(\theta'|\theta) d\theta' = \delta_A R_A$$

as $f(\theta'|\theta)$ is a probability distribution function.

**Proposition 11** Let $\theta \in B_t^c$. Then $R_G < \omega_t(\theta) \le \delta_G^{T-t}$. Thus, $\lim_{T \to \infty} B_T(0) = \Theta$. *Proof:* The proof proceeds via backwards induction. First, at $T$, $\omega_T(\theta) \in B_T^c$ implies by Lemma (11.1) that $\omega_T(\theta) > R_G$. Moreover, since total surplus is 1, $\omega_T(\theta) \le 1$. For $t < T$,

$$\delta_G \int_\Theta R_G f(\theta'|\theta) d\theta' \le \omega_t(\theta) \le \delta_G \int_\Theta \delta_G^{T-(t+1)} f(\theta'|\theta) d\theta'$$

by the inductive hypothesis. Since $f(\cdot|\theta)$ is a pdf, the bound holds. Note the lower bound is sharper, as $\theta \in B_t^c$, by Lemma (11.1).

[9] Should this problem still worry the reader, however, it would be equivalent to consider the more technical subspace topology over $[0, \tilde{\theta}] \subset [0, 1]$ where $\tilde{\theta}$ is the leftmost point for each stage-game equilibrium where $r^t(\theta) = 0$ and restrict Assumption (7) to be injective over $[0, \tilde{\theta}]$. This can be done without loss, and still ensures connectedness in the domain by taking the union later.

**Proposition 13** Fix an equilibrium satisfying Assumption (12). Then for all $t < T$, $\omega_t(\theta) \leq \omega_{t+1}(\theta)$ and $r^{t+1}(\theta) \geq r^t(\theta)$ for fixed $\theta$. For the activist, $\lambda_{t+1}(\theta) \leq \lambda_t(\theta)$.
*Proof:* We split the proof up into steps, as usual. First, consider $\omega_t$ and $r^t$.
Step 1: The Base Case for $\omega_t$, $r^t$. First, note that $\psi_T(\theta) = 1$; since $\xi_T(\theta) = R_G \mathbf{1}_{B_T} + \omega_T(\theta)\mathbf{1}_{B_T^c}$, then $|\xi_T(\theta)| \leq 1$ as $\omega_T(\theta) = 1 - c(\theta)$ and thus $\psi_{T-1}(\theta) \leq 1$. Thus, $\psi_{T-1}(\theta) \leq \psi_T(\theta) = 1$. From here, note

$$\begin{aligned} r^T(\theta) - r^{T-1}(\theta) = &\max\{\psi_T(\theta) - c(\theta) - R_G, 0\} \\ &- \max\{\psi_{T-1}(\theta) - c(\theta) - R_G, 0\} \\ \geq &0 \end{aligned}$$

by considering the argument piecewise and noting that $\psi_{T-1}(\theta) \leq \psi_T(\theta)$ implies that if $r^T(\theta) = 0$, then $r^{T-1}(\theta) = 0$ and otherwise this difference must be nonnegative. Thus, $r^T(\theta) \geq r^{T-1}(\theta)$. Next, for $\omega_T$, note that by Assumption (13), the difference of the functions is

$$\omega_T(\theta) - \omega_{T-1}(\theta) = \begin{cases} R_G - R_G \text{ if } \theta \in B_T \cap B_{T-1} \\ 1 - \psi_{T-1}(\theta) \text{ if } \theta \in B_T^c \cap B_{T-1}^C \\ 1 - c(\theta) - R_G \text{ if } \theta \in B_T^c \cap B_{T-1} \end{cases}$$

The first two cases are trivially nonnegative. The last case follows by noting $\theta \in B_T^c$ gives $1 - c(\theta) > R_G$. This finishes the base case.
Step 2: The Inductive Step for $\omega_t, r^t$. First, $\psi_t(\theta)$ is increasing in $t$. Since $f(\cdot|\theta)$ is time-homogenous, for all $t$, $\psi_{t+1}(\theta) - \psi_t(\theta)$ can be rewritten as

$$\psi_{t+1}(\theta) - \psi_t(\theta) = \delta_G \int_\Theta [\xi_{t+2}(\theta') - \xi_{t+1}(\theta')] f(\theta'|\theta) d\theta'$$

And thus $\psi_t$ increases in time if the function $\xi_{t+2}(\theta') - \xi_{t+1}(\theta')$ is nonnegative. Rewriting this difference explicitly gives

$$\begin{aligned} \xi_{t+2}(\theta') - \xi_{t+1}(\theta') = &R_G[\mathbf{1}_{B_{t+2}}(\theta') - \mathbf{1}_{B_{t+1}}(\theta')] \\ &+ \omega_{t+2}(\theta')\mathbf{1}_{B_{t+2}^c}(\theta') \\ &- \omega_{t+1}(\theta')\mathbf{1}_{B_{t+1}^c}(\theta') \end{aligned}$$

and so there are a few cases to consider. First, if $\theta' \in B_{t+2} \cap B_{t+1}$, then the difference is 0. If $\theta' \in B_{t+2}^c \cap B_{t+1}^c$, then the difference is $\omega_{t+2}(\theta') - \omega_{t+1}(\theta')$, which is nonnegative by the inductive hypothesis. By Assumption (13), $B_{t+1}^c \cap B_{t+2} = \varnothing$ , since $B_{t+2} \subset B_{t+1}$. Thus, the last case if if $\theta' \in B_{t+2}^c \cap B_{t+1}$. Then the difference is given by $-R_G + \omega_{t+2}(\theta')$. Since $\theta' \in B_{t+2}^c$, then $\omega_{t+2}(\theta') > R_G$ at equilibrium, so this term is positive. Since $\psi_{t+1}(\theta) - \psi_t(\theta) > 0$, then the inductive case is *identical* to the argument in the base case, noting the use of Assumption (13) in its full strength to finish the induction.
Step 3: $\lambda_t$ decreasing in $t$. For the base case, note

$$\lambda_{T-1}(\theta) = \delta_A \int_\Theta [R_A - \eta(r^T(\theta'))]\mathbf{1}_{B_T} f(\theta'|\theta) d\theta' \geq \lambda_T(\theta)$$

noting that $\lambda_T(\theta) = 0$ and that $\theta \in B_T$ if and only if

$$R_A - \eta(r^T(\theta)) \geq 0.$$

For the inductive step, a similar argument as in Step 2(A) implies it is sufficient to show $\varphi_{t+2}(\theta) - \varphi_{t+1}(\theta)$ is nonpositive. Expanding, this is

$$\begin{aligned} &R_A(\mathbf{1}_{B_{t+2}} - \mathbf{1}_{B_{t+1}}) - (\eta(r^{t+2}(\theta))\mathbf{1}_{B_{t+2}} - \eta(r^{t+1}(\theta))\mathbf{1}_{B_{t+1}}) \\ &+ \lambda_{t+2}(\theta)\mathbf{1}_{B_{t+2}^c} - \lambda_{t+1}(\theta)\mathbf{1}_{B_{t+1}^c} \end{aligned}$$

The same three cases apply as in the first part of Step 2. First, if $\theta \in B_{t+1} \cap B_{t+2}$, t hen the value is exactly $-(\eta(r^{t+2}(\theta)) - \eta(r^{t+1}(\theta)))$. Since $\eta$ is increasing and by Part (2) $r^{t+2}(\theta) \geq r^{t+1}(\theta)$, this difference is negative. Second, if $\theta \in B_{t+2}^c \cap B_{t+1}^c$, then the value is $\lambda_{t+2}(\theta) - \lambda_{t+1}(\theta) \leq 0$ by the inductive hypothesis. Finally, if $\theta \in B_{t+2}^c \cap B_{t+1}$, (Assumption (13) rules out the other case) then the relevant difference is

$$\begin{aligned} \lambda_{t+2}(\theta) - (R_A - \eta(r^{t+1}(\theta))) \leq \lambda_{t+2}(\theta) - (R_A - \eta(r^{t+2}(\theta)) \\ < 0 \end{aligned}$$

where the first inequality follows by the inductive hypothesis and the increasingness of $\theta$ and the second argument occurring as $\theta \in B_{t+2}^c$. This finishes the entire argument.

### C. *Welfare*

**Proposition 15** A stochastic ex-ante optimal strategy exists.
*Proof:* Existence. It is sufficient to construct a sequence of equilibrium strategy objects using backwards induction on the finite-horizon dynamic programming problem to give an optimal strategy. Consider time $T$ and let $C_T$ be the set under which concession is optimal at time $T$. Then

$$\begin{aligned} s_T(\theta_T) = C &\iff \delta_G^T R_G - \sum_{j=0}^{T-1} \delta_G^j c(\theta_j) \geq \delta_G^T - \sum_{j=0}^{T} \delta_G^j c(\theta_j) \\ &\iff \delta_G^T(1 - R_G) \leq \delta_G^T c(\theta_T) \end{aligned}$$

and so $s_T(\theta_T) = C$ if $(1 - R_G) \leq c(\theta_T)$ and $s_T(\theta_T) = W$ otherwise is optimal at time $T$. Next, the inductive step. Fix $t < T$ and assume that the sets $\{C_j\}_{j=t+1}^T$ have been defined. The time-$j$ strategies induced by $\{C_j\}$ coupled with the probability distribution $f(\cdot|\theta)$ are sufficient to induce a probability measure $\mathbb{P}(\theta, \{s_j\})$ over the terminal nodes still on the path of play from $t+1$ to $T$. Let

$$\kappa_t(\theta) = \delta_G \int_\mathcal{Z} p(z) d\mathbb{P}(\theta, \{s_j\})$$

be the expected utility collected by continuing into time $(t+1)$. Note if $t = T$, then $\kappa_T = 1$ as continuation is not possible. From here, define $C_t$ to be

$$C_t = \{\theta : \kappa_t(\theta) - R_G \leq c(\theta)\}$$

which, by a similar argument given in the base case, is exactly the set where the expected gain in utility from not conceding is smaller than the incurred cost of continuation. Thus the strategy induced by the time-$t$ strategies given by $s_t(\theta) = C$ on $C_t$ and $s_t(\theta) = W$ otherwise is optimal for each $\theta$ and moreover optimal more generally. Note that optimal strategies can without loss be taken as state-contingent, as time-$t$ payoffs are invariant under past realizations of the history (these are "sunk cost") and thus it is without loss to not condition the strategy on the history, just the current state. Thus the induced strategy $s$ above is ex-ante optimal.

**Proposition 16** Fix equilibrium objects $\{c, \eta, f(\cdot|\theta)\}$ and consider the equilibria objects in Proposition (4) and the optimal strategy from Proposition (16). Then if $\{B_t, \omega_t, \lambda_t, r^t\}$ and $\{C_t, \kappa_t\}$ are the respective relevant induced objects,

1) For all $t \le T$, $\psi_t(\theta) \le \kappa_t(\theta)$ and $C_t \subset B_t$.
2) For all $t$ such that $B_t^c \neq \varnothing$, $\{\theta : \sigma_t^A(\theta) \neq 0\}$ is nonempty and $\psi_t < \kappa_t$.

*Proof:* The first statement. Fix $t = T$ First, by definition, $\kappa_T(\theta) = 1 = \psi_T(\theta)$ so $\psi_T(\theta) \le \kappa_T(\theta)$. From here, let $\theta \in C_T$. Then $1 - c(\theta) \le R_G$. Thus, $r^t(\theta) = 0$, so the boundary condition gives $\eta(r^t(\theta)) = 0$, and thus $0 \le R_A - \lambda_T(\theta)$ so $\theta \in B_T$. This gives the base case. For the inductive step, fix $t < T$, and note that $\kappa_t$ can be defined inductively in a manner similar to $\psi_t$; namely, their difference is, for fixed $\theta^* \in \Theta$, is

$$\delta_G\Bigg(\int_\Theta (R_G(\mathbf{1}_{C_{t+1}} - \mathbf{1}_{B_{t+1}}) + (\kappa_{t+1}(\theta) - c(\theta))\mathbf{1}_{C_{t+1}^c} - \omega_{t+1}(\theta)\mathbf{1}_{B_{t+1}^c})f(\theta|\theta^*)d\theta\Bigg).$$

It is sufficient to prove the integrand is nonnegative. Recall over $B_{t+1}^c$, $\sigma_t^A(\theta) = 0$, and $C_{t+1} \subset B_{t+1}$ by the inductive hypothesis. Thus, rewrite the integral as

$$(\kappa_{t+1}(\theta) - c(\theta) - R_G)\mathbf{1}_{C_{t+1}^c \cap B_{t+1}} + (\kappa_{t+1}(\theta) - \psi_{t+1}(\theta))\mathbf{1}_{C_{t+1}^c \cap B_{t+1}^c}.$$

Clearly, the second term is nonnegative by the inductive hypothesis, while over the first term, if $\theta \in C_{t+1}^c$, then $\kappa_{t+1}(\theta) - c(\theta) > R_G$ by definition, giving that this term is positive. Thus $\psi_t(\theta) \le \kappa_t(\theta)$. From here, let $\theta^* \in C_t$. Then

$$\begin{aligned}&\psi_t(\theta^*) - c(\theta^*) \le \kappa_t(\theta^*) - c(\theta^*) \le R_G\\ \Longrightarrow\ & r^t(\theta^*) = 0\\ \Longrightarrow\ & \eta(0) = 0 \le R_A - \lambda_t(\theta^*)\end{aligned}$$

and thus $\theta^* \in B_t$ as well.

The second statement. It is sufficient for $\sigma_t^A(\theta)$ to not be identically 0. To do this, first note that the inequality in Lemma (10.1) is strict and global. Yet by Proposition (6), $\lambda_t(\theta)$ is continuous and thus achieves a maximum as its domain, $[0, 1]$, is compact. Thus, $R_A - \lambda_t(\theta) \ge \varepsilon(t)$ for some $\varepsilon(t) > 0$, dependent on $t$. Fix $t$ s.t. $B_t^c \neq \varnothing$ and let $\varepsilon = \varepsilon(t)$. There exists some $r^t(\theta)$ such that $\eta(r^t(\theta)) < \varepsilon$. Since $\eta$ is continuous, then there exists $\delta > 0$ s.t. $\eta([0, \delta)) \subset [0, \varepsilon)$, and so it is sufficient for $r^t(\theta)$ to be positive at some point $\theta$ (as by the intermediate value theorem, this implies that $r^t(\theta)$ admits a positive interval in the range and thus hits some value in $(0, \varepsilon)$). Assume not. Then $r^t(\theta) = 0$ identically, and so $\eta(0) \le R_A - \lambda_t(\theta)$ always, so $B_t = [0, 1]$, contradicting the fact that $B_t^c$ was nonempty. Finally, the *strict* nature of the welfare effects. By (2), whenever $\theta \in C_t$, rioting is zero and $\theta \in B_t$. Thus, the existence of nonzero rioting implies that $C_t \subset B_t$, and moreover, by continuity of $r^t(\theta)$, that $B_t \backslash C_t$ is of positive measure. Thus, inspecting (1), the set $C_{t+1}^c \cap B_{t+1}$ is of positive measure, and $\kappa_{t+1}(\theta) - c(\theta) - R_G > 0$ is strict, and so one has that

$$\begin{aligned}&\kappa_t(\theta^*) - \psi_t * (\theta^*)\\ &> \int_{C_{t+1}^c \cap B_{t+1}} (\kappa_{t+1}(\theta) - c(\theta) - R_G)f(\theta|\theta^*)d\theta\\ &> 0\end{aligned}$$

borrowing from the equations in (1), for all $\theta^* \in \Theta$. Note the sufficient volatility condition guarantees the lower bound, as the set integrated over is not of zero measure with respect to $f(\theta|\theta^*)$. Thus this inequality is strict so long as the rioting problem is not trivial, finishing the desired proof.

### *D. Weakening the Intratemporal Assumption (7)*

**Proposition A1** Let $E_t = \{\theta : \eta(r^t(\theta)) + \lambda_t(\theta) = R_A\}$. Then if $E_t$ contains at most one point, then $\{B_t\}$ and $\{B_t^c\}$ are connected (and thus intervals). *Proof:* First, note the only connected subsets of $\mathbb{R}$ are intervals, so we need only prove the sets are connected. If either $B_t$ or $B_t^c$ is connected, then the other must be as well, as the intersection of two intervals is likewise still an interval. Without loss, assume that $B_t$ is not connected. Then there exists $x, y \in B_t$ and $z \in B_t^c$ s.t. $x < z < y$. From here, consider $g_t(\theta) = \eta(r^t(\theta)) + \lambda_t(\theta)$, which implies $g_t(x) \le R_A$, $g_t(y) > R_A$, and $g_t(z) \le R_A$. As $g_t(\cdot)$ is continuous, then there exists $\theta_1 \in [x, y)$ and $\theta_2 \in (y, z]$ s.t. $g_t(\theta_1) = R_A = g_t(\theta_2)$, contradicting the assumption that $E_t$ contains at most one point.

Note the converse is not true; it is possible, even with an interval structure, that many values of $\theta$ satisfy the equality given in the above equation. There is a notion of "single-crossing" monotonicity, but it is a weird and not particularly intuitive condition, so we eschew it in favor of the machinery afforded by differentiable functions. In particular, this allows us to consider a sufficient condition for the interval structure of $B_t$ in the context of the *marginal* costs and benefits of rioting, and gives economic intuition behind the assumption. First, a pair of assumptions.

**Assumption A2** The functions $\eta(\cdot)$ and $c(\cdot)$ are differentiable. Moreover, the probability distribution function $f(\cdot|x)$ is differentiable in $x$.

Note Assumption (A2), coupled with Lemma (6.1), immediately implies (in a manner similar to that of Proposition (6)), that the equilibrium objects $\{r^t, \lambda_t, \omega_t\}$ are differentiable; moreover, if $\{\eta, c, f(\cdot|x)\}$ were *continuously* differentiable, then $\{r^t, \lambda_t, \omega_t\}$ would have continuous derivative as well. This is a standard argument, but allows the application of tools from differential calculus (and, by extension, considerations on convexity), to study the intertemporal argument used here. Yet differentiability alone is insufficient to guarantee the singularity of the point of indifference; instead, the *relative magnitudes* of the derivatives themselves.

**Assumption A3** For all $\theta \in \Theta$ and $t \le T$, $\frac{d}{d\theta}[\eta(r^t(\theta)) + \lambda_t(\theta)] \neq 0$.

Note this assumption is not required at $t = T$ given the proof method in Proposition (8), highlighting again how the reason further assumptions are needed are because of the conflicting monotonicities of $\lambda_t$ and $\eta(r^t(\theta))$ when $t < T$ (that is, the tension generated by a higher ex-ante payoff as $\theta$ increases and greater cost as $\eta$ changes as well). However, for all $1 \leq t \leq T$, the following characterization is obtained:

**Proposition A4** Let Assumptions (A2) and (A3) hold. Then for all $1 \leq t \leq T$, $|E_t|\leq 1$. *Proof:* The contrapositive is shown. Assume that $E_t$ contains at least two points, $\theta_1$ and $\theta_2$, with $\theta_1 \neq \theta_2$ and again let $g_t(\theta) = \eta(r^t(\theta)) + \lambda_t(\theta)$. Then one has that

$$\frac{g_t(\theta_1) - g_t(\theta_2)}{\theta_1 - \theta_2} = 0 \implies \frac{d}{d\theta} g_t(\theta^*) = 0$$

for some $\theta^* \in (\theta_1, \theta_2)$ by the mean value theorem, a contradiction.

Fix $1 \leq t \leq T$ and differentiate $g_t(\theta)$. Rearranging, Assumption (A3) can becomes

$$\frac{d\eta}{dr^t}\frac{dr^t}{d\theta} + \frac{d\lambda_t}{d\theta} \neq 0 \iff \frac{d\eta}{dr^t}\frac{dr^t}{d\theta} \neq -\frac{d\lambda_t}{d\theta}$$

where this identity need only hold locally for each $\theta \in \Theta$. However, this inequality can be interpreted in a strict sense.

**Proposition A5** Let Assumption (A2) hold. Then Assumption (A3) holds if and only if, for all $\theta \in \Theta$

$$\frac{d\eta}{dr^t}\frac{dr^t}{d\theta} < -\frac{d\lambda_t}{d\theta} \text{ or } \frac{d\eta}{dr^t}\frac{dr^t}{d\theta} > -\frac{d\lambda_t}{d\theta}$$

*Proof:* This is immediate by the intermediate value theorem for derivatives. In particular, assume not. Then there exists $\theta_1$ and $\theta_2$ satisfying both sides of the inequality, implying that there exists $\theta^* \in (\theta_1, \theta_2)$ such that

$$\frac{d\eta}{dr^t}\frac{dr^t}{d\theta} = -\frac{d\lambda_t}{d\theta}$$

contradicting Assumption (A3). The converse follows immediately by the discussion above prior to the oroposition, noting that the two inequalities must each hold *globally*.

Locally, consider the derivatives as behavior of the system with respect to local perturbations. In particular, $\eta(r^t(\theta)) + \lambda_t(\theta)$ can be interpreted as the opportunity cost of support; thus, Assumption (A3) is equivalent to requiring that, as $\theta$ is perturbed, the problem facing the activist *changes*: at each level of support, the opportunity cost of rioting will be different if $\theta$ is locally perturbed.

Globally, interpret the derivatives as the marginal costs and benefits incurred by various equilibrium objects. The negative sign in front of the static for $\lambda_t$ can be interpreted as the *foregone* cost of inaction incurred by a decision to actively riot, since $\lambda_t'(\theta)$ gives the *increase* (by Proposition (10)) in surplus due to an increase in $\theta$, its additive inverse tracks the utility lost by choosing to force a concession instead of simply waiting to the next period. In this way, our global requirement is that the marginal cost of rioting is either everywhere larger or everywhere smaller than the marginal cost of foregone waiting; the first is believable, in particular when the recoil cost function $\eta$ is sufficiently convex and thus $\frac{d\eta}{d\theta}$ is large and increasing very quickly.

One last thing to note, however, is that Assumption (A3) is not strictly weaker than Assumption (7). Notably, the global condition demonstrates (A3) is actually *equivalent* to (8) when the equilibrium objects are strictly differentiable, though the differential approach is more economically intuitive. Note, however, the assumption that $|E_t|= 1$ *is* weaker than injectivity, and is sufficient for Proposition (10) to hold, and can be interpreted by requiring that the activist is at most indifferent between forcing a concession and waiting at at most one point, and is thus a "weak uniqueness" that is, in spirit, very similar to the assumption $\eta(\cdot)$ is strict. (The government, by contrast, must be indifferent over a set of positive measure always, by Proposition (16)).

### E. Weakening the Intertemporal Assumption (12)

The weakening of assumption (12) follows a similar structure to the approach above, and is thus treated more briefly. As the model is in discrete time, finite differences are taken and differentiability is not assumed.

**Assumption A6** Fix an integral $t < T$. Then

$$\lambda_{t+1}(\theta) - \lambda_t(\theta) \leq \eta(r^t(\theta)) - \eta(r^{t+1}(\theta)) \text{ for all } \theta \in \Theta$$

**Proposition A7** Assumption (A6) implies Assumption (12). *Proof:* We give the contrapositive. Let $\theta \in B_t$, such that $\lambda_t(\theta) + \eta(r^t(\theta)) \leq R_A$. If $\theta \notin B_{t+1}$, then $\lambda_{t+1}(\theta) + \eta(r^{t+1}(\theta)) > R_A$, and thus

$$\lambda_{t+1}(\theta) - \lambda_t(\theta) + \eta(r^{t+1}(\theta)) - \eta(r^t(\theta)) > 0$$

if and only if

$$\lambda_{t+1}(\theta) - \lambda_t(\theta) > \eta(r^t(\theta)) - \eta(r^{t+1}(\theta))$$

implying that Assumption (A6) is false, as desired.

To interpret this proposition, first let Assumption (A6) hold. It will be useful to interpret this assumption through Proposition (13); in particular, note by Proposition (13) both of these terms are negative. In absolute value, then, the requirement becomes

$$|\lambda_t(\theta) - \lambda_{t+1}(\theta)|\geq \left|\eta(r^t(\theta)) - \eta(r^{t+1}(\theta))\right|$$

that is, that the total change in ex-ante expected surplus in waiting due only to temporal effects is greater than the total change of the recoil cost incurred by the change in optimal rioting that is associated with it. Note this is similar to the second global condition in Proposition (A5), where as $\eta \circ r^t$ is decreasing and $\lambda_t$ is increasing in $\theta$,

$$\frac{d\eta}{dr^t}\frac{dr^t}{d\theta} > -\frac{d\lambda_t}{d\theta} \implies \left|\frac{d\eta}{dr^t}\frac{dr^t}{d\theta}\right| < \left|\frac{d\lambda_t}{d\theta}\right|$$

highlighting that the restrictions on both the intratemporal and intertemporal problems are philosophically the same, and related to the *rate of substitutions* between the two objects which compose of the opportunity cost of rioting for the activist, and are meant to deal with, respectively, the nonmonotone nature of their sums.