GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN IN PUBLICA COMMODA
SEIT 1737

—Essays in Political Economy— Drivers of Polarization

Dissertation
zur Erlangung des wirtschafts- und sozialwissenschaftlichen Doktorgrades
"Doctor rerum politicarum"
der Georg-August-Universität Göttingen

vorgelegt von
Johannes Matzat

University of Göttingen
August 2023

BETREUER UND ERSTGUTACHTER DER DISSERTATION

Prof. Dr. Axel Dreher

Professor of Economics

Chair of International and Development Politics

Alfred-Weber-Institute for Economics

Ruprecht-Karls-Universität Heidelberg

ZWEITGUTACHTERIN DER DISSERTATION

Prof. Dr. Krisztina Kis-Katos

Professor of Economics

Chair of International Economic Policy

Department of Economics

Georg-August-Universität Göttingen

DRITTGUTACHTER DER DISSERTATION

Prof. Dr. Andreas Fuchs

Professor of Economics

Chair of Development Economics

Department of Economics

Georg-August-Universität Göttingen

Contents

List of Main Figures

List of Tables

Abstract

This thesis contributes to our understanding of the increasing polarization of societies. The first chapter focuses on immigration inflows as driver of polarization. As societies are faced with an increasing number of immigrants, immigration has become a central and polarized topic of political discourse. This chapter leverages a shift-share instrument to assess the causal impact of U.S. immigrant inflows on political ideologies. It documents that migration increased the polarization of politicians campaigning for the House of Representatives between 1992 and 2016. Subsequently, it focuses on refugees enabled by novel data covering over 3 million individuals. The results echo those for immigrants and suggest that the difference in the moral justification for welcoming refugees does not translate into a different political reaction. The second chapter analyzes whether the arrival of fast internet unites or divides Indian villages. It leverages the largest rural government broadband initiative in the world that aims to connect every Indian village to the fiber-optic network. To identify the causal effect of rural broadband internet, this paper exploits spatial discontinuities, which arose in 2017, between villages getting connected early and late due to the staggered roll-out of the broadband initiative. The paper documents an increase in divisions along several dimensions: First, assaults and riots of supporters of the Hindu nationalist party increase; second, welfare benefits in Jharkhand are increasingly distributed along religious lines; third, in Jharkhand, non-Muslim villages vote for the Hindu nationalist party while Muslim villages vote for the secular parties. The third chapter examines the political influence of labor unions. The workplace is behind family and friends the area most important for political discussions and it is directly influenced by unions. There, they may change the ideological positions of both unionizing workers and their non-unionizing management. This paper analyzes the workplace-level impact of unionization on workers' and managers' political campaign contributions over the 1980-2016 period in the United States. In a difference-in-differences design, it finds that unionization leads to a leftward shift of campaign contributions. Unionization increases the support for Democrats relative to Republicans not only among workers but also among managers, which speaks against an increase in political cleavages between the two groups. These shifts are not driven by compositional changes of the workforce but are also visible at the individual level.

Zusammenfassung

Diese Dissertation trägt zum Verständnis der zunehmenden Polarisierung von Gesellschaften bei. Das erste Kapitel konzentriert sich auf den Zustrom von Einwanderern als Ursache der Polarisierung. Die zunehmende Globalisierung führt zu einer wachsenden Zahl von Einwanderern, daher ist die Einwanderung zu einem zentralen und polarisierten Thema des politischen Diskurses geworden. In diesem Kapitel wird ein Shift-Share-Instrument eingesetzt, um die kausalen Auswirkungen des Zustroms von Einwanderern in die USA auf politische Ideologien zu bewerten. Es dokumentiert, dass Migration die Polarisierung von Politikern, die zwischen 1992 und 2016 für das Repräsentantenhaus kandidierten, verstärkte. Anschließend konzentriert sich die Studie auf Flüchtlinge, was durch neuartige Daten von über 3 Millionen Personen ermöglicht wird. Die Ergebnisse spiegeln jene für Einwanderer wider und legen nahe, dass der Unterschied in der moralischen Rechtfertigung für die Aufnahme von Flüchtlingen sich nicht in einer anderen politischen Reaktion niederschlägt. Im zweiten Kapitel wird analysiert, ob die Einführung schnellen Internets indische Dörfer eint oder spaltet. Es nutzt die weltweit größte öffentliche Breitbandinitiative für den ländlichen Raum, die darauf abzielt, jedes indische Dorf an das Glasfasernetz anzuschließen. Um den kausalen Effekt des ländlichen Breitbandinternets zu ermitteln, nutzt dieses Papier die 2017 aufgetretenen räumlichen Diskontinuitäten zwischen Dörfern, die aufgrund der gestaffelten Einführung der Breitbandinitiative früh und spät angeschlossen wurden. Das Papier dokumentiert eine zunehmende Spaltung entlang mehrerer Dimensionen: Erstens nehmen Übergriffe und Ausschreitungen von Anhängern der hindu-nationalistischen Partei zu; zweitens werden Sozialleistungen in Jharkhand zunehmend nach religiösen Gesichtspunkten verteilt; drittens wählen in Jharkhand nicht-muslimische Dörfer die hindu-nationalistische Partei, während muslimische Dörfer die säkularen Parteien wählen. Im dritten Kapitel wird der politische Einfluss der Gewerkschaften untersucht. Der Arbeitsplatz ist nach Familie und Freunden der wichtigste Ort für politische Diskussionen und wird von den Gewerkschaften direkt beeinflusst. Dort können sie die ideologischen Positionen sowohl der gewerkschaftlich organisierten Arbeitnehmer als auch des nicht gewerkschaftlich organisierten Managements verändern. In diesem Beitrag werden die Auswirkungen der gewerkschaftlichen Organisierung auf die politischen Wahlkampfspenden von Arbeitnehmern und Managern im

Zeitraum 1980-2016 in den Vereinigten Staaten auf betrieblicher Ebene analysiert. In einem Differenz-in-Differenzen-Design wird festgestellt, dass die Gewerkschaftsmitgliedschaft zu einer Linksverschiebung der Wahlkampfspenden führt. Die gewerkschaftliche Organisierung erhöht die Unterstützung für die Demokraten im Vergleich zu den Republikanern nicht nur unter den Arbeitnehmern, sondern auch unter den Managern, was gegen eine Zunahme der politischen Kluft zwischen den beiden Gruppen spricht. Diese Verschiebungen sind nicht auf Veränderungen in der Zusammensetzung der Belegschaft zurückzuführen, sondern sind auch auf individueller Ebene sichtbar.

Introduction

Addressing the world's most pressing challenges requires cooperation and trust, enabling societies to work together towards shared goals. Yet, in an age marked by widespread polarization, developing the necessary willingness to collaborate is an increasingly formidable task. The U.S. stands as a stark illustration of this divide, struggling with levels of political polarization not seen since the Civil War that culminated in the attack on the Capitol on January 6th, 2021 (Hare and Poole, 2014). Every fourth (27%) Democrat and every third (36%) Republican sees in the other party a 'threat to the nation's well-being' (PEW, 2014). This has global implications. More than 70% of foreign policy opinion leaders see U.S. polarization as a critical threat, in a 2018 survey (Busby, 2020). Polarization is not confined to the U.S. but is present in developed and developing countries alike. The world's largest country, India, experiences increasing divisions between Hindus and Muslims under the divisive leadership of the Bharatiya Janata Party (BJP). The consequences are hate crimes and lynch mobs (e.g., Human Rights Watch, 2019). Similar developments can be observed throughout the world, in countries like Turkey, Brazil, and Poland.

One aspect of polarization has been the turn towards cultural issues like immigration, ethnicity, or abortion at the expense of economic issues like taxation or redistribution (Ford and Jennings, 2020; Bonomi et al., 2021). This shift went hand in hand with the rise of divisive populist leaders heavily relying on identity politics take for example Trump in the U.S., Bolsonaro in Brazil, Modi in India, Le Pen in France, Salvini in Italy, and Farage in the U.K. One potential consequence has been the change in the structure of political cleavages, as low-income voters no longer lean to the left but divide based on their education levels between left and right (Gethin et al., 2022). Together, high polarization, concentrated around cultural issues, and changing cleavages in the electorate suggest a shift in identities, such that voters increasingly view complex policy issues through the lens of their cultural background (cultural identity) as opposed to their economic background (class identity) (Gennaioli and Tabellini, 2023).

The consequences of these high levels of polarization are dramatic. As outlined by Carothers and O'Donohue (2019) polarization undermines the function of democracies by deteriorating checks and balances, producing institutional gridlocks and diminishing public faith in various elements of the political process, such as election administration, political parties, and the political establishment more broadly. Beyond the political sphere, polarization is shaping the social fabric of societies, influencing the way people interact with each other. The patterns are best documented in the U.S. where a large share of individuals prefers to befriend people with similar political views (PEW, 2016). Family time has been observed to decline if political views do not align, notably during Trump's 2016 campaign (Chen and Rohla, 2018). In addition, the number of people being at least somewhat unhappy if their child marries someone of the opposite party has been increasing by 35 percentage points over the last 50 years (Iyengar et al., 2012). India shows a similar pattern although empirical evidence is more limited. A large share of individuals do not extend their personal circle to members of a different religion (86% of Hindus have mostly Hindu friends, 89% of Muslims) and 66% would like to stop members of their community to marry someone from a different religion (PEW, 2021). At the same time, an increase in hate crimes and religious violence has been documented (Human Rights Watch, 2019; New York Times, 2019).

The profound consequences of polarization underline the need to understand the origins of divisions. Several studies point to economic shocks, in particular globalization, automation, loss-aversion, and austerity, as contributing factors (e.g., Colantone and Stanig, 2018; Anelli et al., 2019; Fetzer, 2019; Autor et al., 2020; Hübscher et al., 2020; Panunzi et al., 2020). The prevalence of cultural issues in polarized campaigns seems at odds with an economic channel. Advocates of this channel explain it through interaction effects between cultural and economic shocks and the shifting of blame to immigrants (Rodrik, 2021). This line of reasoning is contrasted by a strand of research that view fear of cultural change as the root cause (e.g., Margalit, 2019; Norris and Inglehart, 2019). Immigrants embody cultural change, which can explain the prevalence of immigration in political debates. Due to their economic or cultural influence, the important role of immigration in polarized debates highlights the need to further our understanding of migrants' role in polarized societies. In the U.S., opponents as well as supporters of immigration doubled since 1998, leading to a drastic increase in polarization (Bonomi et al., 2021). Based on this observation, the first chapter of this thesis contributes to the debate on the drivers of polarization by examining the impact of immigration and refugee flows on the polarization of politicians and individuals in the U.S. over the 1992-2016 period. By explicitly taking ideological movements into account it adds to our understanding of the political impact of migrants (e.g., Dustmann et al., 2020; Steinmayr, 2021; Mayda et al., 2022).

Suspect drivers of polarization are not restricted to changes in realities but extend to changes in perceptions. The internet and social media, in particular, are prime suspects. Proponents of this view have argued that it can reinforce particular world views through selected content exposure (termed echo chambers, see, e.g., Bakshy et al., 2015; Halberstam and Knight, 2016 and Levy, 2021) as well as its tendency to spread activating emotional content (e.g., anger, awe, anxiety as opposed to sadness) (Berger and Milkman, 2012). The early phase of the current polarization wave predates social media, and evidence from the U.S. has not found differences in the exposure to selective media sources offline and online (Gentzkow and Shapiro, 2011). The evidence has shifted recently as social media became not only a central information source for large shares of the population but also a tool and platform for political actors (e.g., Campante et al., 2018; Allcott et al., 2020; Levy, 2021; Müller and Schwarz, 2021). The existing evidence has mainly focused on developed countries (a notable exception is Bursztyn et al., 2019 focusing on Russian cities), however, almost every second human lives in rural areas in developing countries without much exposure to media. This is changing rapidly as internet penetration is increasing from 32% (2015) to 58% (2021) in low- and middle-income countries (World Bank, 2021). Whether these rural areas unite behind a broader national identity or divide, is the subject of my second chapter.

As already described, high levels of polarization emerged together with a shift toward cultural issues. To understand the emergence of cultural cleavages at the expense of class-based ones, it is informative to not just analyze the contribution of recently emerging but also disappearing factors. Western societies have experienced rapid declines in former central institutions like the church, labor unions, or the core family (Putnam, 2000; Farber et al., 2021). These serve as central platforms for political discussions and shape social networks (Hertel-Fernandez, 2020). Therefore, they determine who gets in contact and thus exposed to reinforcing or contrasting world views.[1] Moreover, labor unions as well as churches are not neutral spaces for meetings but promote their own ideological positions (see Spenkuch and Tillmann, 2018 for the church's impact in Nazi Germany). The political consequences of their decline are not well understood. My third chapter aims to further our understanding of one of these factors - the labor union. Unions have been portrayed as the main force behind the working class in politics and have played a central role in class conflicts throughout the 20th century (Lipset, 1983). They bring together workers from diverse backgrounds and serve as a link across hierarchies (Frymer and Grumbach, 2021). Workers seem central to our understanding of changing cleavages as it is the group of low-educated workers that shifted away from the left (and therefore more redistribution) towards the right (Gethin et al., 2022). This chapter assesses whether unions not only aggregate but shape ideologies across different levels (workers and managers) at the workplace.

In the following, I will discuss the empirical approach and introduce each chapter separately.

[1] A number of studies have highlighted increased tolerance towards others after exposure in conducive environments (e.g., Mousa, 2020; Lowe, 2021; Corno et al., 2022). Exposure to a diverse set of individuals within a setting where goals are shared is therefore seen as a uniting force.

Empirical Approach: Good decisions by policymakers, as well as individuals, are grounded in understanding the choices we have. Confronted with a decision we ask 'what would happen if?'. The 'credibility revolution' in empirical economics has vastly extended the range of credible research designs to answer these causal questions (Angrist and Pischke, 2010). This thesis relies on a host of quasi-experimental methods including instrumental variables (IVs) and differences-in-differences (DiD) approaches, as well as a regression discontinuity design (RDD) and a difference-in-discontinuity design.

All three chapters introduce new identification strategies. Chapter 1 transfers the logic of shift-share instruments (e.g., as applied by Mayda et al., 2022) to the allocation mechanism behind refugee resettlement. It introduces an instrument that uses the placement of refugee centers in combination with the changing composition of refugee inflows to isolate exogenous variation in refugee inflows in the U.S. Chapter 2 takes advantage of spatial discontinuities in internet availability created by a large Indian government program for rural areas. As a consequence, it proposes an identification strategy that focuses on the periphery in contrast to research designs that rely on first movers (such as Bursztyn et al., 2019; Müller et al., 2022). Chapter 3 exploits the timing of fatal work accidents in the same sector to isolate exogenous variation in unionization in U.S. establishments. Thus, it suggests a new natural experiment within a context that is incompatible with the regression discontinuity designs previously used, as recently demonstrated by Frandsen (2021).

Improvements in empirical research have not solely relied on better research design but also on better data (Einav and Levin, 2014). In this spirit, all chapters rely on large data sets, several of which are novel. Chapter 1 introduces a geocoded individual-level refugee data set covering the universe of refugees between 1975 and 2008. Moreover, it geocodes 3 million U.S. Twitter users interested in politics. Chapter 2 provides village-level data on broadband availability for over 175,000 locations in India. It thus enables for the first time, to the best of my knowledge, the village-level consequences of fast internet in India at a large scale. Furthermore, it aggregates information from over 5 million websites on the distribution of welfare benefits in Jharkhand. This welfare program has been analyzed in many studies at the aggregate level due to its relevance (e.g., Zimmermann, 2021). In contrast, this data set enables individual-level analysis and thereby the assessment of distortions, favoritism, and discrimination within rural communities.[2] Overall, the data sets introduced allow others to answer many more questions than the ones examined in this thesis (see Rosenzweig and Xu, 2023 for an example).

Outline of the Chapters: The first chapter entitled "Immigration, Political Ideologies and the Polarization of American Politics" studies the role of immigration as a driver

[2] Jeong et al. (2023) provide a similar data set for the small state of Uttarakhand.

of polarization in the U.S. between 1992 and 2016.[3] In an era marked by increased global mobility, the phenomenon of immigration has evolved into an integral facet of contemporary societies (Putnam, 2007). In the U.S., the country hosting the largest number of migrants worldwide (Özden et al., 2011), immigration has consistently occupied a central role in the national dialogue (Bonomi et al., 2021). Supporters, as well as opponents of immigration, have increased, resulting in polarized debates (Bonomi et al., 2021). To what extent migrants cause shifts in ideology and contribute to polarization has not been well understood. This chapter adds to the literature on political polarization by focusing on the causal impact of migrants (e.g., Autor et al., 2020; Boxell et al., 2022).

Ideologies significantly change over time, both within and between political parties (Gerring, 2001; PEW, 2014). Mood swings in the electorate can bring different candidates to the top within parties, resulting in ideological shifts of the party. The study of the political economy of migration in the U.S. has mainly focused on vote shares (e.g., Mayda et al., 2022). Vote shares, however, do not capture polarization and ideological shifts within political parties. We add to our understanding of the political influence of immigrants by analyzing their impact on the ideological position of campaign donors and political candidates (e.g., Halla et al., 2017; Edo et al., 2019).

We measure ideology based on 3 million campaign contributions relying on Bonica's (2019) ideology scores which approximate a candidate's ideological position based on donation patterns. What is more, we extend the measure to social media by introducing a new measure based on follower patterns of 3 million Twitter accounts in 2016. To identify causal effects, we follow a recent strand in the literature that has outlined the sufficient conditions for shift-share instruments (Adão et al., 2018; Jaeger et al., 2018; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022). We exploit immigrants' reliance on historic settlement patterns and the changing composition of immigration flows for identification. Our focus on flows mitigates concerns due to the high serial correlation of immigrant stocks highlighted by Jaeger et al. (2018). Pre-trend tests and local randomization tests support the validity of the approach.

The IV results show increases in polarization within two years of the arrival of immigrants. We observe an increase in polarization among donors in tandem with a rightward shift of (right) politicians. The impact declines over time but is still present after 8 years. We highlight a stronger backlash in counties where residents and new immigrants are likely in cultural or economic competition (measured by similarity in education and dissimilarity in origin). Moreover, we find that donors, who are likely colleagues of migrants in jobs with many interactions, are immune to backlash whereas retirees and whites react strongly overall.

The second part of this chapter focuses explicitly on refugees, the migrant group most in

[3]This paper is joint work with Axel Dreher, Sarah Langlotz, and Christopher Parsons. This paper is available as CESifo Working Paper Nr. 8789 (Dreher et al., 2020).

need of a secure location to start a new life. Existing studies consider migrants in general, sometimes including a large share of asylum seekers, but they rarely consider refugees (a notable exception is Dustmann et al., 2020 for Denmark). As refugees forcefully leave their homes, the decision to accept them hinges on a different ethical rationale. It is not well understood whether societies, therefore, react differently to refugees than immigrants in general. In the U.S., the country hosting the largest resettlement refugee program, a lack of data made the explicit study of refugees at a large scale difficult. We alleviate this constraint by introducing individual-level refugee data collected from the Office of Refugee Resettlement and the Bureau of Population, Refugees, and Migration. The data contain over 3 million, the universe of refugees entering the U.S. between 1975 and 2018, which we geocode at the city (and sometimes neighborhood) level.

We develop a new instrument to estimate the causal impact of refugees. In contrast to immigrants, refugees cannot freely choose their initial settlement location in the U.S. Their location is chosen by resettlement agencies or determined by family members if a resettlement center is close by. We exploit this institutional design and construct a new instrument (IV). The IV isolates variation in refugees driven by the distance to the closest refugee center serving co-nationals of a refugee in the past and the changing composition of refugee waves. The IV results echo those for immigrants more broadly. They suggest that the different ethnic rationale behind the acceptance of refugees does not alter the political reaction to foreigners in reality.

The second chapter entitled “Fueling Divisions? The Arrival of Fast Internet in Indian Villages” examines the impact of fast internet on divisions in rural communities in India between 2017 and 2022.[4] In analogy to Chapter 1, it does add to the literature on the polarization of societies (such as Levy, 2021; Boxell et al., 2022). However, in contrast to Chapter 1 which considers actual changes, this chapter studies the consequences of changing information. It assesses fundamental changes in information flows in the largest country in the world, which are representative of trends present in developing countries around the globe. In particular, it examines the consequences of the internet as it connects villages at the periphery to the nation’s core. Does it unite or divide rural communities where people have lived together over decades?

The empirical literature has examined several mechanisms which do not provide a conclusive answer. On one hand, the internet can foster unity by creating national shared experiences and reducing the impact of distance (McLuhan and Powers, 1989; Depetris-Chauvin et al., 2020). On the other hand, it can create polarization through echo chambers, populist messages, and emotional content (Berger and Milkman, 2012; Campante et al., 2018; Levy, 2021). This chapter informs the debate by providing causal estimates of the aggregate impact on villages in a developing country.

[4]This paper is single-authored (Matzat, 2023).

Rural internet penetration has increased drastically in recent years in India. This has exposed villagers to a charged debate driven by the Hindu nationalist BJP mainly targeting the large Muslim minority by spreading inflammatory messages (New York Times, 2019, 2021). To estimate the causal impact of the internet, the endogenous roll-out needs to be accounted for. Internet providers usually connect urban, densely populated, and wealthy places first. Common identification strategies in the literature have used the random spread in the early days of social media or mobile technology for identification (e.g., Bursztyn et al., 2019; Müller et al., 2022; Manacorda et al., 2023). They have thus implicitly focused on first movers that are very different from the millions that have joined the internet in recent years.

This chapter introduces a new identification strategy, which focuses on the uneducated rural population with little prior media exposure. The increase in the internet has been partly driven by a large rural broadband initiative that is connecting every village-council (GP, Gram Panchayat) to the fiber optic network. The broadband initiative is rolled-out in a staggered manner due to capacity constraints and provides public Wifi hotspots. To isolate exogenous variation, this chapter leverages local discontinuities in the internet between areas that receive broadband in the early phase and areas that receive broadband in the late phase.

The data on the location and phase of each broadband connection is based on lists that I collected from Bharat Broadband Network Limited. It represents the first reliable data on internet availability at the village level in India. It is combined with information on the village-council president's allocation of over 300 million workdays between Muslim and other households in Jharkhand within the nation's main social welfare program. The welfare program is supply-constraint such that a large share of eligible households do not receive work. This opens the door to discrimination and favoritism by the village-council leadership responsible for the allocation. The data are publicly available, distributed across over 5 million websites. This chapter combines the information and thus makes it accessible for empirical analysis going beyond the study of single cases. In order to capture other dimensions of divisions, I supplement the data with information on conflict events and voting behavior.

The findings highlight significant divisions in rural communities due to the exposure to fast internet. Non-Muslim village-council presidents allocate fewer work days to Muslims and Muslim village-council presidents allocate more. In sum, local representatives increasingly allocate vital welfare benefits along religious lines as opposed to need. Along with the differential treatment of Muslims, conflicts increase, partly driven by an increase in riots and mobs by BJP supporters. Rural communities with a significant share of Muslims react to these developments with an increase in votes for the secular INC, while non-Muslim villages increase their support for the Hindu nationalist BJP.

The third chapter entitled "Do Unions Shape Political Ideologies at Work?" investigates the political influence of unions.[5] As in Chapter 1, the context of this study is the U.S. where class-based cleavages have given way to culture-based cleavages (Bonomi et al., 2021). It contributes to our understanding of this shift by examining the political consequences of the decline in unions. While Chapters 1 and 2 study the impact of factors that have increased in importance, this chapter examines what we are missing out on.

The long-run decline in unions likely affects the political power balance. The design of fundamental aspects of our welfare systems has been attributed to the influence of unions, like the minimum wage, sick leave, or paid holidays (e.g., Biden, 2021). Unions spend vast resources on mobilizing members to vote, political information campaigns, and lobbying (WSJ, 2012; NILRR, 2021). They have direct connections to millions of members that can constitute a powerful political block if unions manage to influence their political choices. We add to our understanding of interest groups by examining whether unions shape ideologies as opposed to simply aggregating preferences.

We investigate the political influence of unions on workers' political ideologies in unionizing establishments in the United States. While Chapters 1 and 2 both introduce new data sets, this chapter overcomes constraints by developing a new link between two data sets. Previous studies have been unable to assess the political impact of unions at the establishment level due to a lack of matched employer-employee data for political outcomes. We overcome this constraint by linking campaign contributions to establishments holding a union election. We are not aware of any other large-scale data on political behavior with employer information in the U.S. that would allow this link.

The link enables us to study the political impact of unions at the level of the treatment: the establishment level. This comes with several advantages: it enables us to i) compare the treatment group to a reasonable counterfactual and ii) test the plausibility of our identifying assumptions. In particular, we compare changes in the amount and the party composition in campaign contributions between establishments that voted on unionizing in the same year but where some establishments vote pro-unionization while others voted against unionization in a stacked DiD. We demonstrate the suitability of losing establishments as a counterfactual by finding no significant difference in pre-trends, highlighting the robustness of our results for establishments close around the 50% pro-union vote threshold and documenting the absence of trends among losing establishments with differing vote shares. This makes it plausible that trends between losing and winning establishments would have moved in parallel in the absence of unionization, which is what we need to assume to interpret our results causally.

The DiD estimates show a temporary increase in total campaign contributions from workers and a permanent shift of contributions from Republican candidates to Demo-

[5]This paper is joint work with Aiko Schmeißer and available as CESifo Working Paper Nr. 10301 (Matzat and Schmeißer, 2023).

cratic candidates in the six years after the union election. Hence, the findings suggest a short-term political mobilization of workers through a successful union campaign at the workplace. More importantly, they indicate a lasting shift in workers' ideological positions towards the political left. The decline in unionization is thus consistent with a shift of uneducated workers to the right documented by Gethin et al. (2022) due to the diminishing relevance of class-based identification.

Even though unions convince their members, their aggregate impact remains unclear. Unions alter the power between two groups at the bargaining table: workers and managers. The loss of bargaining power for managers can result in a backlash and heightened polarization along class lines. Since we link political outcomes to unionizing establishments we are able to assess the political reaction of the firm's management. We exploit occupational information in the campaign contribution data to estimate the impact of unionization on managers' political ideology in the DiD framework. Thereby in analogy to Chapters 1 and 2, we add to our understanding of drivers of polarization.

The DiD results indicate a leftward shift in campaign contributions not only for workers but also for managers. Unionization moves contributions away from Republican candidates and toward Democratic ones at a slightly faster pace than for workers, without affecting their total spending. These patterns do not align with a rise in tensions between unionized workers and management but instead suggest a convergence of ideological positions. This is consistent with an improvement of the quantity and quality of contact between workers and managers highlighted to be effective in lowering tensions in other areas (Lowe, 2021; Corno et al., 2022) or the persuading power of unions which move even opposing individuals as argued more generally in Coppock (2023).

Summary: The world's central issues necessitate cooperation, unity, and trust, yet many societies, including the U.S. and India, are suffering from significant polarization. This thesis aims to advance our knowledge of three structural changes that affect the political economy of our societies in profound ways: migration, the internet, and unions. Thereby, it contributes to finding solutions by identifying central elements that generate undesirable responses, the first step in a series to design aspirational communities.

This chapter is co-authored with Axel Dreher, Sarah Langlotz and Christopher Parsons

1

Immigration, Political Ideologies and the Polarization of American Politics

1.1 Introduction

Traditionally, polarization refers to the ideological distance among the parties along the political spectrum on specific issues (Sani and Sartori, 1983). In a two-party electoral system, such as the United States, such polarization is "bedevilling…from institutional gridlock…the degradation of checks and balances…the loss of public faith in election administration, political parties and the political establishment more generally" (Carothers and O'Donohue, 2019, p. 66). Ideological polarization increased markedly since the 1970s and accelerated in the 1990s according to a raft of polarization measures.[1] US politics is more polarized today than at any time since the Civil War (Hare and Poole, 2014). A 2018 poll of 588 foreign policy opinion leaders identified political polarization as the greatest single threat facing the United States (Busby, 2020).

In the largest study of its kind, the PEW Research Centre (PEW, 2014) discerned the key compositional shifts in U.S. political ideologies over time, by constructing consistent measures between 1994 and 2014. These include growing minorities holding consistently

[1]These include vote shares (Bond and Fleisher, 2000; Stonecash et al., 2018); measures of party unity voting (Bond and Fleisher, 2000; Stonecash et al., 2018); the voting records of specific interest groups (Stonecash et al., 2018); NOMINATE (D-NOMINATE/DW-NOMINATE) scores based on non-unanimous roll-call votes (Bond and Fleisher, 2000; Fleisher and Bond, 2004; Poole and Rosenthal, 2000, 2001, 2017); campaign contributions (Bonica, 2014) and speech patterns (Gentzkow et al., 2019).

ideological views, Republicans (Democrats) shifting ideologically to the right (left), raised mutual animosity and the rise of ideological silos, wherein individuals surround themselves with like-minded others. Subsequently, 20% (76%) of conservatives (liberals) desired to live in racially and ethnically diverse communities, whereas 57% (17%) of conservatives (liberals) express preferences for residing where most have shared religious faith.

Historically, votes from both sides of the aisle resulted in significant immigration reform in the U.S.[2] The last time this occurred was in support for the 1990 Immigration Act that was signed into law by (Republican) President Bush. Since then, Democrats and Republicans have diverged significantly on issues of migration, culture and race.[3] Asked whether "immigrants strengthen the country because of their hard work and talents" for example, the difference in responses between Democrats and Republicans increased by an *additional* forty percentage points between 1994 and 2017 (PEW, 2017). *A priori*, one might fairly assume therefore that conservatives are on average less amenable to migrants and refugees during our sample period as when compared to liberals—a conjecture supported by recent empirical evidence (Facchini and Steinhardt, 2011; Conconi et al., 2020; Mayda et al., 2022).

Polarization broadly pertains to elites or members of the public. Our focus is on the former, which refers to "high levels of ideological distance between parties and high levels of homogeneity within parties" (Druckman et al., 2013). The relationships between elite and mass polarization remain contentious however. This paper accords with Abramowitz's (2010; 2012) perspective, who highlights the pivotal role played by those members of the public most engaged in politics, since these are recognized as the most highly polarized. Carothers and O'Donohue (2019) argue that it is these individuals that transformed American politics from the bottom up. The available evidence suggests it is these groups of 'most consistent' liberals and conservatives that vote more often (especially in primaries), experienced the greatest increases in polarization between 1994 and 2014, are most likely to contact elected officials, attend campaign events and work for a candidate or volunteer for a political campaign. These so-called 'ideologues' also contribute most frequently to political campaigns (PEW, 2014).

We explore the causal role of migration in fostering the political ideologies of candidates running for the House of Representatives between 1992 and 2016. Our focus is the United States, which hosts the largest number of migrants globally (Özden et al., 2011), in tandem with suffering some of the highest levels of political polarization (Dimock and

[2]Notable instances of bipartisanship both for and against immigration to the U.S. include: the anti-China platforms that both parties adopted in the 1876 and 1880 Presidential elections, which ultimately culminated in the Scott Act of 1888; the eugenicist findings of the Dillingham Commission in 1911 that argued in favor of the racial inferiority of Southern and Eastern Europeans that led to the passing of the Emergency Quota Act (1921), the Johnson-Reed Act (1924) and the Hart-Cellar Act (1965) that abolished national quotas (Tichenor, 2009).

[3]Pat Buchanan's speech on the existence of a culture war for the soul of America is often cited as a relevant turning point in this regard (Fiorina and Abrams, 2008).

Wike, 2020). Using data derived from 3 million campaign contributions, we capture the ideology of campaign donors and their political recipients, in turn calculating a raft of polarization measures. We also provide a complementary analysis for the sake of external validity when implementing Twitter data. We subsequently identify causal effects, employing the familiar shift-share instrumental variable in conjunction with fixed effects for counties and years, such that our identifying variation is within counties over time.

The literatures examining the political economy of migrants on aggregate (Mayda, 2006; Otto and Steinhardt, 2014; Barone et al., 2016; Nikolka and Poutvaara, 2016; Halla et al., 2017; Edo et al., 2019; Lonsky, 2020; Mayda et al., 2022), and refugees specifically (Campo et al., 2020; Dustmann et al., 2020; Steinmayr, 2021) focus on vote shares accruing to (predominantly) far-right parties. As such, they are unable to capture ideological shifts within political parties. Ideologies however significantly change over time, both within and between between political parties (Gerring, 2001; PEW, 2014). In the 1984 Presidential election, for example, Reagan won 59 percent of the popular vote, while Trump won only 46 percent in 2016. An analysis of the Republican vote share alone might therefore imply that the United States shifted politically to the left, whereas the reason why statements based on these vote shares contradict our observations is because Reagan and Trump did not have the same ideological positions simply because they belonged to the same party. Considering the differences in ideology between Reagan and Trump therefore, as well as those of their opponents (Mondale and H. Clinton respectively), would no doubt provide a more nuanced and comprehensive understanding of the shifting ideological sands over time. This is what we do in this paper.

Our work is related to Autor et al. (2020) who exploit local trade exposure from China to provide causal estimates of the effects of imports on American political polarization between 2002 and 2016. We rather examine the role of migration in fostering polarization. According to Bonomi et al. (2021), respondents to a repeated survey by the Pew Research Center mention “race and immigration”—as opposed to trade—as one of the three most important problems facing the United States with the highest frequency in the 2013-2018 period. Migration is therefore expected to affect political polarization more than trade, a proposition we examine conditional on local trade exposure from China.

We also contribute to the literature on interest group politics and political polarization (Cho and Gimpel, 2010; Facchini et al., 2011; Barber, 2016a; Gimpel and Glenn, 2019). Glaeser et al. (2005) argue that candidates holding extreme positions on wedge issues, like immigration, foster both donations and core supporter turnout, ultimately proving politically polarizing. Migration therefore constitutes one candidate to explain the geographical clustering of political contributions (Hopkins, 2017), what has otherwise been termed Partisan sorting (Mason, 2015). Indeed, a large literature in social psychology examines how contact with out-groups of various characteristics affect in-groups (Pettigrew and Tropp, 2006), insights that lend themselves to providing natural heuristics when interpreting the heterogeneity of our results. Finally, we contribute to the literature that examines the determinants of campaign financing (Brown et al., 1980; Mutz, 1995; Gimpel et al., 2006), in our case exploring the role of migration.

Ultimately, we study the ideologies of the universe of candidates running for the House of Representatives as opposed to only those elected to office. Capturing shifts in the prevailing zeitgeist, we leverage "Data on Ideology, Money in Politics, and Elections" (DIME) provided by Bonica (2019) for the 1979-2016 period. The data exploit patterns in campaign contributions to determine candidates' ideologies. Campaign contributions are premised to be driven by ideologies, such that on average contributors give to ideologically more proximate candidates.[4] Based on contribution patterns (i.e., who gives how much to whom) Bonica estimates ideal points for candidates and contributors. The resulting so-called common-space CFscores "represent the most comprehensive ideological mapping of American political elites to date" (Bonica, 2016). We derive a number of polarization measures from these data. Focusing on campaign donors, we measure polarization of campaign finances as donations to extreme candidates relative to moderate candidates. Focusing on candidates, we consider the ideology of election winners, overall, and for Republican and Democratic winners separately. We further measure the ideological distance of election winners relative to losers and the probabilities that moderate or extreme candidates win elections. To test the mechanisms at play, we exploit the differences between residents' characteristics and those of incoming migrants, specifically cultural, educational, occupational and racial disparities.

We identify causal effects using a shift-share instrument, guided by recent advances in the accompanying econometrics literature (Christian and Barrett, 2017; Adão et al., 2018; Borusyak et al., 2022; Jaeger et al., 2018; Goldsmith-Pinkham et al., 2020; Mayda et al., 2022). We predict the number of immigrants in a county and year with an interacted instrumental variable comprising two parts. One element serves to *shift* the number of immigrants from year to year. This is calculated as the change in the number of aggregate immigrants from a particular origin to the United States over an electoral cycle. The second element constitutes the pre-sample *share* of migrants in local labor markets, calculated as the share of foreign-born adults from each country of origin in that country's adult population living in U.S. counties in 1980.[5] Our shift-share instrument is then the product of the shift- and share-components summed over all countries of origin.

We examine how changes in foreign populations differentially affect counties with varying initial shares of immigrants in 1980. Network-effects ensure counties with larger his-

[4]Findings in McCarty and Rothenberg (1996) and Ensley (2009) support this assumption, for example. We exclude arguably more strategic contributors like Political Action Committees (PACs).

[5]1980 constitutes our base year, since it is the first period we observe before our sample period begins. Indeed, the Immigration Act passed in 1990 significantly increased the overall numbers of immigrants permitted to enter the U.S., concurrently introducing family-based immigration, distinct employment visas as well as the diversity lottery.

torical immigrant shares from certain origins are characterized by larger future shares of incoming immigrants from those origins. Counties with higher initial immigration shares are assumed not to be differentially affected by country-wide changes in immigration as when compared to counties with lower initial shares, other than through the impact of contemporaneous immigration, while controlling for county- and year-fixed effects and a battery of controls. This assumption is tested in considerable detail.

Migration on aggregate increases polarization within two years of arrival, inducing political shifts to the ideological right. Campaign contributions to extreme candidates increase relative to those for moderates. Election winners become more conservative when they are Republican. Conservative Republicans are more likely to win elections. Liberal Democrats less so. Our results are similar when we focus on inflows over eight, as opposed to two year time horizons, although they become smaller in magnitude. They become starker as cultural distances between natives and migrants increase or when education levels are similar, one interpretation of which is that natives resent foreigners from different cultural backgrounds and fear competition, while welcoming immigrants with complementary labor market skills. Unpacking our results from the perspective of campaign donors, we demonstrate that our results are driven by the non-working and retirees, those employed in occupations with high proportions, and yet little contact with immigrants, and predominantly whites. These results are robust to an array of alternative econometric specifications and falsification exercises and when we instead rely on an alternative measure of elite polarization, one based on Twitter accounts.

In our final analysis, we examine the specific role of refugees in catalyzing ideological polarization (as opposed to migrants on aggregate). This distinction is likely important. Although traditionally constituting only around one tenth of total immigration, refugees receive disproportionate (both positive and negative) media attention, since refugees constitute *"the most visible, challenging, and morally significant of newcomers"* (Haines, 2012). In part, this is because refugees often represent new populations through the extensive margin along specific migrant corridors (Bahar et al., 2022). Refugees and other migrants represent fundamentally disparate groups, primarily distinguished by their primary motivation for emigrating (forced vs. unforced), their socioeconomic characteristics and their ethnic backgrounds (Chin and Cortes, 2015), in concert with the limited agency refugees have with regards their initial resettlements in the United States.

Whereas immigrants more broadly are free to settle where they choose, refugees, as explained by Bruno (2017), are resettled within 50 or 100 miles—and within the same *state*—as their local 'affiliate', the institution responsible for providing local refugee services.[6] These thresholds do not lend themselves naturally to a regression discontinuity design given the paucity of observations around the relevant cut-offs. Instead we divide

[6]This distance depends upon whether refugees have U.S. ties with friends or family, see Mayda et al. (2022).

the U.S. into 52,341 0.15°*0.15° grid cells and exploit novel data on the precise location and timings of the opening of 313 refugee centers across the United States. This approach yields two sets of instrumental variables, both of which predict the number of refugees at the grid cell level before aggregating to the county level. The first set predicts the number of refugees located in each grid cell based on their distance to the nearest refugee resettlement center, allowing for different coefficients in each year and controlling for cells' distances to the nearest Amtrak station, airport and city with a population over 100,000 (see Figure A.7 in the Appendix). The second set of instruments represents a melding of our two empirical approaches, one in which the predicted numbers of refugees are implemented as our initial 'shares' of the shift-share approach, whilst considering a number of additional aspects of the refugee allocation process, such as accounting for centers specializing in the resettlement of refugees from specific origins and when individual centers began placing them. Exploring the characteristics of campaign contributors, our results focusing on refugees echo our results for immigrants more broadly.

The following section introduces our data. Section 1.3 explains how we estimate the causal effects of immigration on ideological polarization. We discuss our results and their robustness in Section 1.4. The final section concludes.

1.2 Data

1.2.1 Immigrants

County-level immigrant stock data are available in 1980, 1990 and 2000 from the U.S. Census, and biannually from the American Community Survey (ACS) for the years 2006-2016 from IPUMS-USA (Ruggles et al., 2020).[7] The U.S. census and ACS report data on the total foreign-born population, which refers to anyone born outside of the U.S., including U.S. citizens born abroad, shorter term migrants (including foreign-born students), humanitarian migrants (including refugees) and some fraction of the illegal migrant population not otherwise captured (Hanson, 2006). Origin-specific stocks of immigrants in 1980 capture our initial 'shares', while differences in migrant stocks over two-year periods are employed as instrument 'shifters'. Throughout, the term 'migrants' refers to the aggregate foreign-born population. We present results when specifically analyzing the sub-set of 'refugees' in Section 1.4.5.

The number of migrants in the United States increased by 957,554 on average per year between 1990 and 2016. The share of net immigrants relative to the native adult population peaked in the early 2000s (at around 0.06), while turning negative in more recent years (see Figure A.1 in the Appendix). Figure A.2 shows the net increase in the

[7] We use linear interpolation to obtain estimates for the years 1992, 1994, 1996, 1998, 2002, 2004.

number of immigrants over the years of our sample at the county-level, relative to the adult population in the year 1992, with darker shades indicating greater increases.

Our data also detail immigrants' origins and education levels, which we use to derive proxies for cultural and educational distances relative to local native populations. By 2016, some 38 percent of migrants originated from elsewhere in the West, 36 percent from Latin America, 7 percent from Africa and 20 percent from Asia. 32 (21) percent of all immigrants dropped out of (graduated from) high-school, 14 percent spent only 'some' time in college, and 6 percent graduated from college, while 27 percent have more than college education.

1.2.2 Refugees

Our individual-level refugee data derive from two distinct entities of the State Department—the Office of Refugee Resettlement (ORR) and the Bureau of Population, Refugees, and Migration (PRM). The ORR data span the 1975-2008 period and comprise 2.6 million individuals from 136 countries of origin. They are geographically remunerated at the U.S. state, county and city levels. The PRM data comprise 0.6 million individuals from across 99 origin countries between 2009 and 2018.

We geo-code the refugee locations using: Open Street Maps API, Google Maps API, the data science toolkit and manual reviews; relying upon data at the city, county and state levels.[8] To ensure a high degree of accuracy we also reverse geo-code locations, to facilitate comparing the resulting names. Additionally, we manually cross-check a small number of locations receiving at least 10 refugees, in cases in which our county information derived from the raw data conflicted with the county of assigned location. Ultimately, we successfully assign 96.50% of refugees to about 15,200 locations (99.89% at the city level, the remaining at the county level). These locations are then matched to the county-level, 3,141 in total. We provide these data at https://www.refugeeresettlementdata.com.

Relative to immigrants on aggregate, the share of refugees is substantially lower, decreasing from around 0.0012 in 1990 to 0.0006 in 2018. The dilution of these relatively small numbers of refugees across both time and space results in less than ideal identifying variation, in part thereby explaining the conspicuous absence of papers examining a raft of refugee outcomes in the context of the United States. Figure A.3 illustrates the number of refugees arriving in the United States. Figure A.4 plots the same data highlighting refugees' geo-coded locations. The distribution of refugees is comparable to that of migrants more generally since both migrants and refugees are ultimately attracted to larger, multicultural, urban and often coastal locales.

[8] http://www.datasciencetoolkit.org.

1.2.3 Refugee Processing Centers

We obtained data detailing the universe of existing refugee processing centres from the Worldwide Refugee Admissions Processing System (WRAPS) website, previously maintained by the PRM,[9] shortly before they were removed from the public domain during the Tillerson administration. The information provides details of the location of 313 individual refugee resettlement centers run by one of several Voluntary Agencies (Volags) across all U.S. states with the exception of Wyoming.[10] Volags have constituted the backbone of refugee resettlement in the United States from at least 1945, when President Truman passed a directive granting 'Welfare Organisations' the power to sponsor refugees.[11]

The information provides details of the name, address, contact details and voluntary agency to which each processing center is affiliated. Under the Trump administration, dramatic changes were made to the levels and composition of funding to the State Department. In turn, the refugee admission ceiling was reduced from 110,000 in the last year of the Obama administration to 45,000 and ultimately slashing that number to 15,000. As such, significant resources had to be dedicated to confirm the continued existence of each affiliate and if not in the affirmative when they closed, if they changed address and/or if any specific center changed their affiliation; as well as to confirm when each center first opened. Once these details were confirmed, each center was assigned a precise geo-location, as shown in Figure A.4 and as explained above.

1.2.4 Political Ideology and Polarization

We construct several measures capturing political ideologies and polarization from Bonica's (2019) Database on "Ideology, Money in Politics, and Elections" (DIME).[12] These data predominantly leverage campaign contributions registered with the Federal Election Commission (FEC) and state reporting agencies. The data comprise contributors' detailed location, and, for sub-sets of the data, their employment status together with their names and occupations, from which we can subsequently infer their likely origin and degree of contact with immigrants within their workplace. On the receiving end, the data contain information on all candidates running for elected office in the United States

[9] Last downloaded 06/10/2017 from: `http://www.wrapsnet.org/consolidated-placement-plan.`

[10] While it has been commonly reported that Wyoming has never *resettled refugees*, indeed Wyoming did resettle some Vietnamese boat people in 1975. Rather Wyomong never adopted a refugee resettlement program as were ushered in to all other states following the passing of the 1980 Refugee Act. Only very recently have local Wyoming churches taken in Afghan refugees following the refusal of the state Governor to act in this regard (see `https://www.churchtimes.co.uk/articles/2021/22-october/news/world/wyoming-churches-take-in-afghan-refugees-after-state-governor-refuses)`.

[11] This preceded the signing of the Displaced Persons Act of 1948, which acknowledged refugees as a special class of migrant for the first time, together with its extension in 1950, which paved the way for hundreds of thousands of displaced Europeans to subsequently enter the United States.

[12] A number of recent papers implement these data (e.g., Bonica, 2013; Thomsen, 2014; Nyhan and Montgomery, 2015; Barber, 2016a; de Benedictis-Kessner and Warshaw, 2016; Hollibaugh Jr and Rothenberg, 2018; Martin and Peskowitz, 2018; Autor et al., 2020).

that receive such contributions, which arguably holds true for all 'serious' candidates. Bonica (2019) calculates contributors' and candidates' ideologies based on whom they contribute to and from whom they receive contributions, respectively, accounting for factors affecting all contributions across the board, like charisma. The pivotal assumption undergirding these data, and subsequently our analysis therefore, is that contributors donate larger amounts to those candidates they are more ideologically aligned with.[13] Compared to other available data detailing the ideological positions of politicians, those based on roll call votes for example,[14] this approach rather analyzes the entire universe of candidates, including those that failed to win at the ballot. We therefore analyze any and all polarization arising between candidates from the same party, as well as between winning candidates and runners-up from opposing parties. Adopting this methodological approach allows us to capture significant ideological movements *within* parties, even should they fail to win an election. Conversely, omitting losing candidates' ideologies would be akin to treating the 2020 election with Democratic Presidential candidate, Joe Biden—when running against President Trump—as identical to self-styled socialist Bernie Sanders, who would have otherwise run against Trump in Biden's absence.

Bonica (2016) calculates a Campaign Finance (CF) score to measure political ideology, based on campaign contributions.[15] He assumes contributors donate based according to their ideal points, the candidate's ideal point, the utility they derive from donating and the marginal costs involved. The CFscore method applies correspondence analysis, a method similar to principal components analysis that focuses on relative, as opposed to absolute, differences in ideologies between donors and recipients. Bonica calculates ideal points along a single dimension, a typical left-to-right political scale.

This ideological scale is anchored to federal elections. State-level ideological scores are subsequently linked using data on campaign contributors that donate to both federal and state elections. On average between 70 and 90 percent of contributors in any given state contribute to both federal and state election campaigns (Bonica, 2014). These observations therefore serve to 'bridge' and in turn harmonize ideological scores across institutions and political hierarchies.[16] What results is a consistent ideological scale across contributors and candidates, institutions and time periods.

[13] A number of articles validate this assumption (e.g., Ensley 2009; Barber et al. 2017).

[14] E.g., DW-NOMINATE (Poole and Rosenthal, 1985).

[15] Our description of the CFscores draws from Bonica (2014), see in particular his Supplementary Materials.

[16] As Bonica (2014) explains, he first applies correspondence analysis to federal election data. He then scales the resulting data according to the federal-level ideal points that emerge for each individual state. This exercise is based on data of contributions from donors to both state and federal campaigns. This facilitates anchoring state-level scales, such that the resulting state-level CFscores are all based on the same ideological scale as the federal CFscores. Technically, the correspondence analysis applied by the CFscore method scales two-way frequency tables by decomposing a transformed matrix of χ^2 distances (Bonica, 2014). As Bonica (2014) explains, this is almost equivalent to a log-linear ideal-point model, but comes at a much-reduced computational cost.

The *dynamic* DIME scores that we rely upon in this paper are calculated for each time period separately. This allows for idiosyncratic changes in specific candidate ideology over time. We observe few stark movements in CFscores however. Legislator ideal points—as captured by roll call votes, e.g., DW-NOMINATE—are similarly stable over time (Bonica, 2016). Both measures are highly correlated, lending additional plausibility that the CF scores can be interpreted along a liberal-conservative ideological scale. Ideal points, calculated for candidates *prior* to entering office, are typically highly correlated with both candidates' future CFscores as incumbents, as well as their subsequent voting behavior. Bonica (2018) demonstrates DIME scores to accurately predict policy preferences, based on 30 policy items included in the 2012 Cooperative Congressional Election Study (CCES). Candidates' ideal points are also highly correlated with the ideal points of contributions to the political campaigns of *others* (Bonica, 2016), meaning they seemingly represent genuine expressions of ideological preferences.

We analyze all general elections to the House of Representatives between 1992 and 2016. Our main analysis employs biannual changes in migrant stocks. Our focus on the House of Representatives (as opposed to Presidential or Senate elections), is a choice governed by the salience of the topic since for example *"political polarization ... seems to jeopardize Congress's constitutional responsibility for regulatory oversight"* (Farina, 2015), in addition to the resulting identifying variation, which underpins our empirical analysis. During our sample period, our data comprise ideology estimates for 1,089 candidates and 3.7 million contributions, deriving from 186,209 contributors (173,746 individuals, as opposed to corporate donors).

Left-aligned donors include university and college employees, those working in Hollywood and book publishers, as well as the online computer-services industry (Bonica, 2016). Right-aligned donors include those in the oil, gas and coal industries, agriculture, mining and construction. During our sample period, among the top three conservative donors are the *Club for Growth* and the *American Future Fund*. Both support a 'conservative and free-market viewpoint'. Among the three largest liberal donors are *For our Future* and *End Citizen United*, which are 'committed to serving progressive values and causes' and to limit campaign contributions, respectively. Large donors located in the middle of the ideological distribution include the *American Federation of State County & Municipal Employees*, the *Democratic Congressional Campaign Committee* and the *NEA Fund for Children and Public Education*.

We derive a number of polarization and ideology measures from these data. Focusing on (general election) contributions from donors in a specific county—those donated to candidates running for the House of Representatives in any electoral district—we define CF scores for liberal, moderate and conservative donations in that county, based on contributions in 1990.[17] We subsequently rank candidates according to their ideology on a

[17]In essence following Autor et al. (2020).

left-right scale, binning candidates into terciles. Contributions in the right tail of the scale are termed 'conservative'. In analogy, we define 'liberal' contributions as those located in the left tail. Those remaining in the centermost tercile are deemed 'moderates'. Contributions to moderate candidates, according to this nomenclature, substantially declined over time, at the expense of liberal and in particular conservative candidates (please see Figure 1.1).[18]

Figure 1.1 − Share of Contributions to the House of Representatives

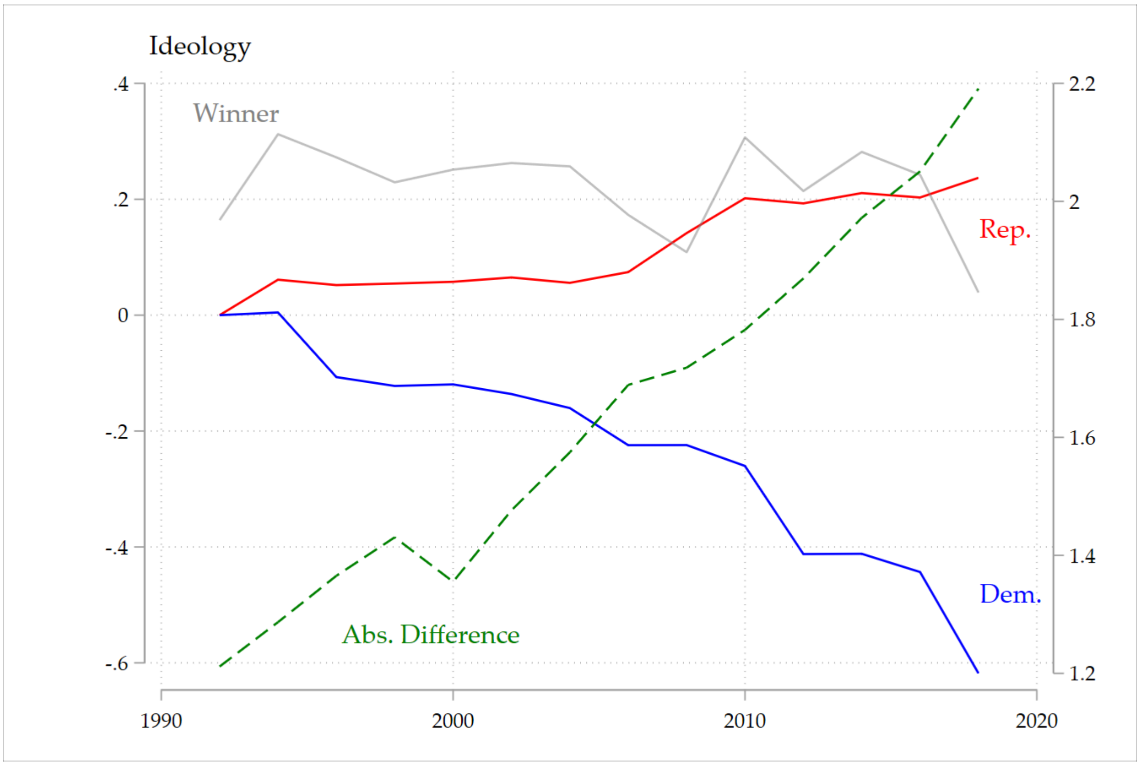
Notes: We rank candidates according to their ideology on the left-right scale and divide the amounts of contributions these candidates received in terciles. For the year 1990, we define the third of the contributions most to the right end of the scale as "conservative" contributions. In analogy, we define "liberal" contributions as those on the left end of the scale and the remaining tercile as "moderates." We then use the resulting cut-offs for ideology scores to categorize amounts of contributions into these three categories of CFscores in each year in our sample.

Our *Extreme vs. moderate* measure of polarization is calculated as the difference in contributions donated to the sum of liberal and conservative (*extreme*) candidates relative to those given to moderate candidates. *Winner* focuses on the ideologies of general election winners. Candidates' ideology scores are assigned to the county-district cell of their victory. We then take the population-weighted average across all county-district cells within a county. Using population weights, we finally harmonize county borders over time to those of 2010.

[18]Group-shares are not exactly equal in 1990 given that candidates at tercile cut-offs do not receive equal amounts.

We proceed by investigating the ideologies of winners conditional on them being Republicans (*Winner if Rep.*) or Democrats (*Winner if Dem.*), which facilitates testing for shifts in ideology within parties. *Winner vs. loser* is calculated as the absolute distance between winning candidates and the runners-up. Once again we calculate these at the county-district cell level and aggregated them up to 2010 county boundaries. Digging deeper, we separately analyze the probabilities that *Conservative Republicans*, *Moderate Republicans*, *Moderate Democrats* or *Liberal Democrats* win at the ballot. We define moderate politicians as centrists within their party, based on their ideology score compared to the party median in 1990; with the remainder constituting conservative and liberal politicians.

Figure 1.2 – Ideology and Polarization



Notes: We depict the ideology of the winners on average (gray line) and by party (red and blue line). Note that we subtract the 1992 party mean of the ideology of the winners by party. The green line depicts the absolute distance between the winner and the runner up. Solid lines refer to the left axis, the dashed line refers to the right axis (both axes represent the ideology score).

Figure 1.2 shows that ideological polarization increased over the years of our sample. While the ideology of winners (left axis) exhibits no clear trend, the absolute difference between winners and runners-up increases over time (right axis). Republican winners move to the right, while Democrat winners move to the left (depicted on the left scale).[19] Specific candidate ideologies, though estimated for each period separately, do not vary substantially over time. The changes that we observe in the data therefore result from

[19]We normalize ideology scores of winning Democrats and Republicans to zero in 1992.

candidates of differing ideologies receiving contributions of varying amounts at different junctures.

We draw on individual Twitter accounts in a supplementary analysis. Updated raw data were obtained from Barberá (2015), in which ideological scores for more than 300,000 users are calculated using a Bayesian Spatial Following model. Barberá (2015) assumes that Twitter users are more likely to follow politicians with shared ideologies. The pre-defined accounts include 318 political accounts of politicians, journalists and political parties, from which 33 million followers are subsequently identified.[20] Ideology scores are subsequently derived from individuals' follower patterns, assuming the existence of a single latent dimension of ideology. The resulting measure is highly correlated with other more established measures (see Barberá 2015). To ensure sufficient numbers of observations per location, we focus on the year 2016. We assign users to counties, based upon the "location" field in their profile, resulting in some 3 million users.[21] We again divide ideological scores into terciles, which we refer to as left, right and moderate users. Our dependent variables detailed at the county level are the shares of extreme users (left or right), left users, right users and moderate users in all Twitter users. Figure A.5 presents these data.[22]

1.3 Methods

1.3.1 Migrant Analysis

The endogenous location decision of migrants likely results in them favoring areas that imbue them with particular advantages, such as better employment prospects. Reverse causality constitutes an additional concern, since newcomers likely choose areas where they are more likely welcomed, as opposed to feared. So too might differential trends exist for treated areas (those that receive immigrants above a particular threshold) and non-treated areas (those that do not). Simply comparing outcomes of locations without recognizing these threats to identification could therefore yield biased estimates.

Our main specification is:

$$\mathrm{Y}_{ce} = \beta \Delta MS_{ce} + \mu_c + \lambda_e + \mathbf{x}'_{ze}\boldsymbol{\gamma} + \epsilon_{cze}, \tag{1.1}$$

[20]Our sample of Twitter users are clearly unrepresentative of the American population as a whole. They have above average education and take greater interests in politics. Whereas the results from this exercise may be deemed to better accord with a definition of mass polarization, the selected characteristics of the resulting sample are argued to be an additional informative source of elite ideology (Barberá, 2015).

[21]Location data are missing or too imprecise for approximately 60 percent of the users in the sample we could retrieve. Figure A.6 shows that the ideology of users with such information has similar distribution in the tails than those without, albeit with lower densities than moderate users.

[22]We report descriptive statistics for all variables in Table A.1 in the Appendix.

where Y_{ce} reflects our measures of political ideology and polarization introduced in Section 1.2, in a county c in election-year e. ΔMS_{ce} is the net change in the number of immigrants relative to (the stock of) a county's adult population. μ_c are county-fixed effects and λ_e are year-fixed effects, which absorb a variety of potential shocks affecting all counties in particular election years. Note that the fixed effects-specification implies that we expect polarization to react to changes in inflows rather than changes in stocks of migrants. This is because we expect populations to become used to levels of migrant inflows, even if these inflows are high, but to react strongest to changes in the flow. In other words, we expect ideology to change temporarily rather than permanently as a consequence of migrant inflows.[23]

In keeping with Mayda et al. (2022), we include a vector of control variables $\mathbf{x}_{ze}$ (all in differences) at the commuting zone level z. These include the shares of low-skilled natives, males, those married, African-Americans and urban residents, in addition to the unemployment rate, the labor market participation rate and the average income per person in the citizen population together with an index proxying import competition exposure to China as defined in Autor et al. (2016).[24] We include a Bartik share control that captures sector-specific local labor market shocks (calculated by Mayda et al. (2022, 365) as the "weighted average of the industry-specific employment in year t, using as weights the employment shares across industries of the commuting zone in 1990"). The error term is ϵ_{cze}. We cluster standard-errors at the state-level and implement population weights in all regressions.

We employ the familiar shift-share instrument to address the endogeneity of immigrant shares in a county's population. In doing so, we closely follow recent work of Mayda et al. (2022).[25] We employ an interacted instrumental variable to predict the change in the number of immigrants in a county and year. We define the number of adults born in the United States that live in county c in the year 1980, as a share of total U.S.-born adults, as $sh_{US,c,80} = \frac{N_{c,80}}{\sum_c N_{c,80}}$.[26] Analogously, we define $sh_{i,c,80} = \frac{M_{i,c,80}}{\sum_c M_{i,c,80}}$ as the share of adults born in country i in that country's adult population living in county c in the year 1980. The number of natives N in county c in year e is then calculated as the product of the county's 1980 population share and the total native adult population in e, $\hat{N}_{ce} = sh_{US,c,80} N_e$. The predicted number of total immigrants residing in a county is $\hat{M}_{ce} = \sum_i sh_{i,c,80} M_{ie}$, the product of the 1980-share of immigrants from a country

[23]When we control for (predicted) stocks of migrants, our results are unchanged. Results for stocks are qualitatively similar, but statistically weaker).

[24]Our source for these data is Mayda et al. (2022), who take them from the U.S. census and the ACS.

[25]See Adão et al. (2018); Borusyak et al. (2022); Jaeger et al. (2018); Goldsmith-Pinkham et al. (2020) for recent contributions. The assumptions discussed in these papers allow us to derive unbiased estimates under assumptions that are, to some extent, weaker than those introduced below. The cost of doing so is in assuming one of the two variables comprising the interacted instrument is exogenous. We return to this point below.

[26]This is in line with Mayda et al. (2022). We define adults as people over the age of 17.

living in a county in the U.S.-total and the number of immigrants from that country to the United States in e, summed over all countries of origin. Our instrument for the change in the number of immigrants as a share of the adult population is then the change in the predicted share of immigrants in the predicted adult population of a county, $\Delta\widehat{M}_{ce}/(\widehat{M}_{ce}+\widehat{N}_{ce})$.

Our empirical set-up therefore examines how changes in foreign populations over time differentially affect counties with varying shares of immigrants in 1980. Due to network-effects, one would assume that counties with larger historical shares of immigrants from a particular country of origin should receive larger proportions of migrants from the same country of origin in any given year. Simplifying somewhat, the exclusion restriction is that counties with higher shares of immigrants in 1980 are not differentially affected by country-wide changes in immigration, as when compared to counties with low initial shares, other than through the impact of contemporaneous immigration, when controlling for county- and year-fixed effects, in addition to our battery of controls. Controlling for county- and year-fixed effects—which capture the levels of the variables that comprise our instrumental variable—initial immigrant shares and country-wide immigration cannot be correlated with the error term and are thus indeed (conditionally) exogenous. We visualize and discuss whether and to what extent counties with higher or lower shares of initial immigration adhere to differing trends in terms of polarization below. We further examine other potential threats to identification as discussed recently by Christian and Barrett (2017); Adão et al. (2018); Borusyak et al. (2022); Jaeger et al. (2018) and Goldsmith-Pinkham et al. (2020). To this end, we conduct Monte Carlo randomization to test for spurious long-run trends, while accounting for potential adjustment dynamics occurring in those years following earlier migrant inflows.

Putting these elements together, we estimate the following first-stage regression:

$$\Delta MS_{ce} = \delta\frac{\Delta\widehat{M}_{ce}}{(\widehat{M}_{ce}+\widehat{N}_{ce})} + \omega_c + \phi_e + \mathbf{x}'_{ze}\boldsymbol{\zeta} + \nu_{cze}, \tag{1.2}$$

where $\mathbf{x}_{ze}$ are the controls from the main equation, ω_c are county-fixed effects, and ϕ_e are year-fixed effects. We then estimate equations 1.1 and 1.2 using Two-Stage Least Squares (2SLS).

Social psychologists have long examined how out-groups (in our context immigrants and refugees) affect in-groups (natives), although theory is conflicting. Knowing members of out-groups personally likely breeds familiarity and empathy, as argued by proponents of contact theory (Allport, 1954). Living in close proximity however, might also result in natives feeling out-competed or threatened, thereby fostering prejudice as proffered by advocates of group threat theory (Sherif et al., 1961; Campbell, 1965).

Motivated by these long-standing hypotheses, we first exploit the richness of our DIME donations data, specifically in terms of splitting our sample along a number of dimensions:

donor's names (from which we can infer origins), occupations (from which we can proxy the degree of contact with immigrants), and donors' employment statuses. Employment statuses are categorized as working, non-working, non-working (student) and non-working (retired). For those in employment, we first harmonized the donor occupations to the Standardized Occupation Codes (SOC), through the application of a number of matching tools (e.g., SOCcer) in addition to significant manual matching over an extended period. Once standardized to the SOC, we further disaggregated these occupations according to the skill components of those jobs, using O*NET[27] so as to identify 'high migrant contact' occupations. In turn, we deemed 'high migrant contact' occupations to lie above the top 70th percentile of a 'contact-score' that we calculated based on three constituent factors from O*NET.[28] We define 'high' immigrant share occupations as being equal to or above the 90th percentile in the 1990 U.S. census. Donors' origins were inferred through an examination of donors' surnames in combination with the relevant census information pertaining to the shares of those of differing ethnic backgrounds and their share of surnames in the U.S. census.[29] This information is available for all surnames appearing at least 100 times in the census.

Finally, we examine the potential roles of cultural and educational distances in mediating the effect of immigration on ideology and polarization with the following regression:

$$\mathrm{Y}_{ce} = \beta \Delta MS_{ce} \times DIST_{ce} + \alpha DIST_{ce} + \delta \Delta MS_{ce} + \mu_c + \lambda_e + \mathbf{x}'_{ze}\boldsymbol{\gamma} + \epsilon_{cze}, \tag{1.3}$$

where $DIST_{ce}$ is either cultural or educational distance. Cultural distances are based on distinguishing immigrant shares aggregated over individual origins, namely: Western, Latin American, African and Asian countries, all of which are available at the commuting zone level. County-level shares are proxied by multiplying commuting zone level shares with the overall increase in the county-level flow of immigrants. We then calculate similar measures for the resident population. Shares of Whites, Blacks, Asians and Hispanics in a county's resident population are obtained from the Census Bureau. The absolute differences in the shares of each group comprising our net immigrant flows, as well as the respective shares in resident populations are subsequently computed. The sum of these shares—which we normalize to one—is our proxy for cultural distance, based on the assumption that similarities in geographic origins correlate with these distances. We adhere to the same procedure to proxy educational differences, but rather rely on the shares of

[27]O*NET is the Occupational Information Network, see https://www.onetonline.org/.

[28]These are: 1) Work Activity: Developing and Building Teams—Encouraging and building mutual trust, respect, and cooperation among team members, 2) Work Values: Relationships—Occupations that satisfy this work value allow employees to provide service to others and work with co-workers in a friendly non-competitive environment. Corresponding needs are Co-workers, Moral Values and Social Service and 3) Work Context: Interpersonal Relationships: Work With Work Group or Team—How important is it to work with others in a group or team in this job?

[29]Examples include: Anderson= 75% White, Jackson = 53% Black, Garcia = 92% Hispanic, Nguyen = 97% Asian.

immigrant and native populations with differing levels of education, as introduced in Section 1.2.[30]

1.3.2 Refugee Analysis

The placement of refugees into one of 313 refugee centers lends itself to an alternative identification strategy. First we divide the U.S. into 52,341 equally sized grid cells of 0.15°*0.15°, which at the equator corresponds to approximately 16.7 km^2. Next we predict the number of refugees located within each grid cell based on their distance to the nearest refugee resettlement center, allowing for annual variations in coefficients (see Figure A.7 for a graphical depiction of refugees by grid cell level, together with the locations of the relevant refugee resettlement centers). Aggregating the predicted number of refugees to the county level, we employ the predicted number of (new) refugees in each year as an instrumental variable. Given that refugees are more likely to settle nearer refugee centers than further away, we expect the instrument to have power. To the extent that the location of refugee centers reflect distances to other locations that might be correlated with polarization through channels other than refugee inflows, this instrument would likely violate the exclusion restriction however. To militate against this possibility, we control for a cell's distance to the nearest Amtrak station, the nearest airport, and the nearest city with a population over 100,000. The zero-stage regression is the following:

$$\begin{aligned} \mathrm{R}_{gt} = \beta_1 dist_{gt} + \beta_2 (dist_{gt} * \lambda_t) + \beta_3 distAmtrak_g + \beta_4 distAirport_g \\ + \beta_5 distCity_g + \mu_c + \lambda_t + \epsilon_{gt}, \quad (1.4) \end{aligned}$$

where R_{gt} is the number of new refugees at the grid-level, $dist_{gt}$ is the distance to the nearest refugee center (in meters), which we include in levels and as interaction with each year λ_t, thereby allowing the effect of distance to vary over time. μ_c are fixed effects for counties. We obtain yearly totals for each county by aggregating the predicted values for incoming refugees based on Equation 1.4. We then use the predicted county-level refugee inflows over two years as instrument in our first-stage equation 1.2.[31]

Despite our regressions controlling for other potentially important distances, a skeptical reader might remain unconvinced that distances to the nearest refugee resettlement centers satisfy the exclusion restriction. In response, we estimate several variants of Equation 1.4, in which we predict the number of incoming refugees at the grid cell level, based on the interaction of distances to a refugee center and the total number of incoming

[30] We use information on education at the commuting zone level, for both immigrant flows and native residents.

[31] We exclude Alaska since it represents an outlier in terms of the relevant distances given the overall size of the state and the fact that Alaska hosts only one refugee processing center in Anchorage. We also exclude Wyoming from this exercise since the state has never been a part of the U.S. refugee resettlement program.

refugees (as variously measured). This approach constitutes a melding of our two empirical approaches, one in which the predicted numbers of refugees are implemented as our initial 'shares'. Additionally, we consider an important aspect of the specific refugee allocation process by taking into account the fact that centers typically specialize in the resettlement of refugees from specific origins.[32]

We estimate three variants of our zero-stage regression, in all cases including grid-cell- (as opposed to county-) fixed effects.[33] Our first variant predicts the number of refugees at the grid-cell level based on a cell's distance to the nearest refugee resettlement center and the *total* number of incoming refugees in a year at the *state-level*. Our second approach considers that some refugee resettlement centers specialize in resettling refugees from specific origins. For each country of origin, we code a binary indicator identifying grid cells that are located within 100 kilometers of a refugee resettlement center that received at least one refugee from that origin in the first year refugees from the country were registered in a state. We then interact this indicator with the number of incoming refugees from that country over the previous election cycle. Aggregating the number of predicted refugees in the same county over all countries of origin yields the total number of predicted refugees at the county level, which we again employ as our instrument. Our third and most conservative approach replaces state-level inflows with (country-of-origin-specific) refugee inflows to the United States at large. The remaining county-level variation is therefore driven exclusively by year-on-year differences in incoming refugees from specific origins to the U.S. and their subsequent allocation across space based on the relative distances to grid cells within 100 kilometers of resettlement centers that had themselves resettled refugees from specific origins in the preceding years. The exclusion restriction is particularly unlikely to be violated in this instance.[34]

1.4 Results

1.4.1 Baseline Results

Table 1.1 reports our baseline results, while omitting coefficient estimates for the control variables for the sake of brevity.[35] Column 1 (Extreme vs. Moderate) adopts the perspective of campaign donors and presents the polarization in donations as the difference

[32]Of the nine voluntary agencies that currently work to resettle refugees across the U.S. all but one—the International Rescue Committee—are affiliated to a specific religious or alternatively faith-based organization, which in turn has naturally resulted in various voluntary agencies developing expertise for clientele from specific origins (Christensen and Ebrahim, 2006). For example, in 1975, the overwhelming majority of Indochinese refugees were resettled by the Catholic Conference (Parsons and Vézina, 2018).

[33]Note that these fixed effects capture all relevant time-invariant distances.

[34]When we construct our instrument in analogy to the analyses of all immigration flows above, results are similar. These results are available on request.

[35]We show our full results in Table B.1 in the Appendix.

in contributions of extreme relative to moderate candidates. Column 2 (Winner) instead focuses on the ideology of the winning candidates, which we contrast with the share of total votes that goes to the Republican candidate (in a county) for comparison (in column 3). Columns 4 and 5 (Winner if Rep/Dem) present results of the ideology of the election winner, given they are Republicans or Democrats respectively. Results defining polarization as the absolute differences between the ideologies of winners and losers are reported in column 6 (Winner vs. Loser). The remaining columns 7-10 focus on binary variables that indicate whether winning candidates are conservative Republican, moderate Republican, moderate Democrat, or liberal Democrat. As these categories are both exhaustive and mutually exclusive, the coefficients from across the four regressions sum to zero. In concert, these variables allow us to test the effect of immigration on polarization, as well as shifts in the overall ideological spectrum.

We report four specifications in each of the ten columns of Table 1.1. Panel A presents the results from ordinary least-squares (OLS) regressions that leverage within county variation. Counties experiencing larger net inflows of immigrants relative to their populations become more polarized in terms of campaign donations originating from those counties in tandem with larger vote shares for the Republican party. Winning candidates experienced a rightward shift in their ideology. Polarization therefore increased as measured by the ideological distance between the winner relative to the loser. The probability of conservative Republicans winning increased significantly, while conversely, moderate Democrats were less likely to be victorious. There is no significant correlation between immigration and the probability of moderate Republicans or left leaning Democrats being elected. The same holds true for the ideology of Republican winners, while Democratic winners shifted leftwards.

Panel B reports the reduced-form estimates for the same set of regressions. Here we regress our measures of ideology and polarization on our instrumental variable (in addition to our controls). If our identification strategy holds in the presence of an effect of immigration on ideology, we should also observe strong reduced-form effects. Indeed, there is a sizable and significant effect of the instrument on ideology and polarization in six of the regressions. This effect will be passed through with the same sign if i) the corresponding first-stage regression is sufficiently strong and ii) the coefficients on our instrument are positive. According to our results, there is no significant reduced-form relationship for the election probability of moderate candidates (for both Democrats and Republicans), the ideology of winning candidates from the Democratic party and the ideology of the winner compared to those of the loser. These insignificant results foreshadow the results of the second stage, to which we turn next.

Table 1.1 – Immigration and Ideology, 1992-2016, Two-year Net Inflows

	(1) Extreme vs. moderate	(2) Winner	(3) Rep. vote share	(4) Winner if Rep.	(5) Winner if Dem.	(6) Winner vs. loser	(7) Right Rep.	(8) Mod. Rep.	(9) Mod. Dem.	(10) Left Dem.
	Panel A: OLS estimates									
ΔImmigrant share	56.858*	8.440**	3.101***	1.759	-3.437**	5.158*	4.557***	1.309	-3.677***	-2.158
	(30.728)	(3.420)	(0.847)	(1.387)	(1.644)	(2.814)	(1.213)	(1.993)	(1.247)	(2.115)
	Panel B: Reduced-form estimates									
Immigrant share IV	9.891***	2.181***	0.507***	0.555***	-0.319	1.217	0.946***	0.082	-0.390	-0.644***
	(2.711)	(0.450)	(0.125)	(0.157)	(0.253)	(0.853)	(0.162)	(0.319)	(0.372)	(0.212)
	Panel C: Second-stage estimates									
ΔImmigrant share	249.685***	55.130***	12.804***	14.488***	-8.667	30.963	23.880***	2.077	-9.840	-16.260***
	(81.515)	(14.404)	(3.805)	(4.388)	(7.850)	(23.902)	(4.727)	(8.177)	(9.958)	(5.507)
	Panel D: First-stage estimates									
Immigrant share IV	0.040***	0.040***	0.040***	0.038***	0.037***	0.039***	0.040***	0.040***	0.040***	0.040***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	40,023	39,514	40,019	27,181	14,287	31,618	39,624	39,624	39,624	39,624
K-P F-stat.	78.22	76.93	78.24	103.6	42.02	66.25	78.68	78.68	78.68	78.68

Notes: The dependent variables are the difference in contributions to extreme compared to moderate candidates (1), ideology of the winning candidates (2), share of total votes that goes to the Republican candidate (3), ideology of the election winner given that they are Republicans (4) or Democrats (5), absolute difference between the ideology of the winner and loser (6), probability the winning candidate is a conservative Republican (7), moderate Republican (8), moderate Democrat (9), or liberal Democrat (10). ΔImmigrant share measures the net inflow of adult immigrants as a share of adult population over the previous two years. All regressions include the full set of control variables, population weights and fixed effects for counties and years (see Table B.1 for the full set of 2SLS results including control variables). Standard errors clustered at the state-level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Panel C in Table 1.1 presents our main results in which we instrument the net inflow of immigrants as a share of the adult population over the two previous years with our shift-share instrument introduced above. As shown in column 1, and in line with our expectations, immigration significantly increased polarization.[36] Evaluated at the sample mean, increasing the share of new immigrants in a county by 1 percent raises the difference between extreme and moderate campaign contributions (in dollars) by 0.89 percent. This coefficient is more than four times the size of the corresponding OLS estimate. Measurement error, reverse causality and omitted variables therefore conspire to bias our OLS coefficients downwards, therein highlighting the need for instrumentation.

Column 2 shows that immigration shifts the ideology of the winner rightwards. Specifically, an increase in the share of immigrants from the 25th to the 75th percentiles shifts the ideology of winners by 0.23 points to the right. This represents an increase of approximately 20 percent of the winners' ideological interquartile range (-0.077 and 1.08). The result could reflect one of two things, or a combination thereof. First an increase in the frequency of Republican candidates winning election, with those candidates being to the right of their Democratic counterparts. Alternatively, the result could capture the Republican candidate moving to the right of their own party. Indeed, the results in column 3 show that the vote share of the Republican party increases with immigration; an increase in immigration inflows from the 25th to the 75th percentile results in an increase in the Republican vote share by 5.42 percentage points. This result is comparable with Mayda et al. (2022), who focus on immigrant stocks as opposed to shares.[37]

To the extent that winning candidates are more likely Republican, the observed rightward shift in ideology in column 2 could follow mechanically. Our results in column 4 however show that the ideology of winning Republicans also moves further to the right. Contrasting the magnitudes of the coefficients in columns 2 and 4 proves informative. The large observed effects in column 2 can be explained by a combination of more Republican candidates winning, in tandem with those winners moving further to the political right. Increasing the immigrant share from the 25th to the 75th percentile shifts the ideology of Republican winners to the right by around 0.06. This is approximately 20 percent of the interquartile range of the ideology of Republican winners (which is 0.83 at the 25th and 1.15 at the 75th percentile). Column 5 demonstrates that the ideology of winning Democratic candidates shifts to the left with larger immigration, although that coefficient is imprecisely estimated.

The same holds for our second measure of polarization, the absolute difference between

[36] This result continues to hold when we focus solely on primary elections, if primary and general elections are combined or if we exclusively include individuals as donors. Falls in moderate contributions drive the result.

[37] According to their results, an increase in low-skilled immigrants of one percent of the population increases the Republican vote share by more than three percentage points (while high-skilled immigrants reduce the Republican vote share).

the ideologies of the winners and losers. According to column 6, the coefficient is positive and substantive, but not significant at conventional levels. The remaining columns of Table 1.1 show that the political spectrum shifts to the right in counties experiencing larger immigration inflows. The probability of conservative Republican candidates winning election increases by more than 10 percentage points when our measure of immigration rises from the 25th to the 75th percentile. This comes at the expense of liberal Democrats, whose probability of winning declines by almost 7 percentage points.[38]

In summary, we provide evidence in line with immigration polarizing campaign donors' contributions, and shifting ideologies politically rightward, particularly among Republican election winners. Given that more extreme Republican candidates also enter office more frequently in response to increased immigration, overall the ideologies of elected politicians turn substantially rightwards. Comparing our second-stage coefficients to our OLS results in Panel A shows they both operate in the same direction, although the OLS coefficients are smaller in absolute terms.

Panel D in Table 1.1 reports our corresponding first-stage regressions. Reassuringly, none of our estimates suffer from a weak-instrument problem. The coefficients are highly significant and all associated first-stage F-statistics exceed 40.[39] As expected, we observe a positive relationship between the shift-share instrument and immigration flows. A typical (one-standard deviation) increase in our instrument—equivalent to around 0.01—increases net immigrant flows by about 4,613 immigrants in a county hosting 109,183 immigrants (the 99th percentile in 1992), but only by approximately 10 immigrants in a county with a stock of 237 immigrants (the median in 1992).

1.4.2 Alternative Measures

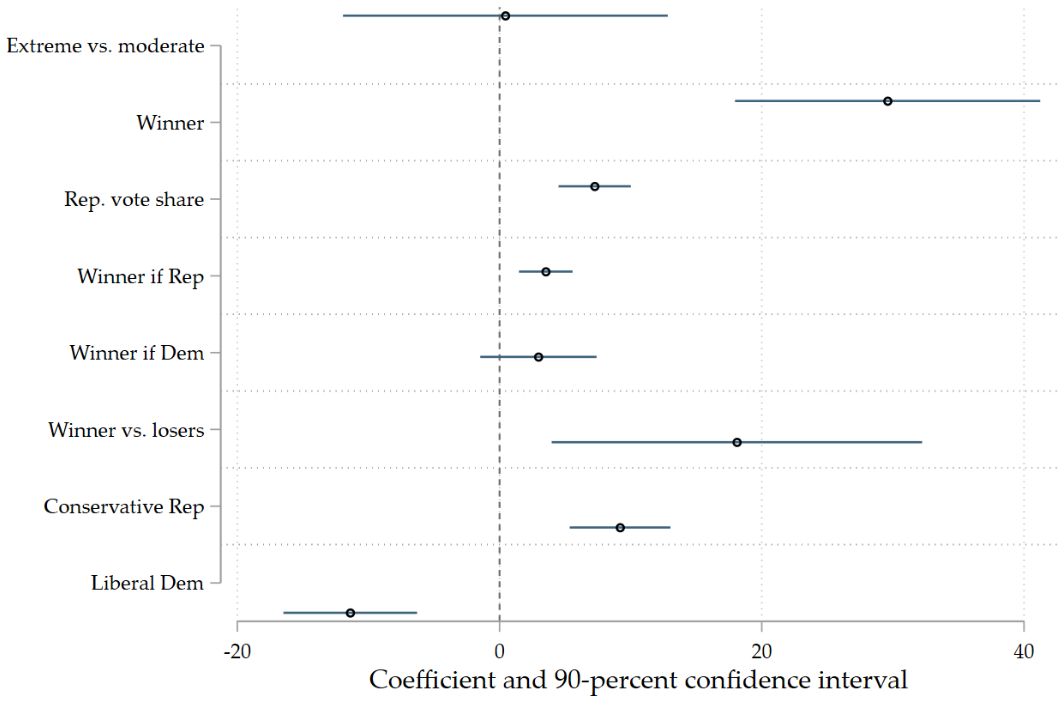
We proceed by testing alternative immigration measures. Figure 1.3 illustrates results of estimates analogous to our baseline in Table 1.1, focusing instead on changes in the stock of immigrants over eight year periods. The figure presents our estimated marginal effects in tandem with the associated 90-percent confidence intervals. The corresponding full regression results are provided in Table B.4 in the Appendix.[40]

[38]While we also observe small gains for moderate Republicans in tandem with (more substantial) losses for moderate Democrats, these effects are imprecisely estimated.

[39]They are thus considerably larger than the conventional rule-of-thumb value of 10. They remain strong when we compute F-statistics that are robust to heteroskedasticity, autocorrelation, and clustering (Olea and Pflueger, 2013). The Montiel-Pflueger effective F-statistic for column 1, for example, is above the corresponding critical value for a 5-percent "worst-case" bias at the 1-percent confidence level (Olea and Pflueger, 2013). The coefficient in column 1 falls also within the Anderson–Rubin 95-percent confidence interval.

[40]The first-stage F-statistics remain strong in these regressions with the exception of those in column 5 of Table B.4.

Figure 1.3 – Immigration and Ideology, 1992-2016, Eight-year Net Inflows



Notes: The figure reports the coefficients of net adult immigration over eight years, in tandem with 90-percent confidence intervals. The coefficient of extreme vs. moderate is multiplied with 0.1. See Table B.4 for details.

Our results for immigration over eight years, as opposed to just two, are broadly similar to our baseline estimates, although the coefficients are smaller in magnitude. The polarizing effects of immigration are therefore attenuated over time, which might be suggestive of some underlying process of acceptance.

We further examine our core hypothesis, i.e., whether or not migrants on aggregate affect political polarization using an alternative data set, namely *individual* data deriving from a 2016 cross-section of Twitter accounts. While we would like to apply the same method as before when testing the effects of immigration on polarization, we are restricted by the availability of data. Given the low uptake of Twitter in earlier years we restrict our analysis to a cross-section for the year 2016. We make use of a first stage analogous to those in column 1 of Table 1 (using the full sample) and estimate second-stage regressions with the same set of control variables included.[41]

Table 1.2 presents the results. Immigration shifts the ideology of Twitter users to the right, concurrently increasing the share of extreme users (and therefore by definition reducing the share of moderate users). In quantitative terms, a typical increase in immi-

[41]When we estimate the first stage for 2016 alone, the power of our instrument is low given the comparably low number of observations per county. For the same reason the second stage includes fixed effects for states as opposed to counties.

Table 1.2 – Immigration and Ideology based on Individual 2016 Twitter Accounts

	(1) extreme	(2) right	(3) left	(4) moderate
ΔImmigrant share	66.367***	114.652***	-48.285***	-66.367***
	(11.330)	(21.998)	(11.855)	(11.330)
Constant	0.386***	-0.084	0.470***	0.614***
	(0.093)	(0.150)	(0.071)	(0.093)
Observations	2,529	2,529	2,529	2,529
Kleibergen-Paap F-stat	78.23	78.23	78.23	78.23

Notes: Extreme indicates the share of left and right users in all users in a county. Right/left/moderate are the respective shares in all county Twitter users. The first stage is estimated over the full sample of Table 1.1, column 1, including control variables, as well as fixed effects for years and counties. The second stage is a cross-section for the year 2016 and includes fixed effects for years and states. We have bootstrapped standard errors and clustered them at the state-level. All regressions include the full set of control variables, population weights and fixed effects for counties and years.

gration from the 25th to the 75th percentile increases the share of right Twitter users by 48.6 percentage points resulting in an increase of 28.1 percentage points of extreme Twitter users. These results corroborate those obtained with our campaign donation-based measures of polarization above.[42]

1.4.3 Robustness

Shift-share design

We test the plausibility of our exclusion restriction along a number of dimensions, guided by recent advances in the related literature. Figure B.1 in the Appendix focuses on non-linear trends. While linear trends would be captured by our set of fixed effects, Christian and Barrett (2017) have shown that non-linear trends can lead to spurious inference, in a setting broadly related to ours. Following Christian and Barrett (2017), we plot the variation in immigration and polarization for different groups that are defined according to the percentiles of the immigrant shares in 1980, in tandem with the yearly values of net immigration. Specifically, Panel A of Figure B.1 presents immigrant net inflows as a share of the adult population. Panel B shows the same variable at the county level, according to percentiles of the initial share of immigrants in 1980 (netting out the effects of our control variables that we include in all regressions). Panel C focuses on extreme versus moderate campaign contributions for the same percentiles. Figure B.1 provides no basis to believe that we violate the parallel trends assumption. The trends in immigration and

[42] Similarly, analyzing GALLUP data ("How would you describe your political views?–very conservative, conservative, moderate, liberal, very liberal") provides additional external validity to our baseline results. In particular, the share of very conservative voters increases as a consequence of immigration. These results are available on request.

moderate versus extreme campaign contributions, respectively, do indeed appear parallel across percentiles.[43] Neither are non-linear trends apparent. Reassuringly, no non-linear trend overlaps the trend in net immigration at the county level (a common trend in all variables that is otherwise indifferent across percentiles would be captured by our year-fixed effects).

We further test the potential importance of pre-trends, following Mayda et al. (2022) and Goldsmith-Pinkham et al. (2020). First, we provide visual evidence in Figure B.2 that plots the correlation between the change in predicted net immigration (1992-2016) and the change in our outcome measure "Extreme vs. moderate" in earlier years (1982-1988). The straight line indicates that the correlation is essentially zero; it is also insignificant at conventional levels. This demonstrates an absence of pre-trends in our outcome which are correlated with changes in predicted immigration.[44]

Second, reverse causality or trends in other variables that are correlated with changes in our instrumental variable could bias our coefficients. Larger Republican vote shares for example could reduce immigration, which in turn could affect the Republican vote share. We therefore test the effect of changes in the same set of (local economic, demographic and ideology) variables over the 1980-1990 period on changes in the shift-share instrument in two-year increments. We again focus on the 1992-2016 period and include the same set of control variables as in the main regressions in addition to year-fixed effects. According to column 1 of Table B.5, the correlations between the changes in our instrumental variable and polarization and ideology measured as the differences between 1982 and 1988 are small and insignificant at conventional levels. The one exception is the difference in "Winner if Republican" between 1982 and 1988, which is marginally significant. Note however that with a 10-percent significance level, one of the 10 regressions in column 1 is significant by chance. Column 2 rather presents analogous (conditional) correlations between changes in our instrumental variable and eleven economic and demographic variables measured as the differences between 1980 and 1990. All are insignificant.

Finally, we consider how the dynamics of our instrumental variable could threaten identification. According to Jaeger et al. (2018), the analysis of immigration responses based on shift-share instruments may conflate the short- and long-run effects of immigration. Jaeger et al. (2018) argue that in order for the instrument to be valid, there should be either no dynamic adjustment process in the outcome variable, or the shifts in (changes

[43] The same holds for our other outcome variables, although we do not report them for the sake of brevity.

[44] We also calculate the correlation between the country-of-origin-specific initial shares in 1980 and changes in local economic, demographic and ideology variables over the 1980-1990 period. Following Mayda et al. (2022) we focus on 14 groups of origin countries to calculate these shares: Mexico, Canada, Rest of Americas, Western Europe, Eastern Europe, China, Japan, Korea, Philippines, India, Rest of Asia, Africa, Oceania and Others. The correlations of these shares with the pre-determined changes in outcome measures are close to zero. All correlations between these shares and the pre-determined local economic and demographic characteristics are smaller than 0.18.

of) immigration at the national level should not be serially correlated. In our sample, the correlation of net immigration at the county level from one year to the next is 0.1 (see also Panel A of Figure B.1). When we further include the instrumental variable in t and $t-1$ in our reduced-form regressions, as in column 1 of Panel B in Table 1.1, we find the contemporaneous effect remains significant, while the coefficient of the lagged instrument is insignificant.[45]

Falsification Exercises

We continue by testing whether our results are driven by omitted variables that are systematically correlated with immigration over time within counties, or across counties at specific points in time. To this end, we randomly assign immigrants across these two dimensions. First, we assign immigrants of each particular year to a random year for the same county. Second, we assign immigrants of one county in each year to a random county in the same year. Third, we randomly assign immigrants across counties and years simultaneously. Figure B.3 (based on the specification of column 1 in Table 1.1) in the Appendix, shows the point estimate coefficients resulting from 5,000 such randomizations for each of the three procedures, in concert with the p-values, which we calculated as the proportion of times that the absolute value of the t-statistics in the simulated data exceeds the absolute value of the original t-statistic. The coefficients are clearly centered around zero and rarely exceed the coefficient of column 1 in Table 1.1 (which is indicated by the dashed vertical lines).

1.4.4 Heterogeneous Effects

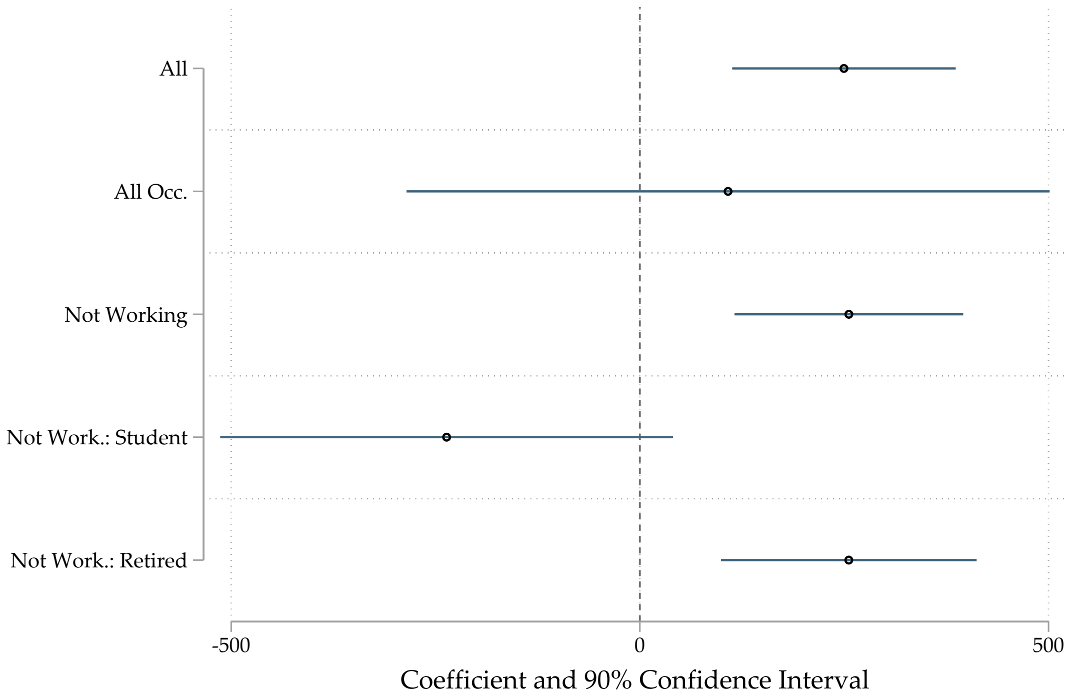
Our analysis captures the *local* effects of immigration, since any country-wide effects are absorbed into our year-fixed effects. Our results can therefore be perceived from the perspective of contact theory (Allport, 1954) and group threat theory (Sherif et al., 1961; Campbell, 1965), which both provide natural heuristics as a means to further interpret our results. The economics literature in this sphere suggests the degree to which native populations feel economically threatened by immigrants depends upon the level of competition for jobs between the two groups, as well as the transfers and public services they receive (Mayda, 2006; Facchini and Mayda, 2009; Cavaille and Ferwerda, 2020).[46] Anti-immigration attitudes have also been related to a taste for cultural homogeneity (Dustmann and Preston, 2007; Card et al., 2012). Cultural threats may depend on the incompatibility of norms and values as well as the size of the incoming group (Brown, 2000; Bansak et al., 2016). Collectively, these theories suggest that migrants can potentially

[45] The coefficient of the contemporaneous instrument falls from 9.89 to 6.35. We do not report these results in a table—details of which are available on request.

[46] Please also refer to Gehring (2022).

increase prejudice if perceived as competitors, a situation that can be reversed should suitable conditions that enhance knowledge be satisfied.

Figure 1.4 – Immigration and Ideology, 1992-2016, Employment Status of Donors

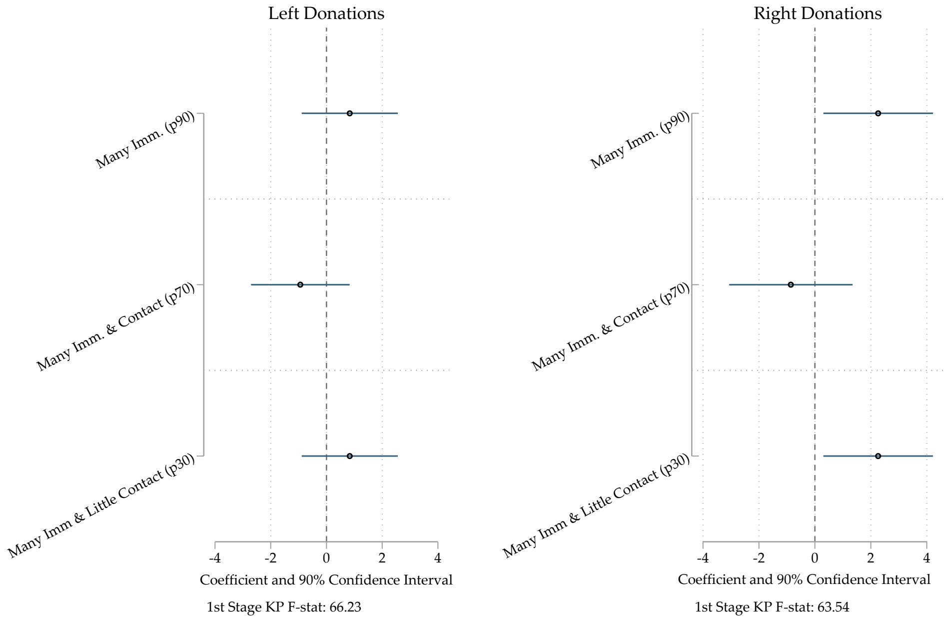


Notes: The figure reports the coefficients of net immigration over two years, in tandem with 90-percent confidence intervals. The outcomes are defined as the inverse hyperbolic sine of the difference of extreme vs. moderate contributions from all, all working, all not working, students and retired contributors respectively. The first stage corresponds to Table 1.1, column 1.

To tease out some of the intricacies at play, we continue by splitting our sample along a number of dimensions of campaign contribution donor characteristics, as detailed in Section 1.3.1. Specifically we exploit donors' employment status, their origins as inferred from their surnames and the degree to which donors likely come into direct contact at work with immigrants as captured by the *nature* of donor's employment in addition to what proportions of immigrants are typically employed in those specific occupations.

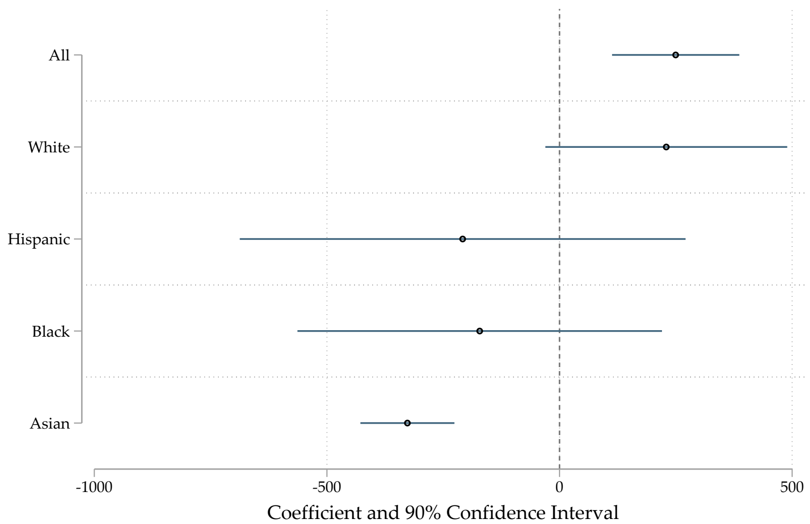
These results are presented in Figures 1.4 to 1.6. Figure 1.4 provides some evidence in favor of our *extreme-moderate* measure of political polarization being driven by those not employed and those in retirement. Unpacking our imprecisely estimated occupation estimate from Figure 1.4, Figure 1.5 digs a little deeper by examining donations to the political left and right, according to the degree to which donors' employment brings them into direct contact with others in addition to the proportion of immigrants typically employed in those occupations.

Figure 1.5 − Immigration and Ideology, 1992-2016, Likely Contact with Donors



Notes: The figure reports the coefficients of net immigration over two years, in tandem with 90-percent confidence intervals. The outcomes in the left graph are defined as the share of donations among left contributions from donors working in occupations with i) many immigrants (90-percentile), ii) many immigrants and much contact (70-percentile), and iii) many immigrants and little contact (30-percentile) respectively. The Kleibergen-Paap F-statistic for the first stage is 66.23. The right graph repeats the exercise for right donations. The Kleibergen-Paap F-statistic for the first stage is 63.54.

Figure 1.6 − Immigration and Ideology, 1992-2016, Race of Donors



Notes: The figure reports the coefficients of net immigration over two years, in tandem with 90-percent confidence intervals. The outcomes are defined as the inverse hyperbolic sine of the difference of extreme vs. moderate contributions from all, White, Hispanic, Black and Asian contributors respectively. The first stage corresponds to Table 1.1, column 1.

This analysis reveals that right-wing donations are increasingly driven by donors working in occupations with high proportions of immigrants, and in cases in which such donors have little contact in their daily work-lives with immigrants. Greater interactions with immigrants attenuate this result. Finally, Figure 1.6 leverages donors' ethnic background, which provides some evidence that our *extreme-moderate* measure of polarization is driven by whites as opposed to other ethnic groups, especially Asians, who rather exert a strong and opposing influence. This result could be explained by Asians having far more familiarity and contact with specific immigrant groups, not least since four Asian countries (Philippines, India, China, and Vietnam) represent four of the top five migrant groups in the United States (with the other being Mexico).

We continue by testing whether cultural and educational distances between incumbents and immigrants mediate or exacerbate our previous estimates.[47] To this end we interact the share of immigrants arriving in a county with indicators of cultural and educational distance (focusing on net immigration inflows over a two year time horizon). We provide full regression results in the Appendix (in Tables B.2 and B.3) and illustrate the results for significant interactions in figures. Since we adopt a control function approach (CFA), the first-stage regressions (and F-statistics) are fundamentally comparable with those reported in Table 1.1.[48]

The effects of immigration on rightward shifts in ideology become more pronounced when cultural distances are greater, since the ideologies of winners shift further to the political right. This effect is due to the increased probability of conservative Republicans winning elections. As shown in Figure 1.7, these interactions result in marginal effects that are significant throughout the ranges of cultural distance for the ideologies of winners and the probabilities of conservative Republicans winning. An increase in immigration from the 25th to the 75th percentile for example increases the probability of a conservative Republican winning by 9.57 percentage points if immigrants are culturally similar to the resident population (the 25th percentile of the distance variable). This effect increases to 12.13 percentage points however when the cultural distance between the two groups increases to the 75th percentile.[49] An increase in immigration over the same interquartile range similarly results in rightward ideological shifts of winners by between 0.21 and 0.29 points, while concurrently increasing the Republican vote share by 5.15 and 6.42

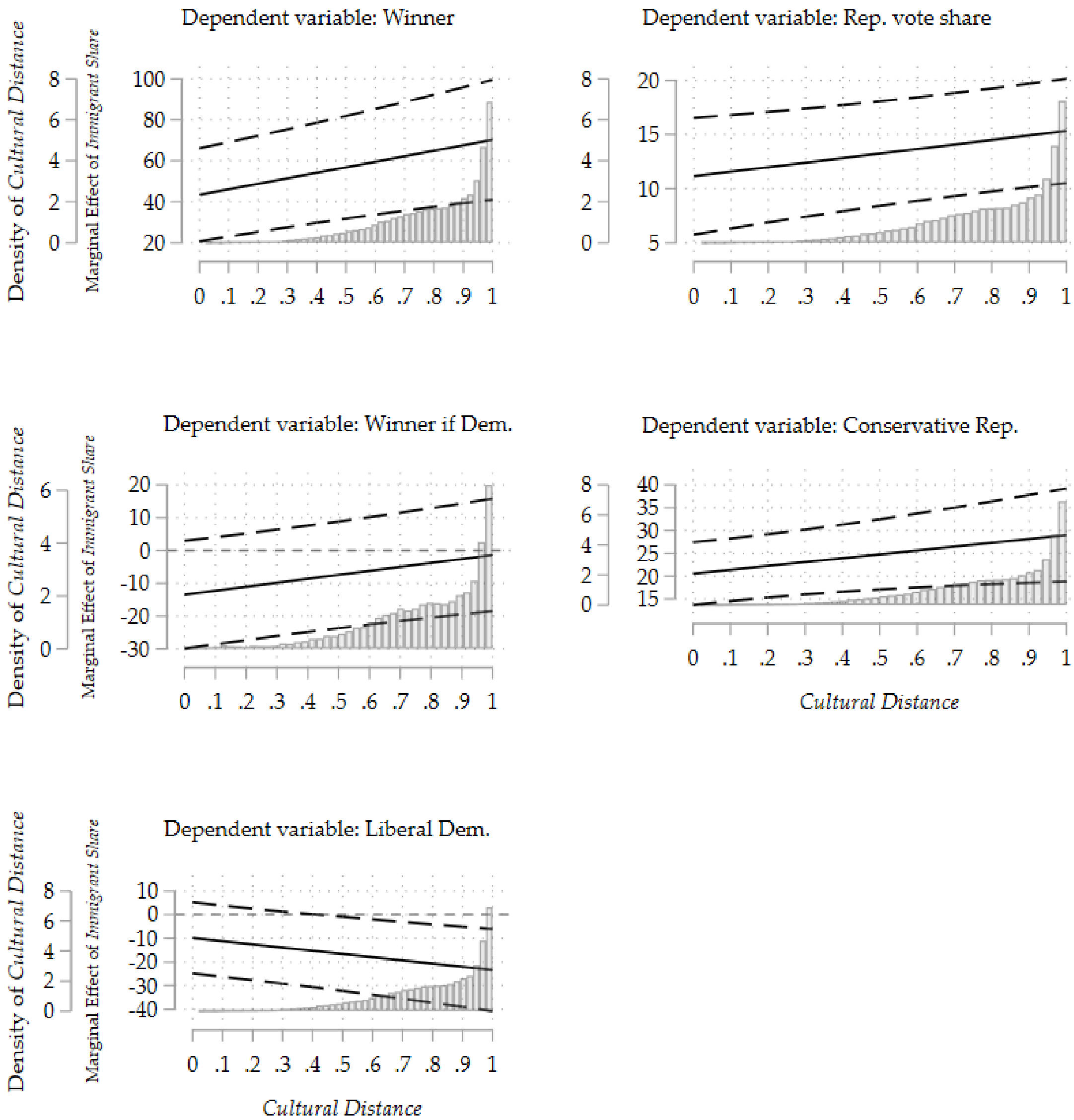
[47] In these additional regressions we no longer report results for the (insignificant) effects of ideology on the probability of moderate candidates winning, to reduce clutter.

[48] The CFA controls for the first-stage regression residual in the second stages. Alternatives to this approach are 2SLS employing the interaction of our instrument with the distance indicators as additional instruments. This would treat the interactions as separate endogenous variables, which "can be quite inefficient relative to the more parsimonious CF approach" (Wooldridge 2015, p. 429). The resulting increase in efficiency comes at the cost of an additional assumption; that is, we need to assume that the bias does not depend on distance. Note that the number of observations falls because we do not have complete data for either distance. Our first stages consequently differ too, but first-stage F-statistics remain sufficiently high (as shown in the Appendix).

[49] Cultural distance takes on the value of 0.24 at the 25th percentile and 0.96 at the 75th percentile.

percentage points, respectively.

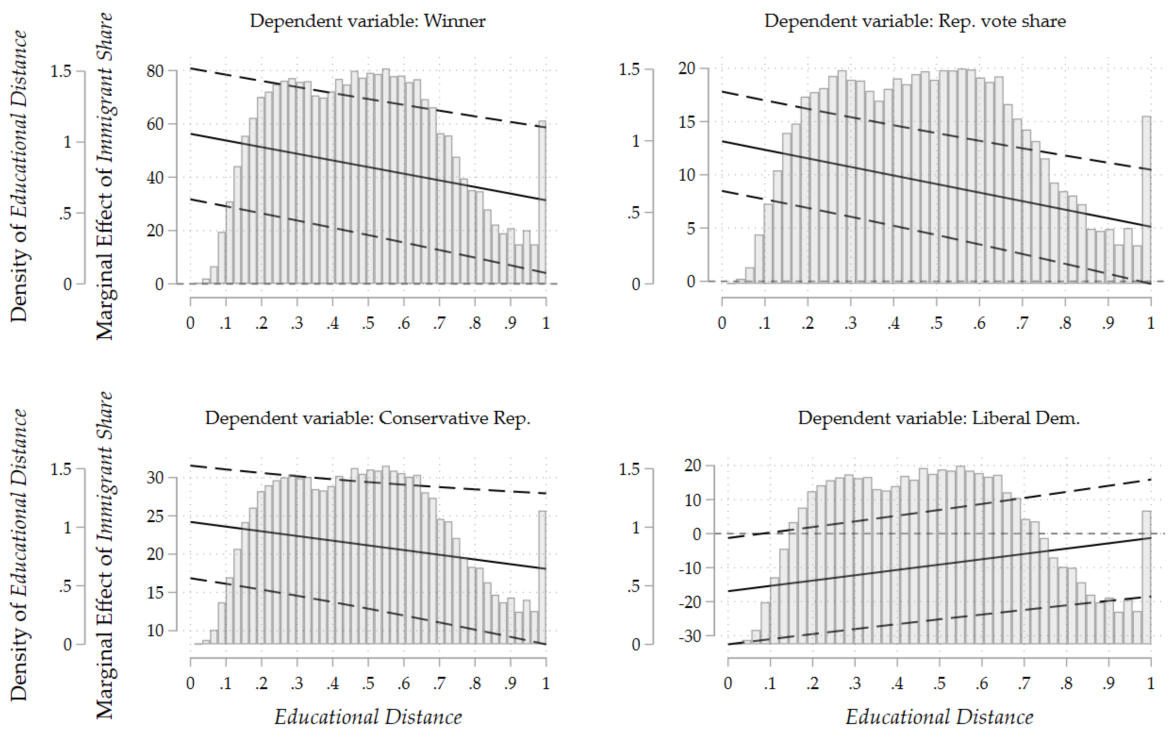
Figure 1.7 – Immigration, Ideology and Cultural Distance, 1992-2016, Two-year Net Inflows



Notes: The figure shows partial leverage plots for the regressions reported in columns 2, 3, 5, 7, and 8 of Table B.2. The dashed lines indicate 90-percent confidence intervals.

Increases in educational distance rather operate in the opposite direction. Figure 1.8 plots the marginal effects for our significant interactions. These show that the probability of conservative Republicans winning election is significant across the full range of our educational distance measure. An increase in immigration from the 25th to the 75th percentile increases the probability of a conservative Republican winning by 9.47 percentage points, if immigrants have a similar educational background compared to the resident population (the 25th percentile of the distance variable). This increase is 8.57 percentage points if immigrants rather herald from different educational backgrounds as when compared to resident populations (the 75th percentile of the distance variable). Similarly, the effect of immigrants on the Republican vote share is positive unless educational distance exceeds about 0.98 (which only holds for some 2.6 percent of our observations). Conversely, the probability that liberal Democrats win elections declines with educational distance (until this distance is smaller than 0.06, which is the case in 0.1 percent of the observations).[50] Similarly, the rightward shift of the winner declines with decreasing similarity in educational background amounting to 0.21 points at the 25th percentile and 0.17 points at the 75th percentile of the distance variable.

Figure 1.8 – Immigration, Ideology and Educational Distance, 1992-2016, Two-year Net Inflows



Notes: The figure shows partial leverage plots for the regressions reported in columns 2, 3, 7, and 8 of Table B.3. The dashed lines indicate 90-percent confidence intervals.

Taken collectively our results are in line with both contact theory and group threat theory. Natives engage more with culturally closer immigrants, while feeling more threatened by newcomers from more distant cultures. Conversely, labor-market complementarities and reduced labor market competition among people with different education reduce the observed shifts to the political right.

[50]Educational distance takes on the value of 0.30 at the 25th percentile and 0.65 at the 75th percentile.

1.4.5 Refugee Results

Panel A of Table 1.3 reports results for our simple distance-IV. Panels B–D rather report results from our interacted instruments, with grid-cell-fixed effects included in the zero-stage regression in Equation 1.4, and county-fixed effects in the first- and second-stage regressions. These regressions employ the absolute numbers of refugees as opposed to population shares. Given that we include fixed effects for counties, population hardly changes from year to year. When we estimate these regressions as population shares, first-stage F-statistics are however weak, so we do not report these results. They are available on request. Though specific levels of significance vary across specifications, we find that refugee inflows increase extreme vs. moderate donations (the exception being the negative coefficient of Panel A), shift the ideology of the winner rightwards and increase the vote share of the Republican party. Republican winners shift to the right, Democratic winners to the left. Winners shift to the right relative to the runner up. Finally, the entire political spectrum moves to the right.

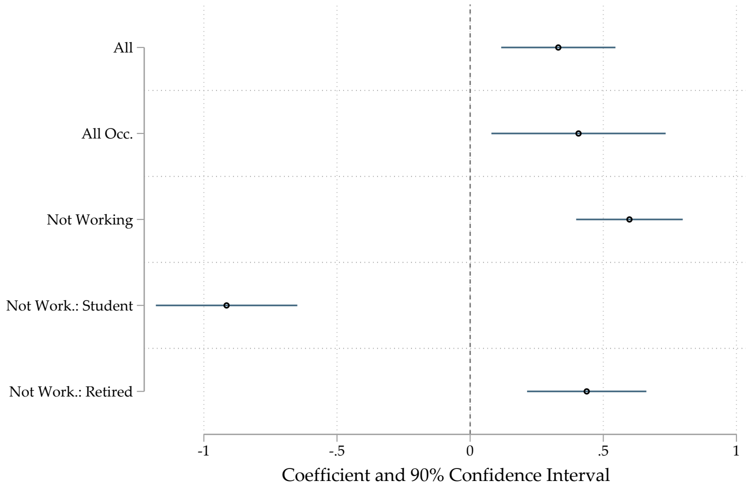
We replicate our donor heterogeneity analysis, assessing which factors play a role in the polarizing response to refugee inflows. This exercise is based on our preferred specification in Panel D. Figures 1.9 and 1.10 explore the potential role of contact by exploiting the occupational characteristics of donors. Once again we find that polarization is driven by those retired and unemployed, as we did for immigrants. In contrast however, the average effect on those employed is also significantly positive. Unpacking compositional changes among donations to the ideological right, we find that donors employed in occupations with high proportions of immigrants and infrequent contact with refugees drive our observed effect, although this effect vanishes in cases in which donors are employed in occupations that involve significant refugee contact. Figure 1.11 reports our results leveraging the ethnic background of donors. These results echo our previous findings for immigrants, highlighting that whites as opposed to other ethnic groups drive the overall polarizing effect in response to refugee inflows. This result is not mechanically determined by whites representing the largest group in our sample.

Table 1.3 − Refugees and Ideology, 1992-2016, Two-year Gross Inflows

	(1) Extreme vs. moderate	(2) Winner	(3) Rep. vote share	(4) Winner if Rep.	(5) Winner if Dem.	(6) Winner vs. loser	(7) Right Rep.	(8) Mod. Rep.	(9) Mod. Dem.	(10) Left Dem.
	Panel A: Distance IV, refugees/population share									
ΔRefugee share	-1190.805	247.595***	112.877***	35.255	-40.472	131.324**	144.622**	15.231	-52.636	-107.160**
	(823.130)	(74.019)	(37.858)	(45.598)	(43.597)	(49.496)	(57.177)	(61.468)	(44.405)	(48.769)
K-P F-stat.	7.07	7.07	7.07	7.62	7.36	6.95	7.09	7.09	7.09	7.09
	Panel B: Interacted IV, state totals									
ΔRefugees	0.27590	0.06236***	0.01372***	0.01382***	-0.02220***	0.04378***	0.03846***	0.01427	-0.03449***	-0.01822
	(0.25281)	(0.01724)	(0.00339)	(0.00426)	(0.00574)	(0.01081)	(0.00803)	(0.01112)	(0.00755)	(0.01166)
K-P F-stat.	70.05	68.67	70.05	104.55	35.54	58.91	68.70	68.70	68.70	68.70
	Panel C: Interacted IV (origin-specific), state totals									
ΔRefugees	0.62091***	0.11303***	0.02014***	0.01965***	-0.02140***	0.08723***	0.05809***	0.02409**	-0.05847***	-0.02368
	(0.18542)	(0.02844)	(0.00552)	(0.00723)	(0.00736)	(0.01740)	(0.01308)	(0.01189)	(0.01119)	(0.01679)
K-P F-stat.	41.64	40.63	41.64	53.90	21.05	33.22	40.74	40.74	40.74	40.74
	Panel D: Interacted IV (origin-specific), U.S. totals									
ΔRefugees	0.32966**	0.09592***	0.02807***	0.00672	-0.02116**	0.06838***	0.02411**	0.03155*	-0.01026	-0.04544***
	(0.12904)	(0.01577)	(0.00602)	(0.00798)	(0.00900)	(0.02288)	(0.00945)	(0.01805)	(0.00789)	(0.01360)
K-P F-stat.	314.95	311.79	314.95	150.90	153.81	260.39	304.41	304.41	304.41	304.41
Observations	39,474	38,966	39,470	26,876	14,466	31,162	39,075	39,075	39,075	39,075

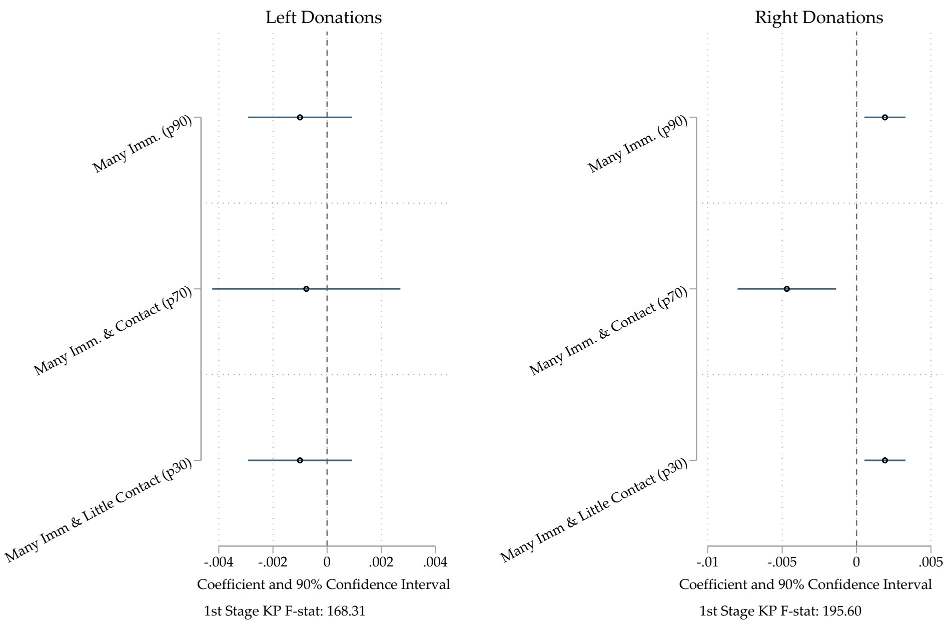
Notes: The dependent variables are the difference in contributions to extreme compared to moderate candidates (1), ideology of the winning candidates (2), share of total votes that goes to the Republican candidate (3), ideology of the election winner given that they are Republicans (4) or Democrats (5), absolute difference between the ideology of the winner and loser (6), probability the winning candidate is a conservative Republican (7), moderate Republican (8), moderate Democrat (9), or liberal Democrat (10). ΔRefugees are gross inflows of refugees over the previous two years. ΔRefugee share are refugee inflows as a share of adult population. All regressions include the full set of control variables, population weights and fixed effects for counties and years. In Panel A, the instrumental variable is the predicted number of refugees relative to county population; Panels B–D use the number of predicted refugees as IV. Fixed effects at the zero stage: county level (Panel A), grid cell level (Panels B–D). In Panels B–D, coefficients and standard errors are multiplied with 1,000. Bootstrapped standard errors (with 500 replications) in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 1.9 − Refugees and Ideology, 1992-2016, Employment Status of Donors



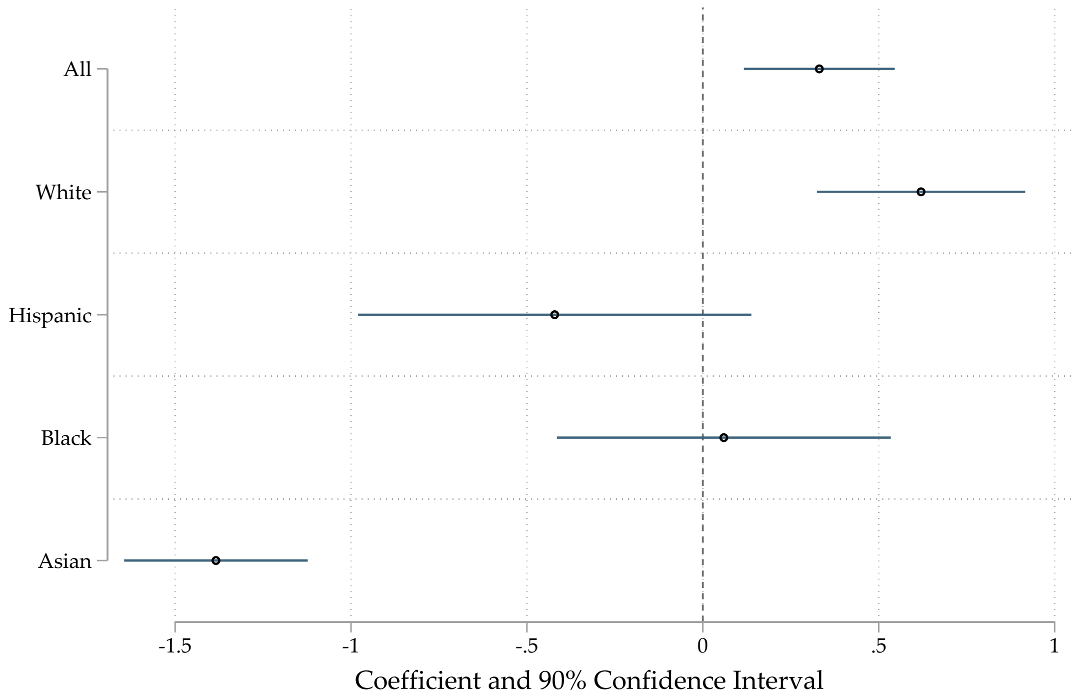
Notes: The figure reports the coefficients of the number of refugees in thousands over two years, in tandem with 90-percent confidence intervals. The outcomes are the inverse hyperbolic sine of the difference of extreme vs. moderate contributions from all, all working, all not working, students and retired contributors respectively. The first stage corresponds to Table 1.3, Panel D, column 1.

Figure 1.10 − Refugees and Ideology, 1992-2016, Likely Contact with Donors



Notes: The figure reports the coefficients of the number of refugees in thousands over two years, in tandem with 90-percent confidence intervals. The outcomes in the left graph are defined as the share of donations among left contributions from donors working in occupations with i) many immigrants (90-percentile), ii) many immigrants and much contact (70-percentile), and iii) many immigrants and little contact (30-percentile) respectively. The Kleibergen-Paap F-statistic for the first stage is 168.31. The right graph repeats the exercise for right donations. The Kleibergen-Paap F-statistic for the first stage is 195.60.

Figure 1.11 – Refugees and Ideology, 1992-2016, Race of Donors



Notes: The figure reports the coefficients of the number of refugees in thousands over two years, in tandem with 90-percent confidence intervals. The outcomes are defined as the inverse hyperbolic sine of the difference of extreme vs. moderate contributions from all, White, Hispanic, Black and Asian contributors respectively. The first stage corresponds to Table 1.3, Panel D, column 1.

1.5 Conclusion

The United States is a nation of immigrants, one profoundly shaped by subsequent arrivals to her shores. Recent decades have ushered in continued high volumes of migrants including refugees, in tandem with significantly diverging, protracted and acute levels of political polarization; so much so, that some argue such polarization represents the single greatest threat to the future of the country. In this paper, we test whether migration causally affects political polarization in the United States. Our data comprise the universe of migrants and refugees as well as the ideologies of 16 million campaign donors and politicians campaigning for election to the House of Representatives in the 1992-2016 period.

Implementing various measures of political polarization, we provide causal evidence that political polarization significantly increases in counties that experienced greater inflows of immigrants over a two-year time horizon. These effects also hold over the longer run, i.e., periods of eight years, although the estimated effects are somewhat attenuated over time. We provide some empirical support for the conjecture that polarizing political campaign donations are driven by whites, the unemployed and those in retirement, with right-wing donations in particular driven by those working in occupations with high proportions of immigrants, especially in which donors have little contact with immigrants in their daily lives. Greater interactions with immigrants attenuate these effects. Our baseline findings are starker the greater the cultural distances between incoming migrants and incumbent natives and the more similar the education levels of the two groups.

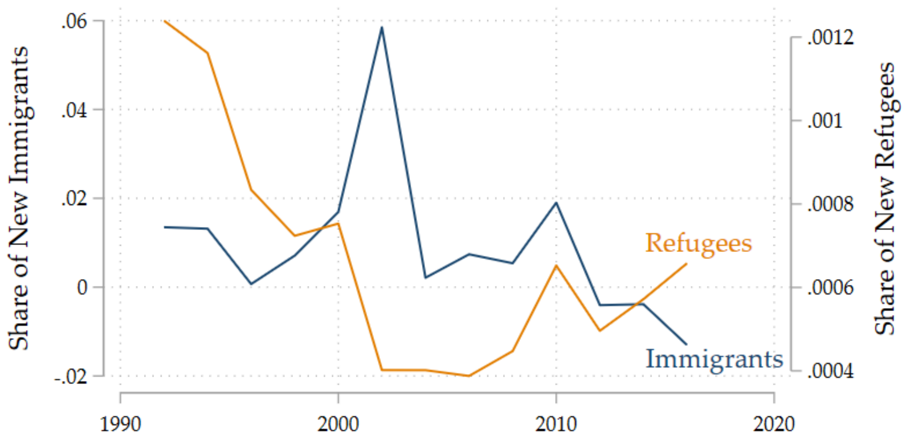
Though refugees differ from other migrants along a number of dimensions, we uncover similar results for refugees and migrants on aggregate, despite adopting an alternative identification strategy; one that leverages the timing and location of refugee processing centers, in tandem with the fact that specific centers specialize in processing refugees from specific origins.

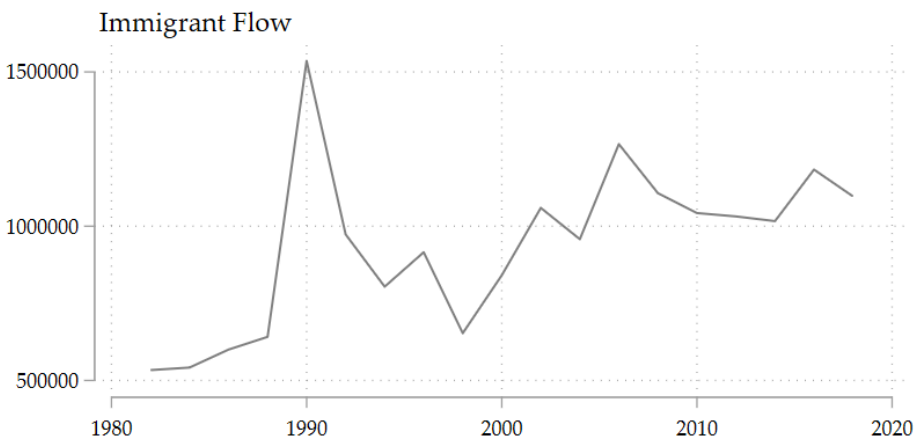
Portes (2011, 424) argues that new immigration is first *"reviled when it is actually taking place and celebrated after a period of time, when the first generation has passed from the scene."* Our results provide some empirical support for the conjecture that this process of acceptance operates more swiftly, but that the local contexts facing immigrants and resettled refugees, including the composition of natives, likely proves pivotal in determining, at least in part, the acute levels of political polarization being witnessed across the United States today.

Appendices

A Descriptives

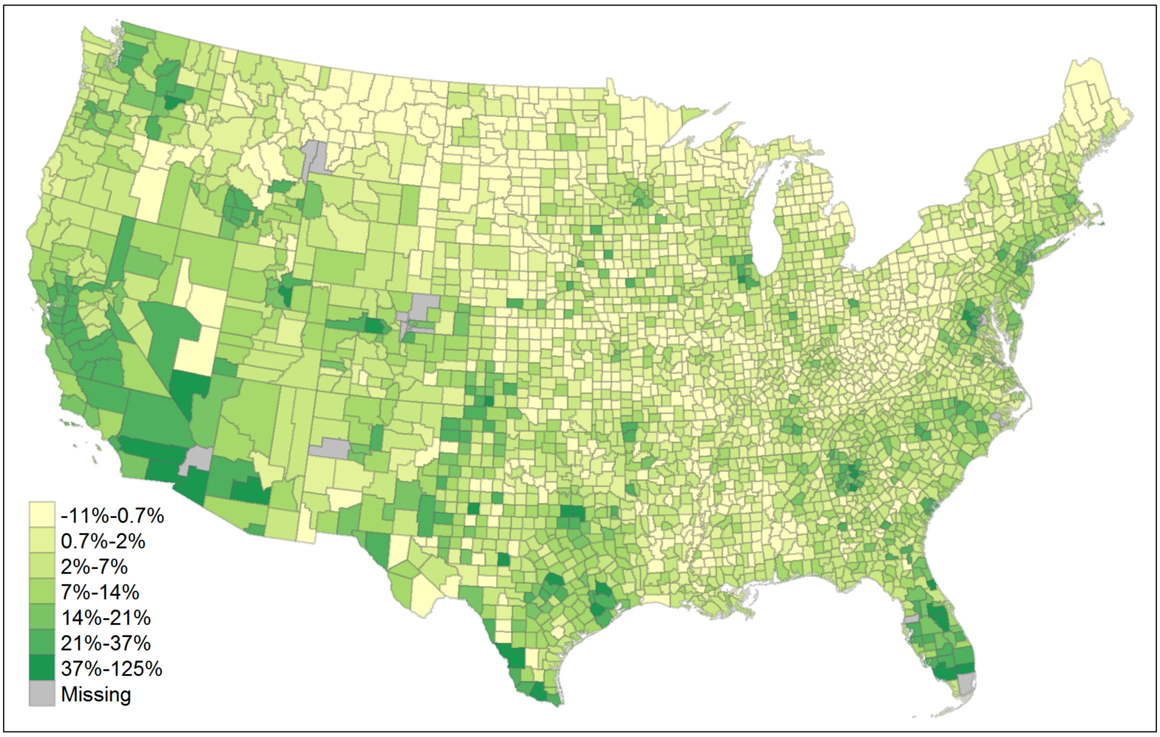
Figure A.1 – Immigrants and Refugees in the United States, 1982-2018, Inflows





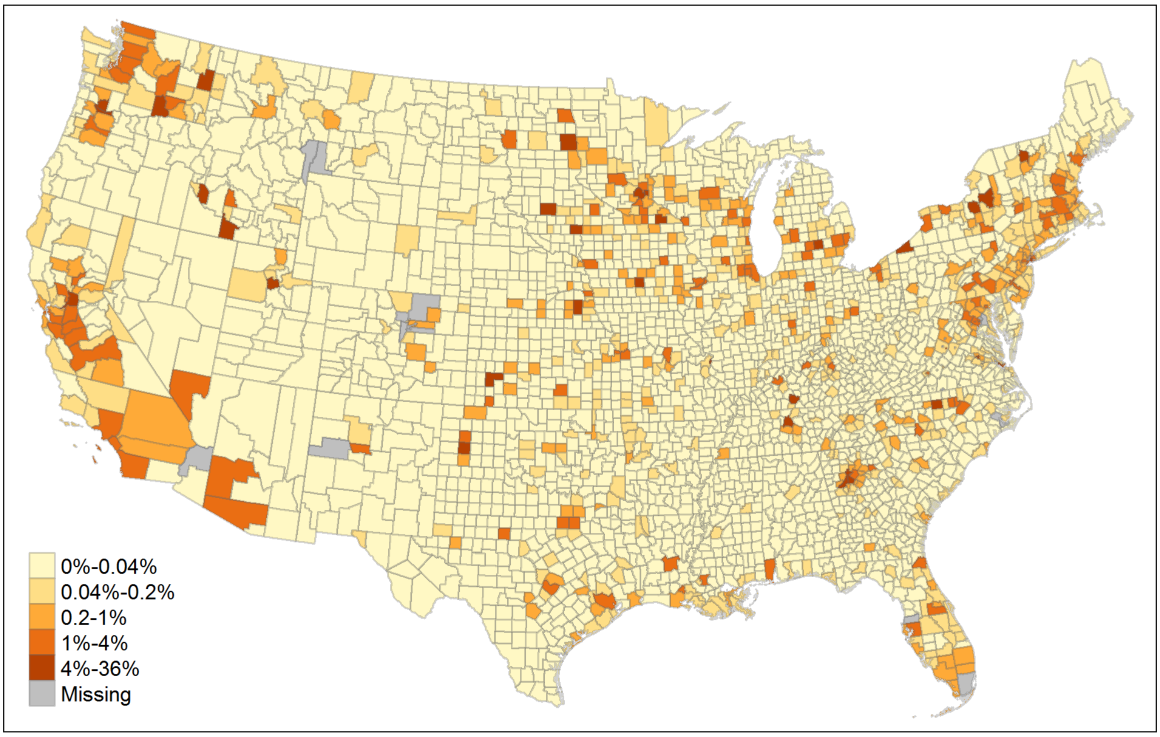
Notes: The upper figure shows net (gross) inflows of adult immigrants (refugees) as a share of the adult population. The lower figure shows the number of foreign nationals that were granted lawful permanent residence.

Figure A.2 – Immigrants in the United States by County, 1992-2016, Net Inflows



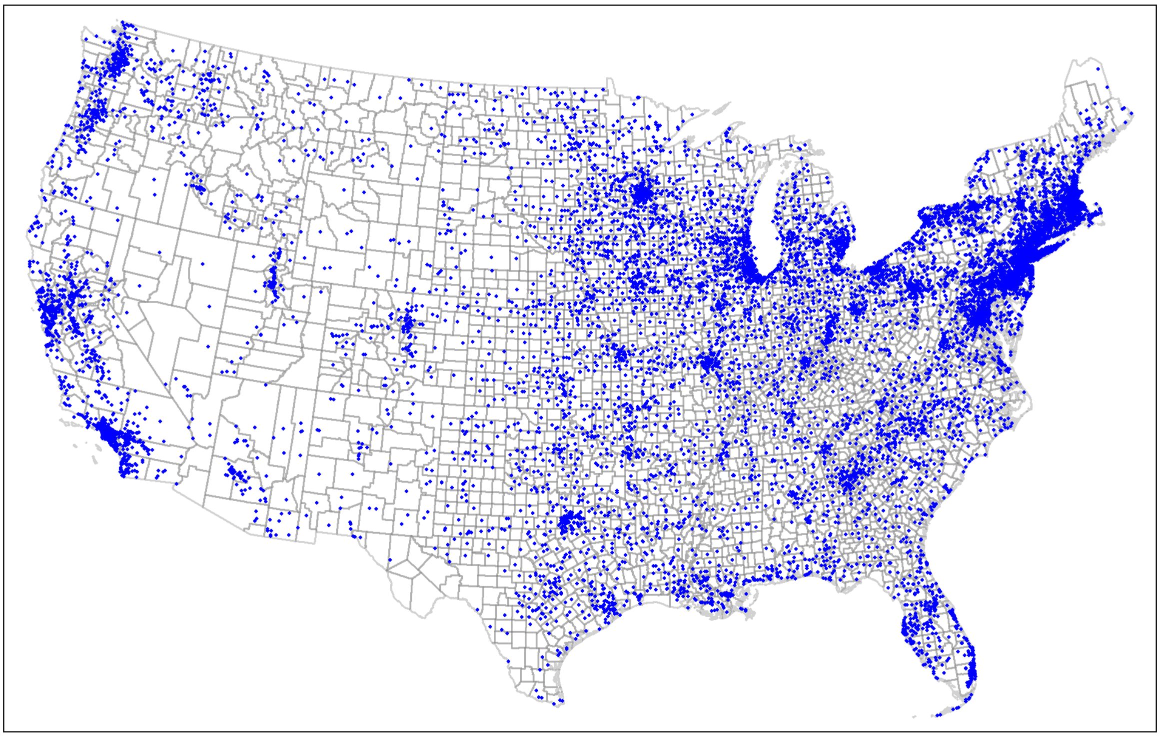
Notes: The map shows the net inflow of adult immigrants over the 1992-2016 period divided by the 1992 adult population. We split groups at the 25th, 50th, 75th, 90th, 95th and 99th percentiles.

Figure A.3 – Refugees in the United States by County, 1992-2016, Gross Inflows



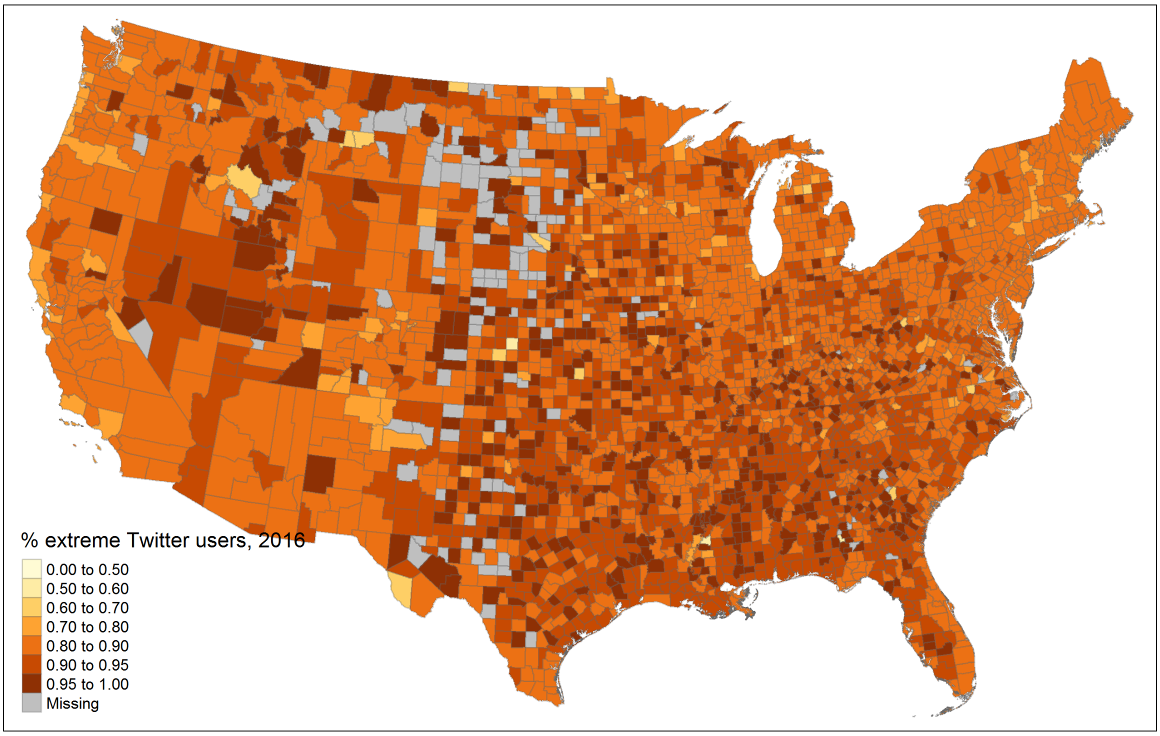
Notes: The map shows the gross inflow of refugees over the 1992-2016 period divided by the 1992 adult population. We split groups at the 75th, 90th, 95th and 99th percentiles.

Figure A.4 – Refugees in the United States by County, 1975-2008, Gross Inflows, Geocoded



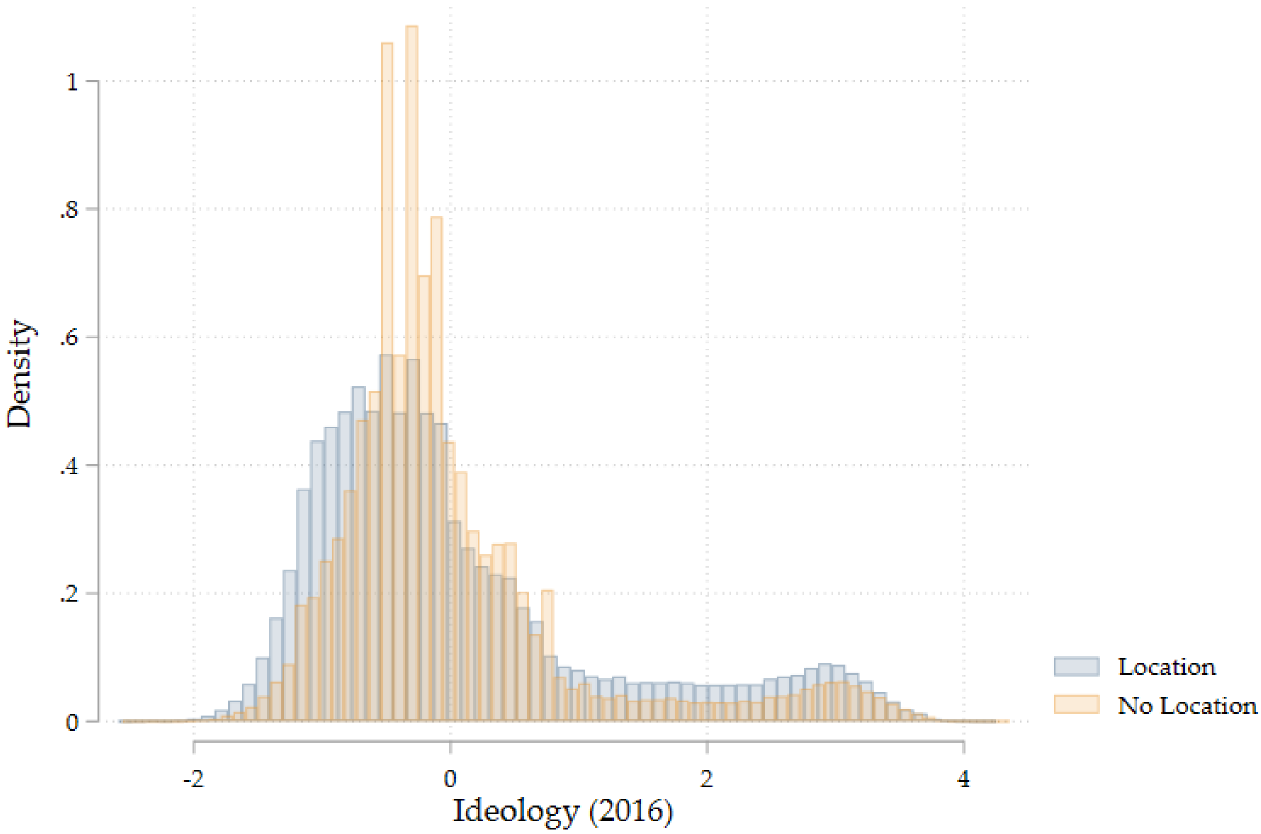
Notes: The map shows the location of first residence of refugees over the 1975-2008 period. We geocoded locations so that they depict a town, city or neighborhood (in large cities). One dot represents one location but can represent several refugees.

Figure A.5 – Twitter polarization 2016



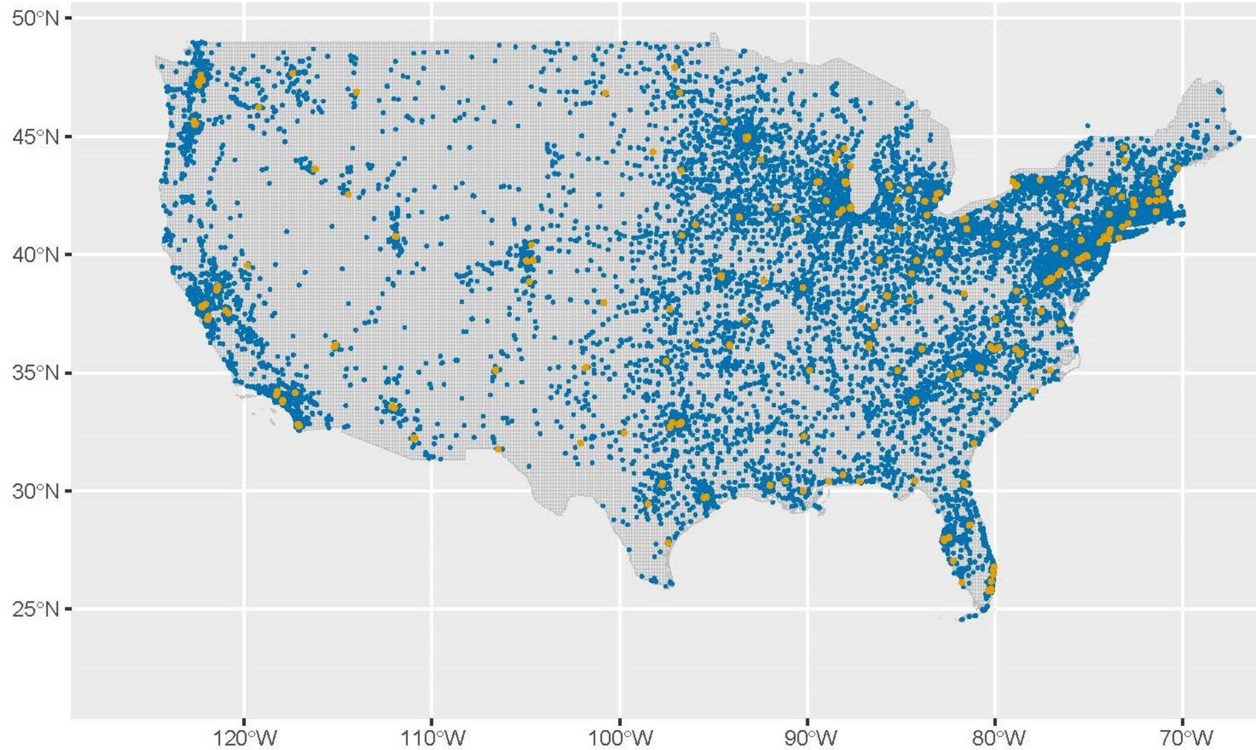
Notes: The map shows extreme Twitter users as a share of all Twitter users at the county level for the year 2016. We get ideology scores of Twitter users from Barberá (2015). We obtain left, right and moderate users by splitting the ideology score into terciles. The map is based on about 3 million Twitter users that provide their location.

Figure A.6 – Distribution of Twitter Accounts with Locational Information



Notes: The graph compares the distribution of ideology scores of Twitter user accounts with county information and accounts with no county information. The graph is based on the full sample of Twitter accounts and their ideology score from the year 2016 that we obtain from Barberá (2015).

Figure A.7 – Refugees and Refugee Resettlement Centers, 0.15° Grid Cells



Notes: The map shows the location of first residence of refugees over the 1975-2018 period in blue. The location of active refugee resettlement centers between 1990-2016 is shown in orange. We aggregate the information into 0.15°x0.15° grid cells shown in the background.

Table A.1 – Descriptive Statistics

	Obs.	Mean	SD	Min	Max
	Panel A: Immigrants and Refugees				
Δ Immigrants*	40023	623.7443	3376.97	-377.1992	126924.00
Δ Immigrant share*	40023	0.0035	0.01	-0.0276	0.12
Immigrant share IV*	40023	0.0024	0.01	-0.0818	0.22
Δ Immigrants (gross)	40023	846.4581	5411.54	0.0000	284252.00
Δ Immigrant share (gross)	40023	0.0051	0.01	0.0000	0.07
Immigrant share (gross) IV	40023	0.0039	0.01	0.0000	0.10
Δ Refugees	40023	44.1943	362.89	0.0000	24549.00
Δ Refugee share	40023	0.0001	0.00	0.0000	0.07
Refugee share IV	40023	0.0001	0.00	0.0000	0.06
	Panel B: Political Outcomes				
Extreme vs. moderate	40023	6.22	5.95	-16.05	17.67
Winner	39514	0.55	0.67	-2.54	2.02
Rep. vote share	40019	0.57	0.22	0.00	1.00
Winner if Rep.	27240	0.98	0.24	-0.90	2.02
Winner if Dem.	14666	-0.32	0.40	-2.54	1.30
Winner vs. loser	31618	1.58	0.56	0.00	5.77
Conservative Rep.	39624	0.20	0.39	0.00	1.00
Mod. Rep.	39624	0.17	0.37	0.00	1.00
Mod. Dem.	39624	0.14	0.34	0.00	1.00
Liberal Dem.	39624	0.49	0.49	0.00	1.00
Sh. Extreme Twitter	2,529	0.89	0.05	0.00	1.00
Sh. Right Twitter	2,529	0.78	0.10	0.36	1.00
Sh. Left Twitter	2,529	0.11	0.06	0.00	0.45
Sh. Moderate Twitter	2,529	0.11	0.05	0.00	0.37
	Panel B: Control Variables				
Δ Cultural Distance	39936	0.80	0.18	0.02	1.00
Δ Educational Distance	39955	0.49	0.23	0.01	1.00
Income*	40023	2.34	0.43	1.35	4.39
Share Afr.-American*	40023	0.10	0.12	0.00	0.65
Share urban*	40023	0.21	0.28	0.00	1.00
Unemployment*	40023	0.04	0.01	0.01	0.12
Share male*	40023	0.49	0.01	0.36	0.56
Share married*	40023	0.57	0.06	0.33	0.71
Import competition*	40023	0.06	0.06	0.00	1.12
Labor participation*	40023	0.63	0.05	0.40	0.84
Share low-skilled*	40023	0.17	0.07	0.04	0.46
Bartik share*	40023	0.01	0.01	0.00	0.13

Notes: We take parts of our data from Mayda et al.'s (2022) replication materials. Those variables are marked with an asterisk in the table. The sample is based on column 1 of Tables 1.1 and 1.2.

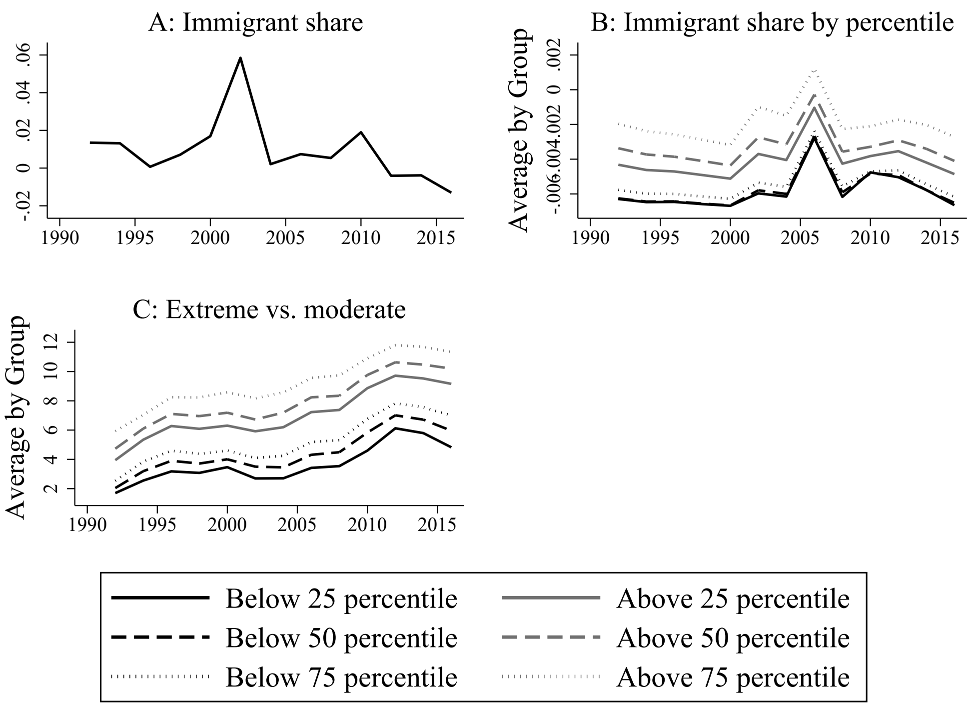
Table A.2 – Description and Sources

	Description	Source
	Panel A: Immigrants and Refugees	
Δ Immigrants (gross)	Change in the county stock of adult immigrants	Census, ACS
Δ Immigrant share (gross)	Change in the county stock of adult immigrants divided by county adult population	Census, ACS, Mayda et al.
Immigrant share (gross) IV	Sum of 1980 share of adult immigrants by country*net flow of immigrants by country divided by 1980 share of adult population*total population	Census, ACS, Mayda et al.
Δ Refugees	Number of new refugees	ORR, PRM
Δ Refugee share	Number of new refugees divided by county adult population	ORR, PRM, Mayda et al.
Refugee share IV	Sum of 1980-90 share of refugees by country*number of new refugees by country divided by 1980 share of adult population*total population	ORR, PRM, Mayda et al.
	Panel B: Political Outcomes	
Extreme vs. moderate	Inverse hyperbolic sine of the difference between extreme and moderate contributions (based on dollar-weighted terciles in 1990)	Bonica (2019)
Winner	Ideology of winner. Winner is the candidate receiving most votes in a county-district cell	EDS, Bonica (2019)
Rep. vote share	Republican vote share	EDS
Winner if Rep.	Ideology of Republican winners	EDS, Bonica (2019)
Winner if Dem.	Ideology of Democratic winners	EDS, Bonica (2019)
Winner vs. loser	Absolute ideological distance between winner and runner up	EDS, Bonica (2019)
Conservative Rep.	Dummy = 1 if winner is a Republican and right of 1990 party median	EDS, Bonica (2019)
Mod. Rep.	Dummy = 1 if winner is a Republican and left of 1990 party median	EDS, Bonica (2019)
Mod. Dem.	Dummy = 1 if winner is a Democrat and right of 1990 party median	EDS, Bonica (2019)
Liberal Dem.	Dummy = 1 if winner is a Democrat and left of 1990 party median	EDS, Bonica (2019)
Sh. Extreme Twitter	Share of right and left Twitter users in 2016. Thresholds for right, left, moderate users are obtained by splitting the 2012 ideology score into terciles.	Barberá (2015)
Sh. Right Twitter	Share of right Twitter users in 2016.	Barberá (2015)
Sh. Left Twitter	Share of left Twitter users in 2016.	Barberá (2015)
Sh. Moderate Twitter	Share of moderate Twitter users in 2016.	Barberá (2015)
	Panel C: Control Variables	
Δ Cultural Distance	Sum of the the absolute differences between the share of Latinos, Asians, Africans and Westerners among residents and new immigrants	Census, ACS, Mayda et al.
Δ Educational Distance	Sum of the the absolute differences between the share of high-school dropouts, high-school graduates, people with some college, college graduates and people with more than college among residents and new immigrants	Census, ACS, Mayda et al.

Notes: We take parts of our data from Mayda et al.'s (2022) replication materials (marked with an asterisk in Table A.1). ACS = American Community Survey, ORR = Office of Refugee Resettlement, EDS = Election Data Services, PRM = Bureau of Population, Refugees, and Migration.

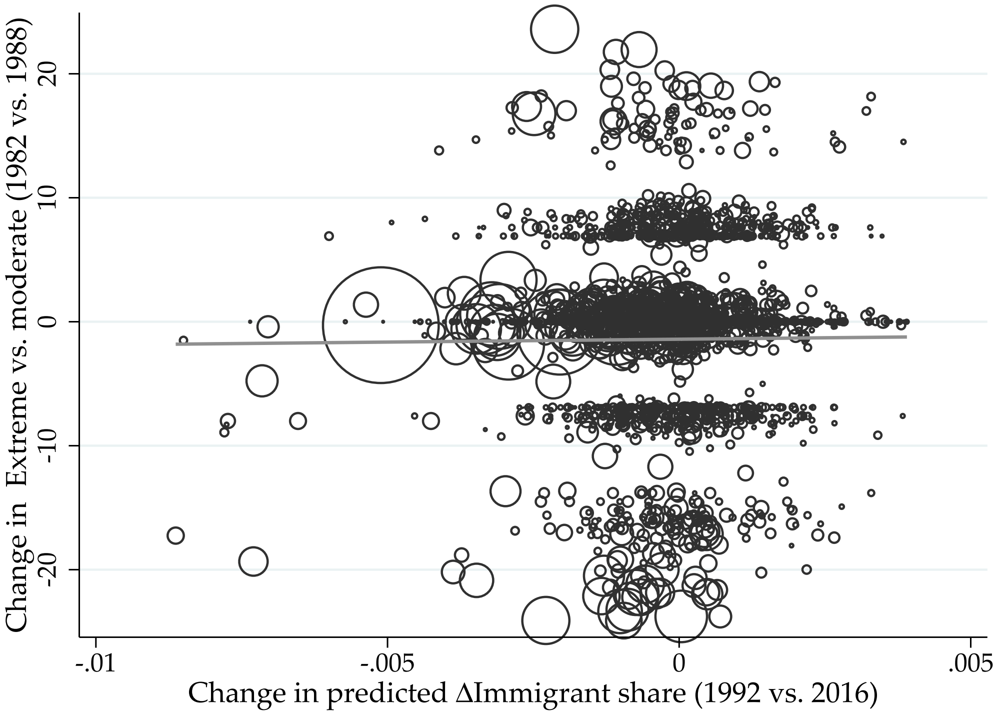
B Further Robustness

Figure B.1 − Parallel Trends—Immigrant Shares by Percentile



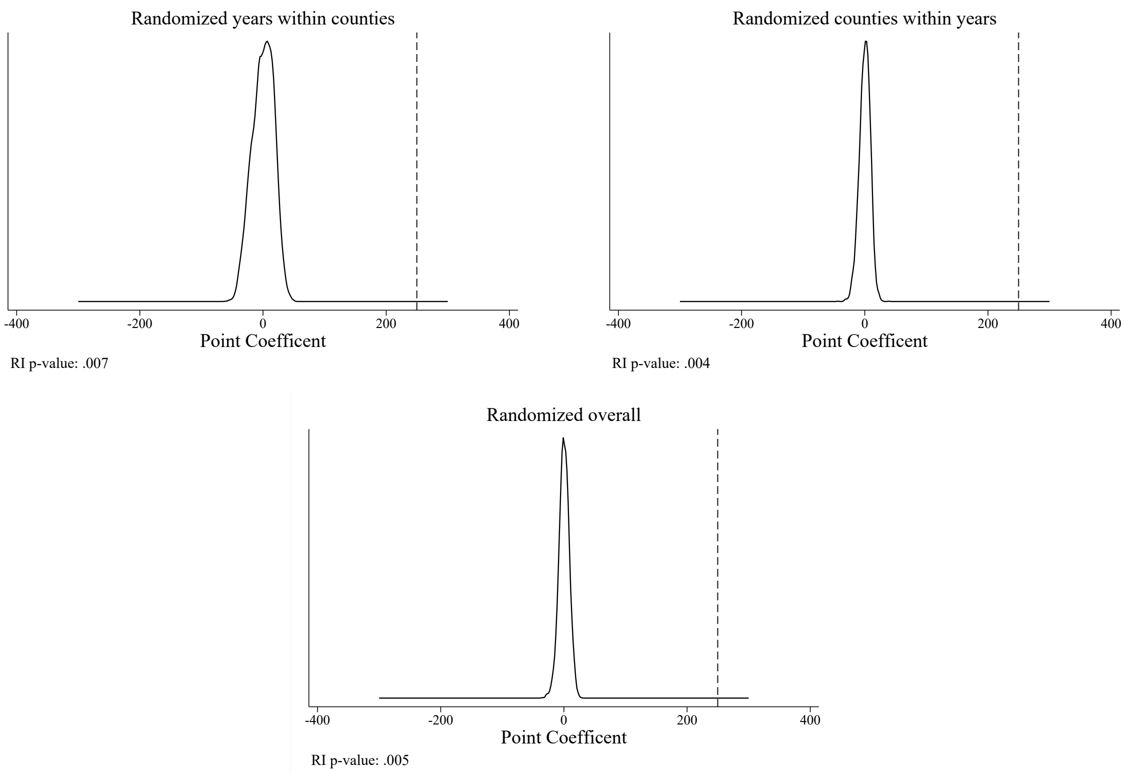
Notes: Panel A shows net inflows of adult immigrants as a share of the adult population. Panel B shows the same variable at the county-level, according to percentiles of the initial share of adult immigrants in the year 1980 (and netting out the effect of the control variables we include in all regressions). Panel C shows extreme versus moderate campaign contributions for the same percentiles.

Figure B.2 − Correlation Between Extreme vs. Moderate Contributions and Changes in Immigration



Notes: The figure shows the correlation between the change in net adult immigration (1992-2016) and the change in extreme vs. moderate campaign contributions (1982-1988). The straight grey line represents fitted values weighted by population, with a slope of 46.21 and standard error of 304.33.

Figure B.3 − Randomized Immigrants, Extreme vs. Moderate Contributions



Notes: The figures show results from regressions based on column 1 in Table 1.1. Each figure graphically represents the coefficients of 5,000 regressions, where we have randomized immigration shares (i) across years within the same county, (ii) across counties within the same year, and (iii) across space and time. The dashed vertical line shows the coefficient for net adult immigration from column 1 of Table 1.1. We calculate the randomization inference (RI) p-value as the proportion of times that the absolute value of the t-statistic in the simulated data exceeds the absolute value of the original t-statistic.

Table B.1 − Immigration and Polarization, 1992-2016, Two-year Net Inflows

	(1) Extreme vs. moderate	(2) Winner	(3) Rep. vote share	(4) Winner if Rep.	(5) Winner if Dem.	(6) Winner vs. loser	(7) Cons. Rep.	(8) Mod. Rep.	(9) Mod. Dem.	(10) Liberal Dem.
ΔImmigrant share	249.685***	55.130***	12.804***	14.488***	-8.667	30.963	23.880***	2.077	-9.840	-16.260***
	(81.515)	(14.404)	(3.805)	(4.388)	(7.850)	(23.902)	(4.727)	(8.177)	(9.958)	(5.507)
ΔIncome	-0.155	-0.030	0.024	-0.104**	0.147	-0.228	-0.114	0.173**	0.006	-0.066
	(1.061)	(0.132)	(0.037)	(0.047)	(0.105)	(0.205)	(0.069)	(0.079)	(0.086)	(0.068)
ΔShare Afr.-American	6.688	0.198	-0.257	-0.383	0.497	-1.558	-0.576	1.428**	-0.694	-0.118
	(11.348)	(1.153)	(0.285)	(0.686)	(1.167)	(2.397)	(0.743)	(0.705)	(0.540)	(0.641)
ΔShare urban	-1.017	-0.054	0.027	0.116*	-0.050	-0.041	0.054	-0.087	-0.062	0.095
	(1.080)	(0.171)	(0.039)	(0.058)	(0.114)	(0.138)	(0.103)	(0.101)	(0.070)	(0.086)
ΔUnemployment	-2.630	1.395	0.732	-1.196**	-0.819	-0.011	-0.720	2.824***	-0.636	-1.475*
	(8.333)	(1.105)	(0.447)	(0.534)	(1.125)	(1.855)	(0.617)	(0.898)	(0.797)	(0.817)
ΔShare male	24.391	-0.999	-0.865*	1.457**	-0.288	3.277	-0.108	-1.654	0.420	1.336*
	(16.328)	(1.498)	(0.433)	(0.625)	(0.915)	(2.010)	(1.202)	(1.115)	(0.874)	(0.676)
ΔShare married	-8.358**	-0.523	-0.071	-0.432**	0.189	-1.309*	-0.784***	0.607**	0.212	-0.038
	(3.662)	(0.451)	(0.137)	(0.177)	(0.431)	(0.774)	(0.268)	(0.260)	(0.248)	(0.266)
ΔImport competition	-9.630*	-0.542	-0.208*	0.514**	-0.038	0.357	-0.324	-0.180	0.325	0.171
	(4.876)	(0.362)	(0.118)	(0.212)	(0.478)	(0.771)	(0.346)	(0.226)	(0.293)	(0.190)
ΔLabor participation	18.391*	1.568	0.235	0.772*	-0.316	3.427	1.509**	-1.353*	0.218	-0.381
	(9.987)	(1.363)	(0.404)	(0.445)	(1.025)	(2.087)	(0.601)	(0.751)	(0.692)	(0.554)
ΔShare low-skilled	-10.063	-0.451	0.275	-0.186	-1.692*	-0.141	0.410	0.288	-0.825	0.143
	(8.944)	(0.725)	(0.265)	(0.266)	(0.954)	(1.834)	(0.579)	(0.576)	(0.543)	(0.401)
ΔBartik share	-13.745	-0.902	-0.574	-1.503	5.696***	-0.239	-1.799	0.459	0.508	0.806
	(18.406)	(1.813)	(0.528)	(1.239)	(1.437)	(2.300)	(1.209)	(1.247)	(1.290)	(1.087)
Observations	40,023	40,019	39,514	27,181	14,287	31,618	39,624	39,624	39,624	39,624
R-squared	-0.017	-0.063	-0.134	-0.085	0.007	-0.035	-0.043	0.006	-0.002	-0.032
K–P F-stat.	78.22	78.24	76.93	103.6	42.02	66.25	78.68	78.68	78.68	78.68

Notes: The table shows the second stages of 2SLS regressions; all regressions include population weights and fixed effects for counties and years; standard errors clustered at the state-level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.2 – Immigration, Ideology and Cultural Distance, 1992-2016, Two-year Net Inflows

	(1) Extreme vs. moderate	(2) Winner	(3) Rep. vote share	(4) Winner if Rep.	(5) Winner if Dem.	(6) Winner vs. loser	(7) Cons. Rep.	(8) Liberal Dem.
ΔImmigrant share	181.751	43.415***	11.152***	16.903***	-13.466	37.614**	20.567***	-9.854
	(143.861)	(13.827)	(3.274)	(6.401)	(10.010)	(16.713)	(4.177)	(9.122)
ΔImmigration * ΔCultural Dist.	145.097	26.852***	4.184**	-4.219	12.108**	-13.084	8.435*	-13.529**
	(116.131)	(9.208)	(2.022)	(4.762)	(5.282)	(15.809)	(4.950)	(5.299)
ΔCultural Distance	-0.425	-0.025	0.008	0.020	-0.053	0.068	0.014	0.044
	(0.602)	(0.057)	(0.013)	(0.024)	(0.047)	(0.071)	(0.030)	(0.033)
ΔIncome	-0.398	-0.080	0.014	-0.101	0.110	-0.209	-0.133*	-0.045
	(1.012)	(0.170)	(0.031)	(0.063)	(0.105)	(0.168)	(0.074)	(0.084)
ΔShare African-American	6.199	0.120	-0.266	-0.356	0.423	-1.487	-0.602	-0.072
	(12.867)	(1.238)	(0.259)	(0.561)	(1.010)	(1.789)	(0.785)	(0.724)
ΔShare urban	-1.040	-0.058	0.026	0.114**	-0.043	-0.039	0.053	0.098
	(1.058)	(0.169)	(0.034)	(0.049)	(0.090)	(0.169)	(0.074)	(0.092)
ΔUnemployment	-3.810	1.129	0.669**	-1.191***	-0.973	0.040	-0.829	-1.384**
	(9.244)	(1.089)	(0.272)	(0.387)	(1.153)	(1.756)	(0.631)	(0.695)
ΔShare male	26.671**	-0.437	-0.745**	1.429**	-0.038	3.088	0.114	1.130
	(11.682)	(1.396)	(0.315)	(0.588)	(1.139)	(2.204)	(0.840)	(0.767)
ΔShare married	-8.497**	-0.566	-0.081	-0.426***	0.217	-1.318**	-0.804***	-0.029
	(4.271)	(0.431)	(0.109)	(0.136)	(0.378)	(0.641)	(0.246)	(0.239)
ΔImport competition	-9.817*	-0.559	-0.208**	0.522**	-0.101	0.389	-0.324	0.189
	(5.012)	(0.480)	(0.100)	(0.205)	(0.319)	(0.578)	(0.310)	(0.235)
ΔLabor market participation	20.181**	1.944*	0.304	0.741*	0.037	3.290**	1.645***	-0.542
	(9.109)	(0.993)	(0.271)	(0.392)	(1.099)	(1.676)	(0.467)	(0.618)
ΔShare low-skilled	-10.848	-0.669	0.227	-0.181	-1.798**	-0.098	0.314	0.213
	(7.850)	(0.723)	(0.199)	(0.287)	(0.735)	(1.587)	(0.474)	(0.459)
ΔBartik share	-11.676	-0.398	-0.472	-1.535	5.958***	-0.376	-1.600*	0.609
	(10.735)	(1.590)	(0.336)	(1.071)	(1.283)	(2.992)	(0.954)	(1.029)
Observations	39,936	39,430	39,932	27,108	14,273	31,560	39,538	39,538
Kleibergen-Paap F	52.21	51.15	52.22	56.89	20.72	41.85	52.28	52.28

Notes: The table shows the second stages of Control Function Approach regressions, including the residual from the first-stage regressions; population weights and fixed effects for counties and years; bootstrapped standard errors clustered at the state-level in parentheses (500 repetitions); *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.3 – Immigration, Ideology and Educational Distance, 1992-2016, Two-year Net Inflows

	(1) Extreme vs. moderate	(2) Winner	(3) Rep. vote share	(4) Winner if Rep.	(5) Winner if Dem.	(6) Winner vs. loser	(7) Cons. Rep.	(8) Liberal Dem.
ΔImmigrant share	249.909**	56.269***	13.152***	13.884	-8.660	29.886	24.207***	-16.918*
	(120.949)	(14.925)	(2.836)	(13.092)	(9.578)	(18.335)	(4.469)	(9.500)
ΔImmigration * ΔEducational Dist.	11.190	-24.915***	-8.034***	5.512	-3.727	19.468*	-6.132*	15.653***
	(66.579)	(5.855)	(1.872)	(3.708)	(4.372)	(11.647)	(3.286)	(4.016)
ΔEducational Distance	0.935	0.420***	0.100***	0.035	-0.061	0.319*	0.194***	-0.141**
	(0.849)	(0.098)	(0.027)	(0.044)	(0.093)	(0.177)	(0.042)	(0.059)
ΔIncome	-0.267	-0.019	0.030	-0.118	0.165	-0.305*	-0.120*	-0.084
	(1.024)	(0.157)	(0.030)	(0.083)	(0.103)	(0.174)	(0.069)	(0.083)
ΔShare African-American	7.955	0.521	-0.198	-0.283	0.342	-1.052	-0.380	-0.168
	(12.521)	(1.098)	(0.240)	(0.529)	(0.965)	(1.769)	(0.746)	(0.667)
ΔShare urban	-1.071	-0.077	0.022	0.115**	-0.048	-0.061	0.043	0.103
	(1.047)	(0.155)	(0.031)	(0.055)	(0.097)	(0.159)	(0.071)	(0.086)
ΔUnemployment	-1.940	1.653	0.790***	-1.142**	-0.870	0.360	-0.593	-1.552**
	(8.885)	(1.050)	(0.255)	(0.464)	(1.062)	(1.685)	(0.611)	(0.670)
ΔShare male	23.006**	-1.187	-0.880***	1.331**	-0.092	2.415	-0.268	1.307*
	(10.645)	(1.292)	(0.290)	(0.610)	(0.965)	(1.940)	(0.818)	(0.720)
ΔShare married	-7.938*	-0.462	-0.063	-0.394**	0.151	-1.091*	-0.735***	-0.026
	(4.131)	(0.414)	(0.102)	(0.163)	(0.384)	(0.611)	(0.242)	(0.235)
ΔImport competition	-9.717*	-0.500	-0.194**	0.502**	0.008	0.271	-0.323	0.140
	(5.019)	(0.449)	(0.096)	(0.230)	(0.353)	(0.568)	(0.300)	(0.225)
ΔLabor market participation	17.732**	1.409	0.211	0.724*	-0.208	3.094**	1.412***	-0.363
	(8.385)	(0.866)	(0.244)	(0.379)	(0.928)	(1.458)	(0.433)	(0.568)
ΔShare low-skilled	-9.973	-0.234	0.338*	-0.196	-1.709***	-0.161	0.476	0.034
	(7.574)	(0.683)	(0.183)	(0.327)	(0.641)	(1.422)	(0.453)	(0.438)
ΔBartik share	-11.804	-0.401	-0.481	-1.326	5.618***	0.471	-1.503	0.722
	(11.813)	(1.573)	(0.312)	(1.055)	(1.326)	(3.062)	(0.949)	(1.025)
Observations	39,955	39,449	39,951	27,124	14,279	31,574	39,557	39,557
Kleibergen-Paap F	86.04	84.39	86.04	101.6	36.46	68.53	86.72	86.72

Notes: The table shows the second stages of Control Function Approach regressions, including the residual from the first-stage regressions; all regressions include population weights and fixed effects for counties and years; bootstrapped standard errors clustered at the state-level in parentheses (500 repetitions); *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.4 – Immigration and Polarization, 1992-2016, Eight-year Net Inflows

	(1) Extreme vs. moderate	(2) Winner	(3) Rep. vote share	(4) Winner if Rep.	(5) Winner if Dem.	(6) Winner vs. loser	(7) Cons. Rep.	(8) Liberal Dem.
ΔImmigrant share	4.430	29.599***	7.255***	3.524***	2.962	18.103**	9.193***	-11.388***
	(73.814)	(6.936)	(1.643)	(1.216)	(2.640)	(8.416)	(2.291)	(3.038)
ΔIncome	3.272*	-0.085	0.010	-0.031	-0.116	-0.136	-0.083	-0.021
	(1.899)	(0.175)	(0.054)	(0.072)	(0.123)	(0.265)	(0.085)	(0.106)
ΔShare African-American	-8.573	-1.020	-0.099	-1.473***	-0.965	-2.278	-1.288	-0.672
	(11.801)	(1.456)	(0.473)	(0.453)	(1.366)	(2.671)	(1.248)	(0.854)
ΔShare urban	-0.828	0.068	0.015	0.054	0.040	0.045	-0.006	-0.027
	(0.516)	(0.109)	(0.025)	(0.033)	(0.093)	(0.108)	(0.044)	(0.052)
ΔUnemployment	29.897	-2.848	0.562	-1.777***	-3.985**	-4.973*	-2.242*	2.090*
	(21.567)	(1.876)	(0.615)	(0.649)	(1.501)	(2.472)	(1.273)	(1.142)
ΔShare male	12.179	3.773	0.661	1.177	-1.015	7.653**	1.874	-0.395
	(11.273)	(3.454)	(1.101)	(0.820)	(2.093)	(3.409)	(1.464)	(1.574)
ΔShare married	-15.352***	-0.914	-0.301	-0.460	-0.763	-1.618	-0.577	0.357
	(5.342)	(0.866)	(0.265)	(0.275)	(0.821)	(1.179)	(0.531)	(0.466)
ΔImport competition	-4.125*	-0.439	-0.173	0.136	-0.441	-0.908	-0.053	0.318*
	(2.420)	(0.492)	(0.118)	(0.139)	(0.518)	(0.931)	(0.385)	(0.172)
ΔLabor market participation	-8.554	3.900	0.978	0.706	1.954	4.644*	1.674*	-1.303
	(13.091)	(2.638)	(0.644)	(0.454)	(1.395)	(2.629)	(0.869)	(1.361)
ΔShare low-skilled	-4.636	-2.231	-0.432	0.090	-1.922	-1.750	0.255	0.997
	(13.469)	(2.229)	(0.548)	(0.402)	(1.624)	(2.455)	(1.043)	(0.840)
ΔBartik share	-5.509	-3.841*	-0.643	-0.279	6.298***	-0.177	-3.927***	1.779
	(25.575)	(1.989)	(0.458)	(1.033)	(1.878)	(2.947)	(1.078)	(1.357)
Observations	9,236	9,138	9,235	5,898	2,408	6,226	9,161	9,161
R-squared	0.020	-0.706	-0.477	-0.040	0.008	-0.140	-0.113	-0.322
Kleibergen-Paap F	18.25	18.16	18.25	15.85	8.744	11.43	18.13	18.13

Notes: The table shows the second stages of 2SLS regressions; all regressions include population weights and fixed effects for counties and years; standard errors clustered at the state-level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.5 – Pre-trends, Shift-Share Instrument

	(1)	obs.		(2)	obs.
Extreme vs. moderate	4.86e-06	36,916	Income	0.000857	36,940
	(1.56e-05)			(0.000548)	
Winner	-0.000172	32,680	Afr.-American	0.00779	36,940
	(0.000169)			(0.0237)	
Rep. vote share	0.000605	36,916	Share urban	0.000265	36,940
	(0.000708)			(0.000832)	
Winner if Rep.	0.00131*	13,772	Unemployment	-0.0259	36,940
	(0.000663)			(0.0216)	
Winner if Dem.	-0.00154	18,908	Share male	-0.0948	36,940
	(0.00110)			(0.0801)	
Winner vs. loser	0.000731	25,950	Share married	0.0139	36,940
	(0.000453)			(0.0197)	
Conservative Rep.	-0.000220	34,840	Import competition	0.00303	36,940
	(0.000301)			(0.00548)	
Mod. Rep.	2.54e-06	34,972	Labor participation	0.0156	36,940
	(0.000245)			(0.00939)	
Mod. Dem.	0.000414	34,840	Share low-skilled	-0.000339	36,940
	(0.000268)			(0.00337)	
Liberal Dem.	0.000135	34,972	Share white low-skilled	0.00511	36,940
	(0.000158)			(0.00413)	
			Share of white male low-skilled	0.0340	36,940
				(0.0216)	

Notes: We define the pre-trend variables as the difference between 1982 and 1988 for column 1 and changes between 1980 and 1990 for column 2, while the dependent variable is the two-year difference of the shift-share instrument in the 1992-2016 period. All specifications include the same control variables as in Table B.1, year-fixed effects (we omit county-fixed effects) and population weights. Each line represents a separate regression with the variables listed as the explanatory variables of interest. Standard errors clustered at the state-level in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

This chapter is single-authored

2

Fueling Divisions? The Arrival of Fast Internet in Indian Villages

2.1 Introduction

Uniting people from different cultures, religions, languages, and ethnicities behind a broader sense of "we" is a fundamental principle of nation-states (Anderson, 1983; Putnam, 2007; Bazzi et al., 2019). Who is encompassed in the "imagined political community" that represents a nation is an ongoing and re-emerging struggle in many countries such as China (Uighurs), Turkey (Kurds), United Kingdom (Scots), India (Muslims) or Myanmar (Rohingya). Information technologies have a central function in spreading a common narrative to create this imagined community where citizens, even without personal interactions, feel a sense of connection. The internet can create shared experiences and connects people as if in a national, or even *global village* (McLuhan and Powers, 1989; Depetris-Chauvin et al., 2020). At the same time, it has the potential to divide by providing a platform for emotionally appealing populists, echo chambers, and misinformation (e.g., Campante et al., 2018; Vosoughi et al., 2018; Levy, 2021). Overall, the *aggregate* impact of internet access on unity in diverse communities remains unclear.

Rural areas have only sparsely received information on politics and social debates via traditional media sources (Correa et al., 1997). National controversies have thus hardly trickled down. Not anymore. Rural areas in developing countries, home to 3.1 billion of the world population, experience a dramatic change in the information environment as fast internet connects them to the nation (ITU, 2020; World Bank, 2021). The leap-frog technology has suddenly exposed a considerable share of inexperienced, mostly uneducated media users to highly sophisticated content creators. A very different experience to most study subjects who are: well-educated, urban, media-experienced individuals and who learned to navigate the increasingly sophisticated online world early.[1] The consequences of connecting the periphery to the core remain to be understood.

This paper focuses on India, which provides a unique setting to causally assess the impact of fast internet on its over 900 million people in rural areas (World Bank, 2021). The largest country in the world is home to a sizeable Muslim minority, making up 14.2% (170 million) of the population according to the 2011 census. Their peaceful coexistence has once more come under threat as the Hindu nationalist Bharatiya Janata Party (BJP) around Prime Minister Narendra Modi rose to power in 2014 (New York Times, 2019). Hate messages and misinformation have spread quickly via social media leading to lynch mobs targeting Muslims, which have forced social media giants like WhatsApp to restrict the viral potential of Indian messages (Time, 2019). Rural India is increasingly exposed to these messages as internet penetration more than doubled within four years from 12% in 2015 to 30% in 2019 (TRAI, 2018, 2019; World Bank, 2021). The sudden exposure of diverse rural areas to the national online discourse surrounding India's identity creates an interesting setting.

In this paper, I study the impact of fast internet on religious divisions in rural communities in India. Specifically, I examine the consequences of broadband connections established in 2017 on local conflict (2018-2022, village council level), the allocation of scarce welfare benefits to Muslims by local officials (2019-2022, individual level in Jharkhand), and voting behavior along religious lines (2019 national election, polling station level in Jharkhand). The study focuses on rural areas at the periphery, which get connected to the core. I introduce new data on the location and roll-out of 175,157 broadband connections as part of the largest rural government broadband initiative (called BharatNet) (Zimmermann, 2014). To identify the causal impact of fast internet, I exploit spatial discontinuities in internet availability that arose in 2017 due to the staggered roll-out of BharatNet. The study of several dimensions elicits the influence on different groups in the population: only extreme individuals resort to conflict, while favoritism by local officials in the allocation of welfare benefits is more common, and lastly, changes in voting behavior paint a picture of the general population. In sum, I aim to further our understanding of the internet's impact on divisions in rural communities through the lens of different dimensions, which I discuss in more detail below.

[1] The difference does not stop at the internet user, the safeguards within the internet are different as well. According to internal documents, Facebook invests 87% of its time resources on fact-checking content in the US, where only 10% of its user base resides (New York Times, 2021). In addition, Facebook has only recently improved its automatic hate speech detection algorithms in Hindi and Bengali - two of the main languages spoken in India (its largest user market).

To identify the causal impact of the internet, I introduce a new identification strategy based on spatial discontinuities in internet availability. The distribution of internet access is highly endogenous since wealthy, densely populated, and urban places are more profitable and thus connected first. I isolate exogenous variation in internet availability by exploiting discontinuities created by BharatNet. BharatNet is a flagship initiative of the Indian government that aims to connect every Gram Panchayat (GP or village council) to the fiber optic network.[2] It is designed to provide all households in a village with internet speeds of 2-20Mbps. Due to capacity constraints, the roll-out was split into two phases. 100,000 GPs were allocated to phase I (connected between 2014-2017), 150,000 to phase II (ongoing as of July 2023). The allocation took place at the block level (third highest administrative level) and was determined by minimizing the length of additional optical fiber that needed to be installed in phase I (Satyanarayana et al., 2015).[3] I obtained data on the exact location of a GP's point of connection to the fiber optic network, as well as in which phase. To isolate exogenous variation in internet availability, I exploit discontinuities in fast internet between neighboring villages on two sides of the phase boundary in a spatial regression discontinuity design (RDD). Further specifications combine these spatial jumps in internet availability with individual-level data differentiating between Muslims and non-Muslims. In particular, they leverage discontinuities in the differential treatment of Muslims in villages on both sides of the boundary in a difference-in-discontinuity framework. In a final step, I compare discontinuities in the differential treatment of Muslims in villages with a non-Muslim GP president to villages with a Muslim GP president.[4] Following Gelman and Imbens (2019), all specifications estimate a local linear RD polynomial based on a sample of GPs, polling stations or individuals located within a small distance of 10km from a boundary as in Dell and Olken (2020). The results are robust to alternative bandwidths, alternative weights, and a quadratic RD polynomial. I check the plausibility of the identifying assumptions by testing for discontinuities in a large number of variables, and outcomes pre-treatment and at a placebo boundary. In sum, combining this novel source of internet data and the unique variation in local internet access creates new avenues for the identification of the consequence of fast internet in a large developing country.

I discuss the impact of fast internet on several dimensions of divisions one by one in more detail. I start by examining the causal influence of fast internet availability on violent conflict. The effectiveness of divisive messages transmitted via radio or movies to create violence against the targeted community has been documented in several contexts (e.g., see Wang, 2021; Ang, 2023; Esposito et al., 2023 for the U.S.; DellaVigna et al., 2014 for Croatia; Adena et al., 2015 for Nazi Germany and Yanagizawa-Drott, 2014 for

[2] A GP consists of 2.5 villages on average and is the lowest level of government.

[3] The roll-out of BharatNet is combined with the set-up of public Wifi hotspots which create extremely local discontinuities in internet access.

[4] The GP president is the elected head of the village council.

Rwanda). Others have shown that uniting messages and shared experiences can lower tensions (e.g., see Blouin and Mukand, 2019 in Rwanda; Armand et al., 2023 in the U.S. and Depetris-Chauvin et al., 2020 in Africa). The impact of messages transmitted online are less well understood but points in a similar direction (e.g., an increase in hate crimes is found in response to Trump's tweets by Müller et al., 2022 and, Russian social media by Bursztyn et al., 2019). In contrast to the literature on traditional media, Bursztyn et al. (2019) attribute the impact to two channels: i) easier coordination and ii) a change in the information set. Altogether, these studies highlight the power of narratives spread via the media but also point to the importance of content. Naturally, this raises questions about the aggregate impact on rural communities.

The endogenous selection of content by the users makes the aggregate impact on uneducated, media-inexperienced villagers unclear. In contrast to the previous literature, this paper studies the impact on the periphery as it gets closer to the core: do villagers unite behind being Indian as the perceived distance to the rest of the nation shrinks or does the internet divide them? The individuals potentially exposed to inflammatory religious content share a local community with members from different religions since generations. These type of frequent personal interactions can lower stereotypes and foster mutual understanding (as highlighted in a different context by Bazzi et al., 2019) and may thus make villagers resilient to divisive online messages. To further our understanding of the overall impact on rural communities, I assess the internet's impact on assaults, as well as riots and mobs by actors related to one of the two main parties: the Hindu nationalist BJP and the secular Indian National Congress (INC). I obtain data on 38,078 assaults over the 2008-2022 period from GDELT, and data on riots and mobs by supporters of the Hindu nationalist BJP and the secular INC from ACLED. I obtain information on 1,052 riots and 914 mobs by BJP supporters, and 382 riots and 281 mobs by INC supporters over the 2016-2022 period. I focus on low-intensity conflict events in order to capture escalating disagreements in local communities that often respond to a trigger (such as heightened divisions online). The findings based on the spatial RDD show a significant increase in the level and change of assaults and an increase in riots and mobs by BJP supporters but not by INC supporters both in the full sample and in the state of Jharkhand (which will be the focus of further results).[5] As in most moderate conflict settings, only a small set of villages is affected. The findings are therefore driven by a small number of extreme individuals who are willing to resort to violence and inflict catastrophic damage. Nevertheless, this can have broader consequences as local conflict can undermine trust and create fear in a larger set of people.

[5] Weidmann (2016) has argued that media availability is related to measurement error in conflict data. Although I cannot rule out the influence of media bias, the differential impact on BJP mobs and riots with the absence of any increase in INC mobs and riots makes it unlikely that the effects can fully be explained by media bias. Moreover, I do not find any increase in peaceful protests (results are available upon request).

These findings serve as a motivation to assess the widespread consequences of rural internet on religious divisions. In particular, I estimate the causal impact of fast internet on the differential allocation of scarce welfare benefits to Muslims by GP presidents (*Sarpanch*). A large literature shows widespread discrimination by individuals that have the power to punish or allocate resources in the name of the public (e.g., Hodler and Raschky, 2014; Burgess et al., 2015; Goncalves and Mello, 2021). It remains unclear to what extent the internet influences the allocation decisions of local officials, but Grosjean et al. (2023) document the amplification of discrimination of law enforcement officers through narratives spread via Trump campaign rallies. I study the allocation of paid work days within the largest rural welfare program in the world (Zimmermann, 2014), the National Rural Employment Guarantee Act (NREGA). NREGA is designed as a social protection mechanism for vulnerable rural households, which have the right to a minimum of 100 days of work for the public at a fixed rate of pay, whereby the distribution of the work days falls within the responsibility of the GP. The program has suffered from supply constraints such that the demand for work cannot be matched, opening the door to discrimination and favoritism by the GP president (Dutta et al., 2012).[6] As the potential consequences are serious and affect a large share of the population, the impact of the internet on more moderate expressions of group divisions is important.

The assessment of NREGA has several advantages. It allows a focus on Muslims, the largest religious minority, which has often been the target of hate. It is a program that large shares of the rural population rely on. Measurement error and spillovers are not likely. Finally, it can contribute to the isolation of an information channel.[7] I estimate a difference-in-discontinuity design to analyze whether the internet changes the treatment of Muslims relative to non-Muslims by GP presidents within NREGA. The outcome is based on the allocation of over 300 million NREGA work days between Muslims and non-Muslims. The information originates from over 5 million websites, which I webscraped and that document the universe of registered households (since 2006) and work days (2019-2022) within NREGA in Jharkhand.[8] I focus on a single state as the collection of the data is time-consuming and depends linearly on the number of registered households.[9] I link the work days to the respective GP president responsible for the allocation. To assess differential treatment by religion, I classify individuals and GP presidents as Muslim or

[6] Jeong et al. (2023) have documented personal favoritism of village heads in NREGA in the context of Uttarakhand.

[7] The literature found it difficult to isolate an information channel as two-way communication also enables better coordination of assaults, protests, or mobs. The preferred solution was therefore to assess only isolated information treatments online, which disguise the overall impact.

[8] In Jharkhand, a large share of households are registered within NREGA. In the median village, the cumulative sum of households that signed up between 2006 and 2022 is 1.16 times the number of households in the 2011 census.

[9] The analysis focuses on Jharkhand due to the following characteristics: it has a sizeable Muslim minority of 14.5%, similar to the Indian average, and it has a large BharatNet-phase boundary distributed across several areas in the state.

non-Muslim based on their names. This allows me to obtain a unique perspective on the treatment of a minority by public officials at the individual level. The discontinuity in internet availability results in a sudden decrease in the number of new registrations of Muslims in NREGA and a significant decrease in the number of work days by 10.9% in villages with registered Muslims. Significantly fewer work days are allocated to Muslims in areas with internet compared to neighboring areas without internet, but only in villages with a non-Muslim president. The sign reverses in areas with a Muslim president, where significantly more work days are allocated to Muslims. Together, this is in line with the GP increasingly allocating scarce welfare benefits based on religious group identities.

Finally, I turn to political divisions created by the internet differentiating between polling stations in Muslim and non-Muslim villages. Identity-based voting can have detrimental consequences (e.g., Banerjee and Pande, 2007). The results discussed so far have highlighted an increase in discrimination and conflict brought about by the internet. Increasing discrimination and violence can result in increases in group-based voting as promised benefits are allocated along group lines (e.g., Carlson, 2015; Hadzic et al., 2020). At the same time, the content consumed online can have a direct influence on voting behavior (see Zhuravskaya et al., 2020 for a review). In the U.S. and Europe, studies show that populists have benefited especially from the internet and voters were exposed to a high degree of online misinformation (Mocanu et al., 2015; Campante et al., 2018; Grinberg et al., 2019; Guriev et al., 2021). Assessments of the spread of misinformation find that especially old (inexperienced) users re-share political misinformation (Guess et al., 2019). This makes an impact on political divisions involving many voters new to the internet seem likely.

I test the political implications of access to fast internet in the context of Jharkhand. Therefore, I assemble the 2019 national election results for 29,464 polling stations in Jharkhand. Although Muslims are a sizeable minority in India, there is no Muslim party that pools their votes behind them. Rather secular parties, most importantly the INC, are the main opponents to the rising movement of Hindu nationalist parties led by the BJP. The evolution of political polarization in India is fundamentally shaped by the dichotomy between secularism and Hindu nationalism (Sahoo, 2020). Therefore, I assess the political consequences along these lines in addition to assessing the changes in the differential support of Muslim candidates by Muslim villages. Overall, villages with internet shift towards Hindu nationalist parties and away from secular parties in non-Muslim villages. The shift is absent in villages with a high Muslim share, where they increasingly favor secular parties and Muslim candidates. These patterns are in line with a higher resilience of Muslim villages against Hindu nationalist messages, which can be explained by more common interactions between members of different groups.

So far, I have attributed the exacerbation of divisions to content exposure online. Alternatively, conflicts could arise if the potential economic impact of the internet favors

Muslims. Mitra and Ray (2014) have documented an increase in violence against Muslims after they fared better economically relative to other groups. If Muslims benefited disproportionately from internet access in villages with a non-Muslim president and benefited less in villages with a Muslim president, then the patterns shown could be attributed to an economic mechanism. I test the plausibility of this alternative mechanism based on DHS data (2019-2021) covering over 2 million individuals (and over 280,000 Muslims). The difference-in-discontinuity estimates do not show any differential increase in wealth for Muslims and no significant difference in the economic status of Muslims in GPs with and without a Muslim president. This alleviates concerns that the increase in divisions is driven by the internet's economic impact. In contrast, poor Muslims seem to fare worse—a pattern consistent with an increase in the discrimination of Muslims in general (and within NREGA) due to the exposure to new information online.

This paper makes several contributions to the literature. First, it documents the divisive influence of fast internet on rural areas in a large developing country. The internet has transformed developing economies, creating new jobs (Hjort and Poulsen, 2019) and increasing consumption (Bahia et al., 2020).[10] These economic effects have been accompanied by changes in political dynamics. While mobile internet has been shown to coordinate protests (Acemoglu et al., 2018; Manacorda and Tesei, 2020), decrease government approval (Guriev et al., 2021) and crowd out offline interactions of politicians (Bessone et al., 2022), I highlight the implications for communal relations in rural communities. The consequences for extreme individuals, local representatives, and the general population have not been well understood.

Second, I further our understanding of group divisions. A large literature has investigated shifting identities and the determinants of tensions between groups in diverse nation-states (e.g., Laitin, 1998; McGuirk and Burke, 2020; Bazzi and Gudgeon, 2021; Gehring, 2022). Economic rewards and ethnic polarization increase the likelihood of conflict, while a common enemy brings groups closer together. Shared experiences and personal interactions can form a common identity and build trust among groups (Bazzi et al., 2019; Depetris-Chauvin et al., 2020; Lowe, 2021). This paper highlights the internet's influence as it spreads into poor rural areas. It, therefore, documents the influence of a structural change in the information set on divisions. It shows that tensions at the core can spill over fueling conflict in rural communities even though personal interactions and personal knowledge are likely high.

In the following section, I give an overview of recent changes in the political discourse in India regarding Muslims offline and online. I present the data and outcomes in section (3.3), followed by the empirical strategy in section (3.4). The results are described in section (3.5), an alternative mechanism is discussed in section (2.6), while section (3.7)

[10] A related literature on mobile phones has similarly documented widespread gains (for a review see Aker and Mbiti (2010).

concludes.

2.2 Background

Many nation-states struggle to define their relation to religion (e.g., Turkey, Iran, or Afghanistan). India, a state with a large minority of 170 million Muslims, is a prime example. The rise of the BJP in recent years has ignited discussions about the secular fabric of India. A party deeply rooted in Hindu nationalist ideology, the BJP has sought to redefine India's identity along religious lines.[11] As a consequence, the notion that Indians are defined based on their place of birth is more frequently replaced with a definition based on Hindu culture (Sahoo, 2020). Secularism, a principle enshrined in the Indian constitution, guarantees equal rights and freedom to all religions, thereby separating religion from the state's matters. Following a period dominated by the secular Indian National Congress (INC) governing India for 54 years since independence, the rise of the BJP has challenged this principle, leading to tensions within India's diverse population.

These developments have been linked to increases in religious violence and hate crimes (New York Times, 2019). The rise of the BJP has made divisive and inflammatory language common among high-ranking government officials. A case in point are debates around killings (of mostly Muslim herders) by Hindu mobs in the name of cow protection.[12] The chief minister of Chhattisgarh (Raman Singh, BJP) proclaimed in 2017, for example: "We will hang those who kill cows" and a BJP lawmaker in Uttar Pradesh (Vikram Saini) stated a month earlier: "I had promised that I will break the hands and legs of those who do not consider cows their mother and kill them" (Human Rights Watch, 2019, p. 5). The Wire (2019) analyzed 34 campaign speeches of Uttar Pradesh's chief minister (Yogi Adityanath, BJP) and found over 100 instances of hate speech and religious polarization.

These inflammatory messages have reached a rapidly increasing rural audience. Previously isolated villages have rapidly adopted digital technologies. Rural India has gained 53 million new internet users every year since 2017, totaling 399 million in 2022 (Kantar, 2021, 2023). Internet consumption skyrocketed to 17 GB per day and user in 2021 according to the India Mobile Broadband Index, as an increasing network was combined with the worldwide fifth lowest prices per Gigabyte (0.17$ in 2022 according to cable.co.uk) (Nokia, 2022). The new users are uneducated, inexperienced and not informed about

[11] The party's leader, Narendra Modi, faced international criticism for his role in the 2002 Gujarat riots during his tenure as the state's Chief Minister. In 2005, he was denied a diplomatic visa to the U.S. on the grounds of "severe violations of religious freedom." This incident spotlights the BJP's stance towards religious minorities.

[12] The cow has a sacred status under Hinduism. A Human Rights Watch Report documents 44 killings in the name of cow protection between 2015 and 2018.

politics (CSDS, 2022).[13] They are confronted with a political discourse that is dominated by the BJP. Narendra Modi has been labeled India's first "social media Prime minister" in 2014 (Financial Times, 2014) and won the first "WhatsApp election" in 2019 (Financial Times, 2019). His party outspent its main rival the INC by a factor of 15 on social media in 2019 (Hindustan Times, 2019) and relies on a network of millions of volunteers who spread the BJPs messages in customized WhatsApp groups throughout the country (Time, 2019). This has likely impacted the online discourse. Leaked internal documents from Facebook document widespread misinformation including hate messages against Muslims and an internal memo from Facebook employees identifies "misinfo that are connected to real world harm, specifically politics and religious tensions" as the main request from users in India (New York Times, 2021). In sum, the inexperienced villagers are confronted with an information environment that combines state-of-the-art online communication strategies with lax data protection rules during a time when religious tensions are high on a platform with limited tools to delete hate messages.

2.3 Data and Outcomes

Rural areas in the developing world have only very recently gained access to fast internet. I leverage the following data sources to capture emerging divisions in rural communities.

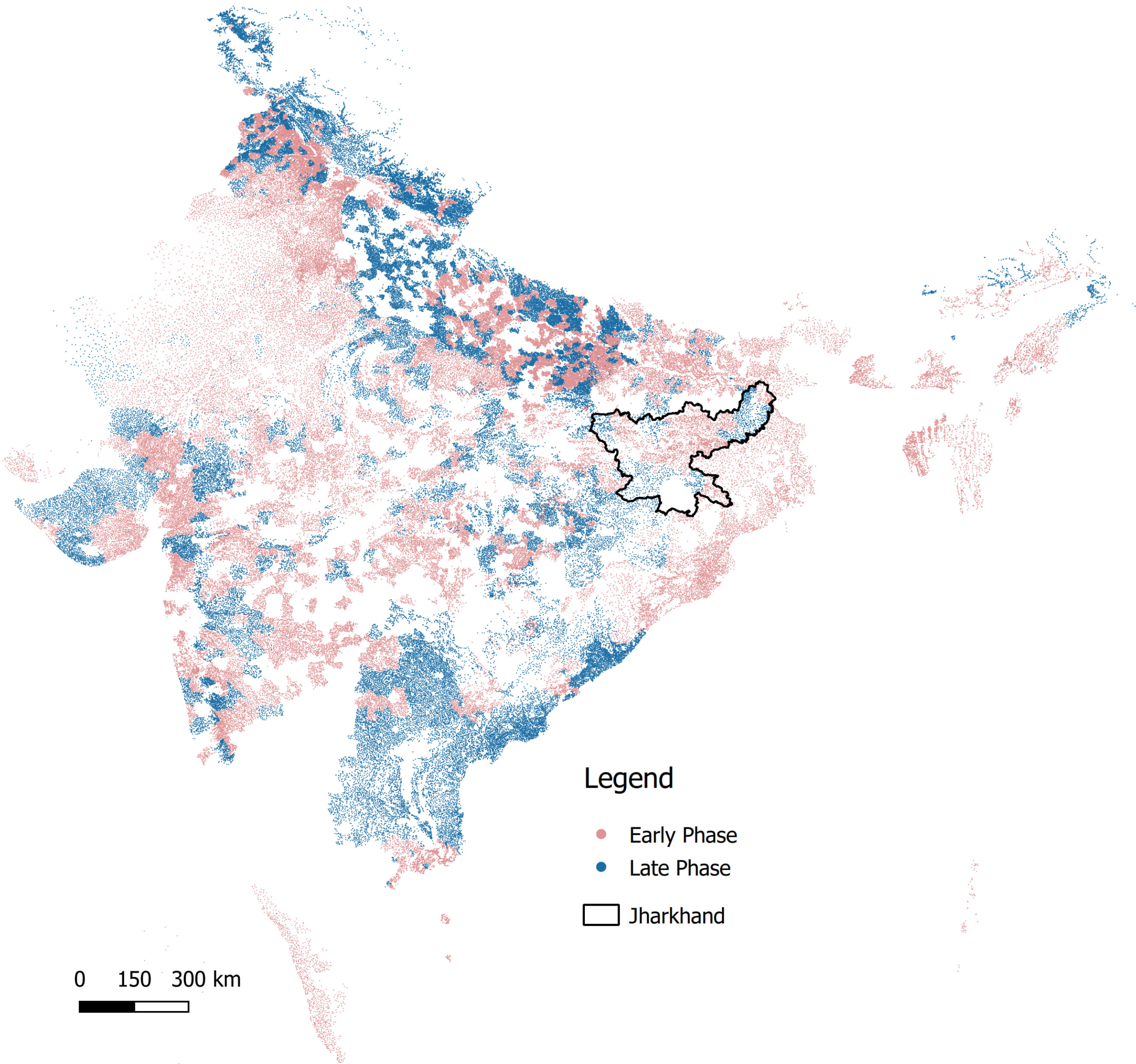
Internet Data: The Indian government proposed in 2011 to integrate its lowest level of government (Gram Panchayats) into its fiber optic infrastructure. This initiative now known under the name BharatNet, is proclaimed to be the largest rural government broadband connectivity program in the world. It aims to connect all 250,000 GPs in India to fiber optic internet.

Bharatnet is a central pillow of the National Telecom Policy of 2012 that aims to provide all households the opportunity of a broadband connection between 2 Mbps and 20 Mbps on demand (Satyanarayana et al., 2015). Telecommunication providers and other companies can use the fiber optic infrastructure at highly subsidized rates. Major telecommunication providers have made use of the broadband infrastructure and more than 100,000 Wifi hotspots have been installed (Economic Times, 2017; Ministry of Communications, 2021).

The connection of GPs to broadband internet is implemented in a staggered manner due to capacity constraints. In phase I, which was completed in December 2017, 100,000

[13] Via their phone, they consume religious, political, and ethnic information in specialized chatgroups that spread national cleavages throughout the country (CSDS, 2022). Another issue has been violent acts based on misinformation spreading online. The BBC analyzed English-language media reports and identified a rapid increase in fatal mob attacks triggered by rumors originating from WhatsApp. While they found 0 in 2015 and 2016 the number increased to 31 in 2017 and 2018.

Figure 2.1 – BharatNet GPs by Phase



Notes: Each dot on the map represents a GP that gets connected to broadband internet via the government's rural broadband initiative BharatNet. Salmon dots denote GPs in phase I that got connected between 2014 and 2017. Blue dots denote GPs in phase II which is still ongoing (as of 2023). I make these data available at https://sites.google.com/view/johannes-matzat/data.

GPs were connected (Krishnan, 2018). The connection of the remaining 150,000 GPs is still ongoing as of July 2023.

I obtained lists of GPs by phase from Bharat Broadband Network Limited including the exact location where the GP is connected to the broadband internet. The raw data include 208,512 connections for GPs located in 37 states and union territories.[14] Information is sparse for Karnataka, Tamil Nadu and Goa.

[14]Some GPs are listed in both phases. I follow the instruction of Bharat Broadband Network Limited and delete duplicate GPs from phase II.

I cross-check the locations of the GPs by automated searches for a given GP on onefivenine—a village repository that provides local information about villages including their location. I update the location information if the location is unique in the repository and the sum of the absolute difference between the coordinates is larger than 0.05 (5.5km at the equator). In a second step, I manually check all GPs that were located outside of their district or were more than 1 degree (about 111 km at the equator) away from the mean location of the other villages in the block. The final dataset consists of 175,157 GPs—94,023 in phase I and 81,134 in phase II—as illustrated in Figure (2.1).

Internet Usage: I collect data on internet usage, in order to confirm that the rural broadband program led to an increase in internet consumption. Measuring internet usage directly is inherently difficult due to a lack of fine-grained internet usage data, which is partially why I collect and rely on variation in internet availability in the first place.[15] Nevertheless, I use a feature in the Facebook Marketing API provided by Meta to commercial users for targeted advertising. Specifically, Meta shows a commercial user the number of active users on Meta's social media platforms in the last month for parameters (including location) that can be specified. Therefore, I feed the location of each village to the API and set all other parameters to the most general values to retrieve the number of active Facebook and Instagram users between 13 and 65 years that live (or have recently lived) in the location in 2020.[16] A limitation of that approach is that Meta has set the minimum value it reports to 1,000 users to protect the privacy of its users; after 1,000 the number of users is reported in steps of 100. Since 1,000 active users is a prohibitively high number for many villages, I use a workaround to improve the accuracy of the number of users. I request the number of users in an Indian GP and a specific town in the US (which approximately has 1,000 active users) jointly. This lowers the de facto minimum number of reported users to 0. Since the Marketing API restricts the shape of the area to a circle, I approximate the area of a GP as follows: I ask for the number of users within a 2km radius which captures space slightly smaller than the average GP.[17]

The Meta data have the advantage of highlighting variation in internet usage that exposes individuals likely to national political debates around religion. However, their relative coarseness makes slightly stronger assumptions for identification necessary. Therefore, I supplement the data with information on internet availability in all schools of Jharkhand in 2019.[18] The data originate from the District Information System for Ed-

[15] Studies evaluating the internet's impact in developing countries at a large scale have relied on mobile coverage data based on the GSM Association's data which is not available in India or temporal variation due to connections via undersea cables (e.g., Hjort and Poulsen, 2019; Guriev et al., 2021).

[16] API calls were made between October and December 2020.

[17] The average GP in Jharkhand covers an area of 18.39km^2 while I collect Facebook activity for an area of 12.57km^2.

[18] The data on schools was obtained from http://schoolgisjharkhand.nic.in/education and last accessed on April 4, 2023. Although information on internet availability in schools exists country-wide, their

ucation (DISE) and include the exact location of all schools. In total, there are 45,782 schools, 29.61% reporting internet access. The high spatial granularity in combination with the indirect measure of usage, complement the strengths and weaknesses of the Meta data.

Conflict: To capture the most severe form of group divisions, I obtain data on assaults over the 2008-2022 period from the GDELT 1.0 Event Database (Leetaru and Schrodt, 2013). Assaults include physical and sexual assaults, destruction of property, torture, and death by physical assault. Assaults committed by the state (i.e., the police or the military) are excluded. The data highlight a drastic increase in assaults over time. While there are 7,050 recorded assaults in the 3 years before the start of BharatNet (2011-2013), there are 11,994 in the three years after phase I was completed (2017-2019). I complement the data with mob and riot events by supporters of the two main parties: the Hindu nationalist BJP and the secular INC from ACLED (Raleigh et al., 2010). The data only start in 2016 and record 1,195 mobs and 1,434 riots over the sample period (2016-2022). Based on the data I construct the following outcomes: the natural logarithm of one plus the number of assaults, riots, or mobs (by party) within 1km around the broadband connection over the 2018-2022 period.

Social Welfare: To capture subtle but relevant consequences of changing group divisions, I obtain data on the largest public employment program worldwide: NREGA (Zimmermann, 2014). Its primary goal is the "social protection of the most vulnerable people" in rural India by guaranteeing every household a minimum of 100 days of wage employment per fiscal year (Ministry of Rural Development, 2014, p.1). The program is implemented at the local level, where the GP is responsible for generating enough public works, registering households, and allocating work. The program is known to face supply issues such that demand for work days exceeds supply in all states as highlighted in Dutta et al. (2012). Assessing excess demand based on the National Sample Survey 2009-10, they find that 51,7% of rural households in Jharkhand wanted to work in NREGA, of which 62.8% did not get any work.[19] Excess demand leaves the allocation of work among villagers at the discretion of the GP, opening the door to favoritism. The presence of which is highlighted by Jeong et al. (2023) in the state of Uttarakhand, where GP presidents that barely won the election assign themselves three times more work days than those that barely lost.

latitude and longitude is only accessible in Jharkhand.

[19] Such high levels of excess demand are characteristic of India's poor states where local and regional supply constraints bind. Although the main fiscal costs are provided by the national level, the supply, organization, and allocation of work and workers require skilled labor at the local and regional levels. While no state meets the demand for public employment within NREGA, others report more modest supply shortages resulting in a national average of 44.4% excess demand (Dutta et al., 2012).

To obtain public employment via NREGA, members of a household in a given village can register for a job card within their GP. A job card, in turn, enables members of the household to request public employment. To foster transparency, job cards are in the public record online. They include the name of the household head, the village and GP, the date of registration, and list the days on which public employment was provided. I webscrape 5.41 million websites, including the universe of job cards in every village of Jharkhand.[20] 57.86% percent (3.13 million) of job card holders received public employment between 2019 and 2022, totaling 324 million days of employment.

To assess changes in group divisions, I capture the differential treatment of Muslims in NREGA. Therefore, knowledge of the religious affiliation of individuals seeking employment is necessary. This type of information is scarce in Indian data and only available at large aggregates - the district - in the Indian census. Therefore, similar to Ash et al. (2021), I infer the religion based on the first and last names of villagers. I employ machine learning trained on over 41 million land records in Bihar, which Jharkhand used to be part of until the year 2000. The algorithm has been shown to predict whether an individual is Muslim with an accuracy of over 97% on unseen names in the context of Bihar (Chintalapati et al. 2022).[21] In total, 648,068 of the households (11.97%) registered in NREGA are Muslim.[22] 88% of GPs have at least one registered Muslim and they make up more than 10% of registered individuals in 33% of GPs. To further test whether the differential treatment of Muslims in NREGA is driven by deepening discrimination, I exploit the group affiliation of the GP president. Strengthening group identities along religious lines would lead to Muslims being disadvantaged in villages governed by a non-Muslim GP president and advantaged in those governed by a Muslim president. To test this hypothesis, I webscrape the GP president's name at the very end of their 5-year term in 2022 for all villages in Jharkhand reporting it online—3,928 out of 4,345.[23] Again, I use the machine learning algorithm to classify presidents into Muslim and non-Muslim. Overall, 1,427 villages (4.41%) are governed by a Muslim president.

Finally, in order to locate the GPs I fuzzy merge them based on their name and the block and district information to the SHRUG village shapefile (Asher et al., 2021). I am able to identify 89.4% of the locations. I then calculate the minimum distance of the centroid of the GP to the next phase boundary segment.

I construct the following outcomes: the share of Muslims that registered in a year for

[20] Webscraping a large number of websites can be time-consuming due to computational constraints on both ends (the hosting server, as well as the CPU of the scraping machine). I webscraped the data between January and May 2023.

[21] The algorithm was developed by Rajashekar Chintalapati, Aaditya Dar, and Gaurav Sood and can be accessed via the pranaam package in Python.

[22] This is close to 14.53%—the Muslim share reported in the 2011 census for Jharkhand.

[23] The information was obtained from https://gpdp.nic.in/PPC/sarpanchWithDetailsReport.html, last accessed on April 21, 2023. I obtain the information based on the last year in office as reporting in previous years was low. The election of that term took place between September and December 2017 and thus mostly before phase I was completed.

NREGA, as well as the natural logarithm of one plus the number of work days allocated to a specific household. This enables me to capture the allocation of work days to Muslims relative to others within NREGA, differentiating between villages governed by a Muslim and villages governed by a non-Muslim GP president.

Voting: Next, I obtain voting data of the 2019 national election for Jharkhand at the polling station level.[24] Overall, there are 29,464 polling stations documenting the choices of 14.8 million voters. There are 57 parties most of which are small. The three most successful parties were the BJP with over 7.5 million, the INC with more than 2.3 million, and the Jharkhand Mukti Morcha (JMM) with over 1.7 million votes. Form-20 reports votes by candidate and polling station and includes a unique polling station ID. It does not report the name or location of a polling station, however.[25] In order to link each polling station to a location, I identify the name of each polling station through the Chief Electoral Officer in Jharkhand after verifying that the IDs did not change between the national and state elections in Jharkhand that both took place in 2019.[26] To assign a location based on a polling station name, I exploit that the majority of polling stations are situated in school buildings. I recover the precise coordinates of 17,031 polling stations via a fuzzy match with the District Information System for Education (DISE) - a data set containing every school and its coordinates.[27] In order to capture the political consequences of rural internet access which can accelerate group divisions, I assess voting behavior at polling stations in villages based on the share of Muslims present. I use the share of registered Muslims in NREGA since 2006 as a proxy. I group votes as follows: First, shift in votes towards Hindu nationalist parties.[28] Second, the share of votes received by secular parties. Third, the share of votes received by Muslim candidates.[29] Finally, I present the vote shares received by the two main national parties: the Hindu nationalist BJP and the secular INC. Thereby, I aim to capture emerging divisions in voting behavior along religious lines.

[24] The data have been accessed via Form-20 of the Election Commission on October 7, 2019.

[25] I tested several data sets containing the ID as well as latitude and longitude for polling stations that are based on https://gis1.jharkhand.gov.in/Election_GIS, but have found that the quality of the coordinates was too low.

[26] Source: https://ceojh.jharkhand.gov.in/mrollpdf1/aceng.aspx.

[27] The data of schools have been accessed via http://schoolgisjharkhand.nic.in/education on April 4, 2023. In order to match the building name of the polling station, I first assign weights to each word based on the term frequency-inverse document frequency (tf-idf)—a technique commonly used in text analysis. Thereby, I aim to increase the quality of the match by assigning low weights to words that are common across observations (like *school*, *primary* or *secondary*) and high weights to words that are rare across observations and thus particular informative (usually the name of the village). Then I use fuzzy matching based on a nearest neighbor algorithm.

[28] Hindu nationalist parties in the 2019 general election in Jharkhand: BJP; secular parties: INC, All India Trinamool Congress, Communist Party of India, Communist Party of India (Marxist, Liberation), Communist Party of India (Marxist-Leninist, Red Star).

[29] Again, I have used the same procedure as described above to assign a religion to each candidate running based on their name.

2.4 Empirical Strategy

Identification: To further our understanding of the causal impact of the internet on the cohabitation of religious communities in India, several challenges to identification need to be addressed. In particular, internet providers usually roll-out their services gradually under demand and supply side considerations. On the demand side, more dense and wealthy areas are connected first. On the supply side, right-of-way, distance, and already available infrastructure are taken into consideration. Therefore, a simple comparison between areas with and without internet access usually implies a comparison of very different individuals. In contrast, I assess variation in rural internet availability due to the staggered roll-out of BharatNet, a large rural broadband connectivity program by the Indian government. The program divided villages into early and late receivers based on supply-side considerations at the block level (third administrative unit) as stated by Satyanarayana et al. (2015) p.25:

> "In Phase I, the Blocks to be connected were selected based on the least length of incremental optical fibre to be laid."[30]

To the extent as supply and demand side factors are uncorrelated, this mitigates concerns related to vast differences in wealth, education, and population density between early- and late-treated areas. Nevertheless, early-treated areas are more populated and educated (see Table A.2 in the appendix). The roll-out was combined with the set-up of over 100,000 Wifi hotspots which makes the internet readily accessible and creates highly localized variation in internet access. I isolate quasi-exogenous variation in fast internet by exploiting the discontinuous change in internet availability at the phase boundary. I use a range of approaches to examine the impact of fast internet on group divisions, which I describe one by one below.

2.4.1 RDD

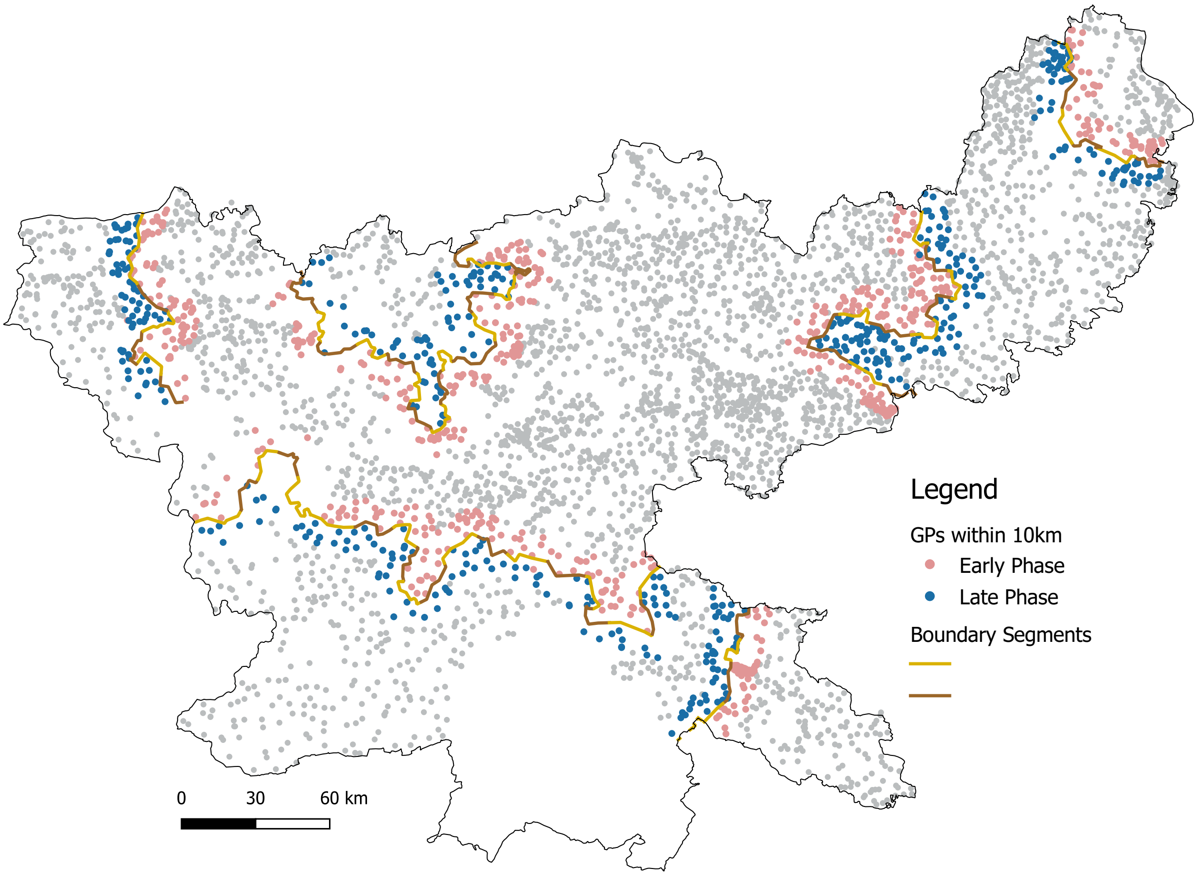
To eliminate concerns regarding the correlation of supply and demand side factors, I exploit the discontinuous change of internet availability at the boundary in a spatial RDD:

[30]Incremental optical fiber refers to the length of connections that need to be built to connect the existing optical fiber network between cities to the GPs within a block.

$$y_v = \alpha + \beta_1 phase_v + f(dist_v, phase_v) + g(long_v, lat_v) + \sum_{s=1}^{1,614} seg_{iv}^s + e_v, \tag{2.1}$$

where y denotes one of several outcomes capturing divisions at the GP or polling stations level v. $phase$ is a binary variable that takes on the value one if the GP received a broadband connection in the first phase and is zero if it will receive a broadband connection in the second phase. I follow Gelman and Imbens (2019) and account for smooth changes in the outcomes by including a local linear polynomial that is estimated separately for both phases. In addition, $g(long_v, lat_v)$ is a local polynomial that controls

Figure 2.2 – RDD Visualization, Jharkhand



Notes: Each dot on the map represents a GP in Jharkhand that gets connected to broadband internet via the government's rural broadband initiative BharatNet. Colored dots are within 10km of the boundary. Salmon dots denote GPs in phase I that got connected between 2014 and 2017. Blue dots denote GPs in phase II which is still ongoing (as of 2023). The boundary is divided into 20km segments as highlighted by the two different colors.

for the location of the GP in the two-dimensional latitude-longitude space. In order to exploit local discontinuities, I divide the boundary into 1,614 segments of 20km length (100 segments in the case of Jharkhand). This ensures that I only compare neighboring GPs located on either side of the same segment. To avoid comparisons across state boundaries, I add state-fixed effects in the full sample and I apply a small bandwidth of 10km as in Dell and Olken (2020) (and similar to Dell et al., 2018; Lowes and Montero, 2021; Méndez and Van Patten, 2022). β_1 captures then the causal impact of fast internet if divisions would have evolved smoothly in the absence of BharatNet. Section (2.4.4) assesses the plausibility of this assumption.

2.4.2 Difference-in-Discontinuity

Several outcomes allow me to differentiate between Muslims and non-Muslims. This enables an analysis of the impact of fast internet on Muslims relative to non-Muslims. In particular, I estimate the following model in a difference-in-discontinuity framework (similar to Grembi et al., 2016 and Bluhm and Pinkovskiy, 2021):

$$\begin{aligned} y_{iv} = \alpha + \beta_1 phase_v + \beta_2 phase_v \times muslim_{iv} + \beta_3 muslim_{iv} + \\ f(dist_v, phase_v) + g(long_v, lat_v) + \sum_{s=1}^{100} seg_v^s + e_{iv}, \end{aligned} \tag{2.2}$$

where y_{iv} is the natural logarithm of one plus the number of work days received by individual i living in GP v, $phase_v$ is a binary variable that denotes whether the individual lives in a GP part of the first roll-out phase, whereby $muslim_{iv}$ identifies Muslims. Consequently, β_1 captures the average impact of broadband internet for non-Muslims, whereby $\beta_1 + \beta_2$ estimate the influence for Muslims. My coefficient of interest is then β_2 capturing any change in the differential treatment of Muslims (e.g., in NREGA) due to internet availability. As in model (2.1), $f(dist_v, phase_v)$ is a local linear polynomial that is estimated separately for both phases and $g(long_v, lat_v)$ is a local polynomial that controls for the location of the GP in two-dimensional latitude-longitude space. 20km boundary-segment-fixed effects restrict the comparisons to close-by villages. Observations are weighted using triangular kernel weights. The model is estimated within a small bandwidth of 10km. The causal interpretation of β_2 requires slightly different assumptions compared to model (2.1). Namely, the absence of other discontinuities that affect Muslims and non-Muslims differently and a constant difference in outcomes between Muslims and non-Muslims. Thus, the difference-in-discontinuity design accounts for factors that affect Muslims and non-Muslims in the same way. This mitigates concerns regarding compound treatment effects due to the overlap with a low-level administrative boundary.

2.4.3 Difference-in-Difference-in-Discontinuity

The analysis of the allocation of NREGA benefits allows me to assess how fast internet changes the treatment of Muslims relative to non-Muslims depending on the religion of the GP president. Therefore, I estimate the difference-in-difference-in-discontinuity between GPs with a Muslim president and a non-Muslim president. I adapt model (2.2) as follows:

$$
\begin{aligned}
y_{iv} = \alpha + \beta_1 phase_v + \beta_2 phase_v \times muslim_{iv} + \beta_3 phase_v \times muslim_{iv} \times muslim\_pres_v + \\
\beta_4 muslim_{iv} \times muslim\_pres_v + \beta_5 muslim_{iv} + \beta_6 muslim\_pres_v + \\
f(dist_v, phase_v) + g(long_v, lat_v) + \sum_{s=1}^{100} seg_v^s + e_{iv},
\end{aligned}
\tag{2.3}
$$

where $muslim\_pres_v$ identifies a Muslim GP president and all other variables are defined as in model (2.2). The main identifying assumption changes accordingly. To interpret β_3 causally, in the absence of BharatNet, there should be no discontinuity in the difference-in-difference in outcomes between Muslims and others in villages with a Muslim president and without a Muslim president. If the assumptions are valid, then β_3 captures whether Muslim GP presidents change their treatment of Muslims due to internet availability differently relative to non-Muslim GP presidents.

2.4.4 Plausibility of Identifying Assumptions

The central assumptions behind models (2.1), (2.2), and (2.3) build on each other. While model (2.1) requires the absence of all discontinuities at the boundary that impact the outcome, model (2.2) and model (2.3) require the absence of discontinuities at the boundary that affect the difference in outcomes between Muslims and non-Muslims (or the difference in difference for model (2.3)). Models (2.2) and (2.3) thus do not require the absence of discontinuities that affect all individuals equally. However, any test of model (2.1) increases the plausibility of models (2.2) and (2.3).

Absence of other discontinuities: I test for discontinuities in a range of variables at the village level, which I obtained from SHRUG (Asher et al., 2021). These include 2001 and 2011 census data on total population, number of households, size of the scheduled caste population, size of the scheduled tribal population, size of the literate population, number of primary, middle, secondary, and senior secondary schools, and number of colleges—in addition to, total nightlights, average nightlights, electricity, rural consumption, as well as the information on the timing and length of rural road upgrades. The results reported in Tables (2.1) and (B.2) do not show any evidence for discontinuities for

Jharkhand and none for the full Indian sample except for the growth rate of the scheduled tribe population (significant at the 5% level) and the number of colleges (significant at the 10% level). This is to be expected by chance when testing for discontinuities in 72 different variables in total. It is therefore no evidence for a systematic deviation from the continuity assumption.[31]

Table 2.1 – Test for other Discontinuities - Census

	Panel A: India			*Panel B: Jharkhand*		
	(1) 2001	(2) 2011	(3) 2011-2001	(4) 2001	(5) 2011	(6) 2011-2001
Total Population	-64.889	-83.201	-13.352	-511.283	-801.588	-172.453
	(55.720)	(65.448)	(12.867)	(612.645)	(702.171)	(115.820)
No. of Households	-10.907	-14.838	-2.917	-92.519	-158.546	-43.192
	(10.082)	(13.020)	(3.392)	(112.877)	(143.854)	(33.613)
Literate Population	-34.608	-48.769	-10.851	-367.713	-529.999	-94.456
	(36.262)	(46.134)	(11.788)	(447.453)	(527.233)	(96.150)
Scheduled Caste Population	-10.154	-15.603	-4.567	-26.110	-82.953	-29.869
	(8.451)	(10.284)	(2.975)	(47.178)	(67.264)	(22.929)
Scheduled Tribe Population	1.193	-2.994	-3.828**	70.809	70.248	-0.281
	(4.578)	(5.780)	(1.851)	(59.707)	(73.110)	(18.501)
No. of Primary Schools	-0.023	-0.052	-0.026	-0.254	-0.263	0.059
	(0.025)	(0.035)	(0.025)	(0.180)	(0.335)	(0.220)
No. of Middle Schools	0.007	0.002	-0.003	-0.117	-0.436	-0.304
	(0.014)	(0.025)	(0.020)	(0.183)	(0.367)	(0.217)
No. of Secondary Schools	-0.001	0.008	0.010	-0.072	-0.105	-0.031
	(0.008)	(0.016)	(0.013)	(0.091)	(0.113)	(0.071)
No. of Sr. Secondary Schools	0.004	0.015	0.011	-0.026	-0.058	-0.028
	(0.006)	(0.012)	(0.009)	(0.020)	(0.039)	(0.033)
No. of Colleges	0.002	-0.008	-0.009*	0.004	-0.012	-0.007
	(0.002)	(0.006)	(0.005)	(0.020)	(0.013)	(0.013)
Observations	50,476	50,586	50,476	1,193	1,201	1,193

Notes: This table tests for pre-existing discontinuities at the boundary. It estimates the main coefficient of interest (β_1) of model (2.1) for a number of outcomes from the 2011 and 2001 censuses. The model is estimated at the GP level, restricted to locations less than 10km away from the boundary. Panel A reports coefficients for the full sample and includes 20km segment- and state-fixed effects. Panel B is restricted to Jharkhand and includes 20km segment-fixed effects. Observations are weighted using triangular kernel weights. $^*p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$

Absence of pre-treatment discontinuities: I test whether there are any detectable differences in the outcomes before the start of BharatNet. This can be seen as a test for

[31] Nevertheless, it does present a random deviation. Table (B.3) in the appendix highlights that the results do not depend on it.

the presence of any pre-existing discontinuities that potentially affect the outcome. Limited data availability in the years before BharatNet restrict this test to conflict outcomes based on GDELT data and the registration of NREGA households. Columns 1 and 2 of Table (2.2) show the results and do not highlight any discontinuities.[32]

Falsification (Placebo Boundary): I construct a plausible placebo, not across time, but across space. This allows testing for the absence of discontinuities for outcomes that are not available before the start of BharatNet. Furthermore, it serves as a test of the additional assumptions in models (2.2) and (2.3).

A simplistic approach would just shift the boundary by a given amount or use random block boundaries as a placebo. This would ignore the non-random selection of blocks, however. Therefore, to construct a realistic counterfactual, I leverage information on the selection process of blocks into the treatment.

BharatNet will connect all GPs to the fiber optic network but split the project into two phases. It selected blocks, containing in sum 100,000 GPs, for phase I based on minimizing the length of additional cables that need to be built. The selection process can thus be visualized in two steps: first, all blocks are sorted based on the length of additional cables that need to be built, and second, the number of GPs is summed up across blocks moving from the block with the shortest additional cable need to the longest until 100,000 GPs are reached. That is then where the cut-off is, which determines the geographic phase boundary used in the spatial RDD. Now, imagine a thought experiment in which we move further up the sorted sequence of blocks, changing the arbitrary number of GPs that determine the cut-off to 175,000 (see Figure (2.3) for an illustration). Like any other number, this would then generate a plausible counterfactual boundary.

Since I lack information on the need for additional cables by block, I have to approximate this thought experiment. First, I use all 2011 census variables listed in Table (2.1) to estimate a block's propensity to be in phase I. Second, I split all district boundaries dividing phase II areas that do not intersect with state boundaries into 20km segments and assign the block on each segment with the higher predicted propensity to the placebo treatment.[33] I then test for discontinuities across the placebo boundary using models (2.1), (2.2), and (2.3).

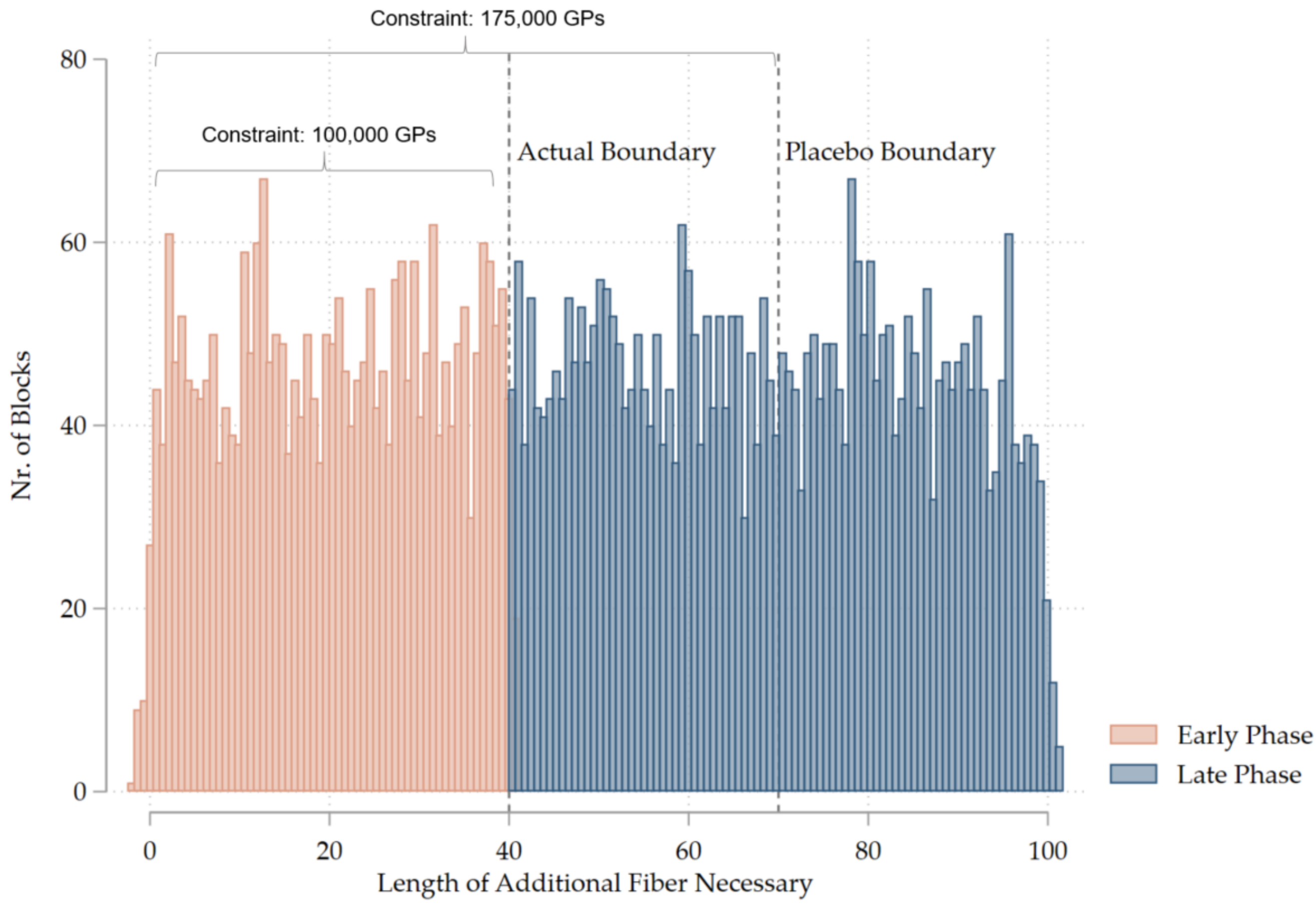
The results reported in Table (2.2), columns 3-6 do not show any evidence for discontinuities. This supports the main identifying assumption of models (2.1), (2.2) and (2.3), namely the absence of other discontinuities and constant differences for models (2.2) and

[32]Another approach to indirectly test for pre-existing discontinuities assesses the smoothness in the density of observations around the cutoff (Cattaneo et al., 2020). Figure (B.1) presents the density of NREGA GPs and shows, confirmed by a formal test, no evidence for a discontinuity at the cutoff.

[33]For specifications that are restricted to Jharkhand, I use all district boundaries where the treatment status does not change as opposed to only non-treated areas to construct the counterfactual. This ensures enough (placebo) identifying variation.

(2.3). It also mitigates concerns about compound treatment effects due to the overlap between phase boundaries and low-level administrative boundaries.

Figure 2.3 − Thought Experiment behind Placebo Boundary



Notes: The histogram illustrates the thought experiment behind the placebo boundary. It exemplifies the selection of the 6,697 blocks in India into phase I and II. According to Satyanarayana et al. (2015), all blocks were ranked based on the number of kilometers of fiber optic cable needed to connect each GP in a given block to the existing fiber optic network. The blocks with the lowest need were then selected subsequently into phase I until the number of GPs reached 100,000. A different constraint would have resulted in a different boundary, which still follows the same logic, however. Changing the constraint, therefore, generates plausible counterfactuals (the figure illustrates a counterfactual with a constraint at 175,000 GPs for phase I). Note, that the actual number of kilometers that each block needs to add is not known to me and the variable used here is hypothetical to illustrate the argument.

Table 2.2 – Internet, Conflict, and Public Works, Placebo

	(1) Assault	(2) Δ Assault	(3) BJP Riot	(4) BJP Mob	(5) INC Riot	(6) INC Mob
	Panel A: Conflict, India					
Placebo	-0.846	-0.941	0.003	-0.005	-0.046	-0.041
	(0.610)	(0.599)	(0.058)	(0.057)	(0.056)	(0.055)
Observations	19,772	19,710	19,772	19,772	19772	19,772
Bandwidth	10km	10km	10km	10km	10km	10km
Period	Pre	Pre	Post	Post	Post	Post
Placebo	Time	Time	Space	Space	Space	Space
	Panel B: Conflict, Jharkhand					
Placebo	-2.097	-2.256	0.086	0.086	0.086	0.086
	(2.637)	(2.633)	(0.065)	(0.065)	(0.065)	(0.065)
Observations	960	956	1,843	1,843	1,843	1,843
Bandwidth	10km	10km	20km	20km	20km	20km
Period	Pre	Pre	Post	Post	Post	Post
Placebo	Time	Time	Space	Space	Space	Space
	Registrations: Muslim Share	Registrations: Muslim Share	NREGA Work Days	NREGA Work Days	NREGA Work Days	NREGA Work Days
	Panel C: Public Works, Jharkhand					
Placebo	0.007	0.018	0.058	0.097**	-0.255	0.073*
	(0.021)	(0.043)	(0.040)	(0.038)	(0.156)	(0.038)
Muslim × Placebo			0.051	-0.085	0.063	-0.098
			(0.079)	(0.066)	(0.205)	(0.066)
Muslim Pres. × Muslim × Placebo						0.273
						(0.301)
Observations	2,735	763	5,650,320	4,285,068	429,824	4,749,420
GP presidents	All	All	All	Non-Muslim	Muslim	All
Bandwidth	10km	10km	10km	10km	12.5km	10km
Period	2009-2011	2011	2019-2022	2019-2022	2019-2022	2019-2022
Placebo	Time	Time	Space	Space	Space	Space

Notes: This table shows a falsification test. Columns 1-2 test for discontinuities in the pre-treatment period, while columns 3-6 test for discontinuities at a placebo boundary. Panels A and B estimate model (2.1) at the GP level. Columns 1-2 are based on GDELT and estimate the impact of BharatNet on the natural logarithm of one plus the (change in) number of assaults. The outcomes in columns 3-6 are based on ACLED and measure the natural logarithm of one plus the number of riots or mobs initiated by BJP or INC supporters, respectively. Mobs are a subset of riots. The bandwidth is always 10km except if there is not sufficient variation, in which case I double the bandwidth. All coefficients in Panels A and B are multiplied by 100. In Panel C, columns 1 and 2 estimate model (2.1) at the GP-year level. The outcome is the share of Muslims among newly registered households in NREGA. Columns 3-5 estimate model (2.2) at the individual-year level. The outcome is the natural logarithm of one plus the number of workdays allocated within NREGA. Column 6 estimates model (2.3). The bandwidth is always 10km except if there is not sufficient variation in which case I add 2.5km. Standard errors are clustered at the GP-year level for columns 3-6. All specifications include 20km segment-fixed effects. Triangular kernel weights are applied. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$

RD bandwidth: In the baseline, I restrict the sample to a small bandwidth of 10km (following other spatial RDD applications such as Dell et al., 2018; Dell and Olken, 2020; Lowes and Montero, 2021; Méndez and Van Patten, 2022). The choice of the bandwidth can be seen as a trade-off, where small bandwidths allow for a good approximation of any functional form by the linear RD polynomial, while larger bandwidths increase the power (Cattaneo et al., 2019). Tables (B.4) and (B.5) show the results for alternative bandwidths of 7.5km and 12.5km. Figure (B.2) extends the test for my preferred specification, estimating model (2.3) for a range of bandwidths starting with 30km and moving down in 1km steps to the point where the power is insufficient at 5km.

RD Kernel Weights: The baseline specifications apply triangular kernel weights. Thus, they put more weight on observations close to the boundary. Table (B.6), Panel B highlights that the main results, based on NREGA outcomes, are robust to giving equal weight to all observations.

RD Polynomial: All specifications apply local linear RD polynomials. A low-order polynomial is recommended in the literature to avoid overfitting and bad performance at the cut-off (Gelman and Imbens, 2019). Table (B.6), Panel C applies a quadratic RD polynomial to the main specifications, based on NREGA outcomes. The results are overall robust, only the discontinuity in the share of Muslim registrations loses its significance. The magnitude and the sign of the coefficients of interest remain very similar.

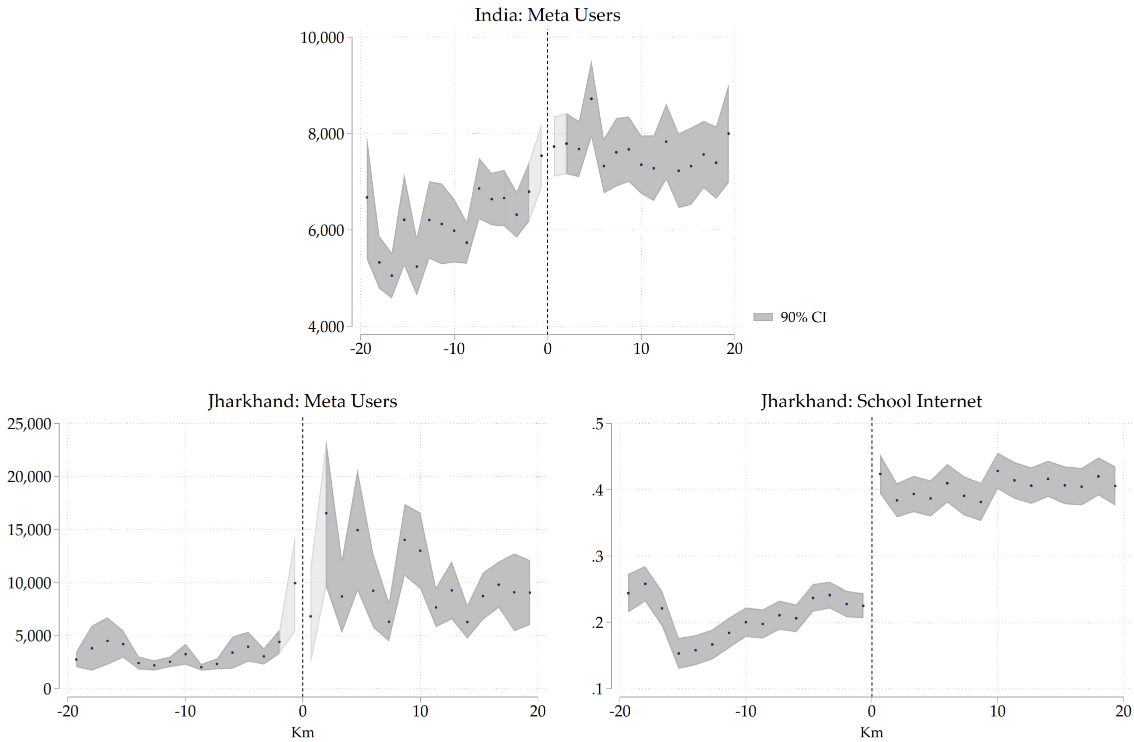
2.5 Main Results

I begin by documenting the impact of internet availability on usage in Section (2.5.1). I then examine whether fast internet divides rural communities. Section (2.5.2) assesses the influence on conflict. The treatment of the main religious minority - Muslims - in India's biggest public works program is analyzed in Section (2.5.3) and finally voting behavior in villages with and without Muslims is assessed in Section (2.5.4).

2.5.1 Internet Usage

The connection of phase I GPs to the fiber optic network does not necessarily cause a discontinuous jump in internet access at the phase boundary. I confirm the presence of a discontinuity based on social media usage and school internet data. First, Facebook and Instagram (Meta) usage in 2020 within 2km of the point of connection are examined. Points closer than 2km are dropped in the estimation to avoid an overlap between the radius and the boundary. Figure (2.4) highlights a strong discontinuity in monthly active Meta users at the boundary. The number increases by 1,135 on average India-wide and by 6,050 within Jharkhand.[34] Second, I test for discontinuities in internet availability in Jharkhand's schools. Precise locations of the schools allow me to test this at a very local level. Figure (2.4) documents a discontinuous increase in the likelihood of internet availability. Schools in phase I close to the boundary are 19.9 percentage points more likely to have internet compared to neighboring schools just across the boundary.

Figure 2.4 − BharatNet and Internet Usage



Notes: The figure depicts on the y-axis the number of active monthly Meta users (Facebook and Instagram) within 2km of the connection point of the fiber-optic cable within a GP in a binned plot. The x-axis denotes the distance to the nearest boundary between phase I and phase II in kilometers. Positive values denote the distance for GPs in phase I (early internet); negative values denote the distance for GPs in phase II (late internet). Note that the number of active Meta users includes users on both sides of the boundary if the connection point of the fiber-optic cable lies within 2km of the boundary (area with confidence intervals in light grey). When I estimate model (2.1) in a donut-RDD such that observations within 2km of the boundary are excluded, internet availability increases active monthly Meta users by 1,135.0* for the full sample and by 6,050.0** for Jharkhand. The probability that the internet is available in a school in Jharkhand increases by 19.9*** percentage points at the boundary.

These results confirm the uptake of the broadband infrastructure documented in several media and government reports. The Economic Times, for example, announced in 2017 that major telecommunication providers like Reliance Jio, Airtel, Vodafone India, and

[34]The average in the full sample is 6,924. The higher increase in Jharkhand mirrors state-wide numbers. A McKinsey Global Institute (2019) report includes Jharkhand among the five states with the fastest growth in internet penetration between 2014 and 2018.

Idea Cellular have made use of the broadband infrastructure and until 2021, 104,310 GPs had public Wifi hotspots installed, 510,559 homes were connected and 36,000km of unused fiber ("dark fiber") were leased (Ministry of Communications, 2021).

2.5.2 Conflict

This section reports the impact of the fast internet on extreme forms of divisions. It estimates model (2.1) on the level and change of assaults, as well as the number of riots and mobs by BJP and INC supporters. Table (2.3), Panel A shows overall an increase in the number of assaults by 0.5% on average in the full sample, as well as an increase in the growth rate by 0.43%. The increase is considerably more pronounced in Jharkhand, where assaults increase by 12.79% and the growth rate rises by 13.08%.[35] These estimates mirror the large increase in internet usage documented in Jharkhand in Section (2.5.1).

Table 2.3 – Internet and Conflict

	(1) Assault	(2) Δ Assault	(3) BJP Riot	(4) BJP Mob	(5) INC Riot	(6) INC Mob
	Panel A: India					
Internet	0.499*	0.430*	0.111**	0.104**	-0.095	-0.108*
	(0.267)	(0.234)	(0.044)	(0.043)	(0.060)	(0.058)
Observations	55,707	55,487	55,707	55,707	55,707	55,707
Bandwidth	10km	10km	10km	10km	10km	10km
	Panel B: Jharkhand					
Internet	12.785***	13.075***	3.006***	3.006***	-0.099	-0.099
	(4.093)	(4.099)	(1.049)	(1.049)	(0.069)	(0.069)
Observations	1,294	1,287	1,294	1,294	2,372	2,372
Bandwidth	10km	10km	10km	10km	20km	20km

Notes: This table estimates model (2.1) at the GP level. Columns 1-2 are based on GDELT and estimate the impact of BharatNet on the natural logarithm of one plus the (change in) number of assaults. The outcomes in columns 3-6 are based on ACLED and measure the natural logarithm of one plus the number of riots or mobs initiated by BJP or INC supporters, respectively. Mobs are a subset of riots. The bandwidth is always 10km except if there is not sufficient variation in which case I double the bandwidth. All coefficients are multiplied by 100. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$

Columns 3-6 report the impact on riots and mobs by party. If inflammatory online messages by the BJP drive parts of the increase in divisions, one would expect an increase in conflict events by their supporters. Columns 3 and 4 show an 0.11% increase in the number of riots and an 0.1% increase in the number of mobs by BJP supporters India-wide, significant at the 5% level. Again, the increase is more dramatic in Jharkhand

[35] Note that the magnitude but not the sign of these coefficients depends on the bandwidth. The increase is 8.89% and 9.83% for a 12.5km bandwidth.

where BJP riots and mobs rise by 3.01%. In contrast, fast internet does not increase riots or mobs by INC supporters if anything the number of mobs declines slightly by 0.1% in the full sample.

These patterns are consistent with an increase in violent divisions driven by the exposure to divisive debates on the internet. This increase seems to be partly driven by the BJP, although the increase in BJP riots cannot explain the full increase. Fast internet does not impact divisions driven by the INC. Although I cannot rule out that measurement bias influences the magnitude of the estimates, the negative point estimate for INC violence makes it unlikely that they explain the full result. These findings serve as a motivation to explore emerging religious divisions in other dimensions in more detail in the next sections.

2.5.3 Internet and Public Works

The impact of the fast internet on the allocation of scarce welfare benefits is examined in this section. If the exposure to internet affects the salience of group identities, it can affect the allocation decisions of local officials. Analyzing divisions based on administrative data on welfare benefits has several advantages: First, it allows me to isolate the information channel from the coordination channel potentially present in the conflict results. Second, it provides an opportunity to explicitly focus on Muslims. Third, measurement errors are unlikely. Fourth, variation in the GP president responsible for the allocation enables an assessment of the supply channel.

I start with the examination of entries into NREGA before assessing the allocation of NREGA work days among registered households. To capture group divisions along religious lines, I always differentiate between Muslims and non-Muslims. Every adult resident of a GP is eligible to register for NREGA. Entries are therefore not supply constrained such that fast internet can affect the differential entry of Muslims due to multiple reasons. Most importantly, entries highlight the internet's impact on the religious composition of registered households among which NREGA work days are distributed. The RDD results are reported in Table (2.4) and show a 2.8 percentage point decline in the share of Muslims among households that registered between 2017 and 2022. The result is significant at the 5% level and appears immediately in 2017. The differential decline in registrations could be explained by a decrease in need or a decrease in the expected benefit (if discrimination is anticipated). The quick materialization of the pattern speaks to a psychological rather than an economic channel (I examine further evidence for the economic channel in Section (2.6) and find no support).

Table 2.4 – Internet and Public Works

	(1) Registrations: Muslim Share	(2) Registrations: Muslim Share	(3) NREGA Work Days	(4) NREGA Work Days	(5) NREGA Work Days	(6) NREGA Work Days
Internet	-0.028**	-0.056*	-0.013	-0.067	-0.494	-0.074
	(0.011)	(0.031)	(0.054)	(0.050)	(0.317)	(0.050)
Muslim × Internet			-0.056	-0.109**	0.221***	-0.125***
			(0.047)	(0.047)	(0.065)	(0.047)
Muslim Pres. ×						0.417***
Muslim × Internet						(0.086)
Observations	6,697	1,128	6,865,104	5,589,676	487,120	6,091,712
GP president	All	All	All	Non-Muslim	Muslim	All
Segment FE	Yes	Yes	Yes	Yes	Yes	Yes
Period	2017-2022	2017	2019-2022	2019-2022	2019-2022	2019-2022

Notes: Columns 1 and 2 estimate model (2.1) at the GP-year level. The outcome is the share of Muslims among newly registered households in NREGA. Columns 3-5 estimate model (2.2) at the individual-year level. The outcome is the natural logarithm of one plus the number of workdays allocated within NREGA. Column 6 estimates model (2.3). Standard errors are clustered at the GP-year level for columns 3-6. All specifications include 20km segment-fixed effects and are restricted to a bandwidth of 10km. Triangular kernel weights are applied. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$

Next, I turn to the allocation of scarce NREGA work days. Since demand for work days considerably exceeds supply, it is the GP president who determines which household benefits. Changes in the religious composition of households after the arrival of fast internet reflect then the internet's impact on the GP president's allocation decision. Table (2.4), column 3 reports the estimates based on the difference-in-discontinuity model. The point estimate is negative but insignificant such that there is no evidence for a differential treatment of Muslims on average (conditional on being registered). The imprecision in the estimate may disguise a sizeable decrease of 5.6% in the distribution of work days to Muslims relative to non-Muslims.

The average picture hides important differences, however. Any disadvantage of Muslims in NREGA due to internet exposure would likely hold not for all GP presidents but only for non-Muslim presidents. Therefore, I focus on GPs with at least one registered Muslim resident but with no Muslim president first. Column 4 highlights the result. Muslims receive 10.9% fewer NREGA days than non-Muslims in villages that have fast internet, significant at the 5% level. Thus, the internet results in fewer benefits for Muslims in GPs where the president does not share their religion. Next, I turn to GPs with a Muslim president. If the internet affects Muslims in general differently, the effect should remain. This is not the case: the sign of the coefficient reverses in GPs with a Muslim president. In particular, Muslims get 22.1% more work days relative to non-Muslims, significant at the 1%-level. Hence, the internet leads to relatively more benefits for Muslims in GPs where the president is of the same religion. These patterns are in line with strengthening religious identities ultimately distorting the distribution of welfare benefits.

Finally, I re-estimate the relationship in model (2.3) on a sample of all villages with at least one registered Muslim. I directly contrast the treatment of Muslims (relative to non-Muslims) by Muslim GP presidents in neighboring villages across the internet boundary and compare it to the treatment of Muslims (relative to non-Muslims) by non-Muslim GP presidents in neighboring villages across the internet boundary. The results in column 6 show that the internet results in a 12.5% decline in NREGA work days for Muslims (relative to non-Muslims) in GPs without a Muslim president. In contrast, they benefit relatively to non-Muslims from internet exposures in GPs with a Muslim president. Overall, the difference-in-difference-in-discontinuity design confirms the earlier pattern.

2.5.4 Political Impact

I turn to the political implications of rural internet availability next. The exposure to divisions online and offline (documented before) can impact voting behavior. Around the world, ethnically fractionalized countries vote often based on identity as opposed to merit. Therefore, I examine the internet's impact on the extent to which the vote share for Muslim candidates is related to the share of Muslims in a GP (approximated by the share of Muslims registered in NREGA since 2006). The results are presented in Table (2.5), column 3 and show a significant increase in the relation. The results are consistent with an increase in voting based on religious identity, such that Muslims more likely vote for a Muslim if exposed to fast internet.[36] The impact is limited, however. Muslim candidates do not play an important role nationally.

Therefore, I examine vote shifts between the two main poles in the debate around India's religious identity: the Hindu nationalist BJP and secular parties (these two groups represent 68.1% of total votes). Again, I differentiate by the share of Muslims in a GP. It is a viable strategy for a targeted minority to combine their voting power behind an inclusive broader franchise as offered by the secular parties. An increase in religious divisions can then lead to an increase in votes for that franchise, while non-Muslims start favoring the other pole. Table (2.5) shows that polling stations in non-Muslim areas with fast internet report a shift of votes to the Hindu nationalist party. The shift is sizeable and declines with the share of Muslims. The discontinuity in the shift to the BJP is consistent with a general increase in Hindu nationalism due to the internet, but could also be explained with voters rewarding the government for the set-up of fast internet. The latter does not explain the absence of an increase in BJP votes in villages with Muslims, however. Although these areas do not increase their support of the BJP, the results suggest that they shift away from secular parties in villages with a small share of Muslims (below the state average). The sign switches in GPs with a considerable Muslim

[36]The small but significant effect of internet availability on votes for Muslim candidates in non-Muslim villages could be due to the imperfect measurement of a village's Muslims (as I rely on NREGA registrations).

share. Polling stations in areas with many Muslims increase their support of secular parties when they get exposed to the internet. The absence of individual-level voting data makes a conclusive interpretation difficult. Nevertheless, these patterns are consistent with the notion that a minority rallies behind a larger beneficial franchise if it cannot get the majority.

Table 2.5 – Internet and Voting

	(1) Hindutva-Shift	(2) Secular Votes	(3) Muslim Cand.	(4) BJP	(5) INC
			Panel A: All Villages		
Internet	0.056**	-0.032**	0.002***	0.024	-0.033***
	(0.027)	(0.012)	(0.000)	(0.019)	(0.012)
Muslim × Internet	-0.174***	0.123***	0.001**	-0.051	0.131***
	(0.058)	(0.024)	(0.001)	(0.049)	(0.022)
Observations	3,240	3,240	3,240	3,240	3,240
			Panel B: Non-Muslim Village		
Internet	0.081**	-0.031**	0.001***	0.049*	-0.028*
	(0.038)	(0.015)	(0.000)	(0.030)	(0.015)
Observations	1,187	1,187	1,187	1,187	1,187
			Panel C: Muslim Village		
Internet	0.018	-0.013	0.003***	0.005	-0.015
	(0.036)	(0.017)	(0.000)	(0.024)	(0.017)
Observations	2,044	2,044	2,044	2,044	2,044
			Panel D: Minority Muslim Village		
Internet	0.069	-0.065***	0.002***	0.004	-0.065***
	(0.046)	(0.023)	(0.000)	(0.028)	(0.023)
Observations	1,317	1,317	1,317	1,317	1,317
			Panel E: Majority Muslim Village		
Internet	-0.083	0.059***	0.003***	-0.024	0.047**
	(0.058)	(0.021)	(0.001)	(0.051)	(0.020)
Observations	724	724	724	724	724

Notes: Panel A estimates model (2.2) at the polling station level in Jharkhand. Panel B estimates model (2.1) on different subsets. In particular, polling stations located in non-Muslim villages (Panel B), Muslim villages (Panel C), and villages with a Muslim share below (Panel D) and above (Panel E) state average. The outcome in column 1 is the number of votes for Hindutva parties minus the number of votes for secular parties as a share of total votes. Column 2 assesses the impact on the vote share of secular parties, column 3 on the vote share of Muslim candidates, column 4 on the vote share of the BJP (the main Hindutva party) and column 5 on the vote share of the INC (the main secular party). All specifications include 20km segment-fixed effects and are restricted to a bandwidth of 10km. Triangular kernel weights are applied. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$

2.6 Alternative Mechanism

The results presented so far show an increase in divisions in rural communities. The increase can be attributed to a change in the information set as formerly isolated communities gain access to heated national debates online. The patterns can also be explained through a different mechanism, however. Hjort and Poulsen (2019) have shown an increase in employment in Sub-Sahara Africa after an area gets connected to broadband internet. Economic gains if unequally distributed can increase tensions between groups as shown by Mitra and Ray (2014) for the case of Muslims in India. Did the internet create inequalities that produced the patterns highlighted above? In the following, I discuss the evidence for this alternative mechanism.

Table 2.6 – Internet and Economic Effects

	(1) BPL	(2) Wealth	(3) BPL	(4) Wealth
	India		*Jharkhand*	
Internet	-0.013	50.818*	-0.036	-142.152
	(0.017)	(30.674)	(0.064)	(102.243)
Muslim × Internet	0.043*	75.029	0.195***	-56.591
	(0.026)	(54.515)	(0.072)	(108.823)
Observations	288,866	289,380	26,557	26,614
Segment FE	Yes	Yes	Yes	Yes
Internet			-0.043	-162.923
			(0.063)	(108.137)
Muslim × Internet			0.181***	-71.077
			(0.069)	(116.716)
Muslim Pres. × Muslim × Internet			0.056	30.975
			(0.197)	(198.850)
Observations			26,557	26,614
Segment FE			Yes	Yes

Notes: The upper panel estimates model (2.2) at the individual level; the bottom panel estimates model (2.3). The outcome is a dummy denoting whether a household owns a below poverty line card in columns 1 and 3. The outcome is the rural DHS-Wealth index for columns 2 and 4. Columns 1-2 are estimated on the full sample, columns 3-4 only consider Jharkhand for comparability. All specifications include 20km segment-fixed effects, are restricted to a bandwidth of 10km, and standard errors are clustered at the DHS-cluster level. Triangular kernel weights are applied. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$

In order to attribute the increase in NREGA work days for individuals of the same religion than the GP president to the internet's economic impact, the economic effects need to follow specific patterns. First, they need to materialize quickly as a decline in the share of Muslims is observable from 2017 onward. Second, the impact of the internet needs to reverse its sign in villages with compared to villages without a Muslim president. Third, Muslims need to gain disproportionately in villages without and gain less than average in villages with a Muslim president. Although it seems *a priori* unlikely that the impact of the internet would follow these patterns, I test for economic effects below.

Therefore, I leverage the latest round of DHS data covering the 2019-2021 period to test the internet's differential impact on households' wealth and poverty status. I rely on individual-level data on 2,062,660 adults after restricting the sample to rural areas and residents of the interviewed household. To identify the impact of fast internet on Muslims relative to non-Muslims, I estimate models (2.2) and (2.3) on both the full sample and Jharkhand for comparability. Table (2.6) presents the results. Columns 1 and 3 do not show any evidence for the hypothesis that poor Muslims benefit disproportionately from rural internet (India-wide and in Jharkhand), nor does it show that there are differential effects on Muslims in Jharkhand's DHS clusters with and without a Muslim president. In contrast, the likelihood that a Muslim is below the poverty line increases by 5.1 percentage points in the full sample and by 18.2 percentage points in Jharkhand. To take the whole distribution into account I assess the internet's impact on wealth using DHS‘s rural wealth index. I do not find any significant differential impact of the internet on Muslim's relative to non-Muslims' wealth. The same holds true for the difference in the difference between DHS clusters with and without a Muslim president.

The interpretation of the magnitudes has to be considered carefully, however, since DHS clusters are randomly shifted in space. As this paper relies on fine-grained spatial discontinuities, measurement errors in the location can considerably impact the result. To protect the respondent's confidentiality DHS shifts a cluster in a random direction and random distance between 0 and 5km from the true location for 99% of rural DHS clusters and up to 10km for the remaining 1%. Importantly for this paper, DHS assures that the clusters are not displaced across district boundaries. This ensures that points are never shifted across the internet boundary in the case of Jharkhand where the boundary intersects with district boundaries. One can therefore expect that the coefficients are upward biased (increasing the likelihood of detecting non-existent economic effects) as the continuity assumption might be violated.[37]

2.7 Conclusion

Billions of people live in rural areas in developing countries. As they get access to the internet, they are joining national conversations that were once far removed. How does that impact the cohabitation of different groups in rural communities that have lived

[37] Table (B.1) tests for significant differences in outcomes based on the 2011 census within 5km on both sides of the boundary. There are no significant differences in Jharkhand but slightly more schools in early treated areas in the full sample. This mitigates concerns regarding sizeable bias due to the displacement of DHS clusters. If GPs would have a circular shape their average diameter would be around 5km. Since I use the location of the DHS cluster to link it to a GP president, the GP president's religion is subject to random measurement error.

together for decades?

This paper analyzes the impact of the largest rural government broadband initiative worldwide on group divisions in rural communities. It collects new data on the location of 175,157 broadband connections that aim to connect every GP to the fiber optic network. In combination with Wifi hotspots and a staggered roll-out, the initiative creates spatial discontinuities in internet usage between villages receiving fast internet in phase I and those which receive it in phase II.

This paper examines several dimensions of divisions. It considers extreme outcomes and shows that the sudden increase in internet results in an increase in assaults, as well as an increase in riots and mobs by supporters of the Hindu nationalist BJP. These findings motivate more detailed assessments of moderate but widespread forms of divisions in the state of Jharkhand. I document increasing distortions in the allocation of NREGA welfare benefits along religious lines by GP presidents in a difference-in-discontinuity design. Non-Muslim GP presidents allocate fewer work days to Muslims. The reverse is apparent for Muslim presidents, which favor Muslims over others. Further specifications rule out a differential economic impact that could explain these patterns. Thus, the evidence suggests that the change in the information set brought by the internet leads elected representatives to allocate public goods based on group identities. In a final step, this paper explores the political consequences and shows an increase in votes for the Hindu nationalist BJP in villages without (in NREGA registered) Muslims and an increase in votes for the secular INC and Muslim candidates in villages with a sizeable Muslim share. Although the evidence varies in depth, in sum, it paints a coherent picture that reinforces itself. Fast internet divides rural communities.

The results suggest vast consequences for the developing world as rural communities gain internet access. They highlight that changes in the information environment can transfer national divisions to local rural communities. These results may hinge on the national debate and the design of (social media) algorithms, but they fuel doubts that the internet can strengthen national identities in diverse nation-states by reducing the cost of distance.

Appendices

A Descriptives

Table A.1 – Summary Statistics

	N	Sum	Mean	SD	Min	Max
	Panel A: India					
Early Phase	174,736	93,687	0.54	0.50	0	1
Distance (km)	174,736	2,196,091	12.57	72.68	-368.02	1,662.12
Absolute Distance (km)	174,736	6,452,036	36.92	63.85	0.01	1,662.12
Assaults (ln)	175,157	1,531.90	0.01	0.14	0	7.39
Δ Assaults (ln)	174,421	1,208.14	0.01	0.12	0	6.83
BJP Riots (ln)	175,157	449.86	0.003	0.06	0	3.26
BJP Mobs (ln)	175,157	403.15	0.002	0.05	0	3.22
INC Riots (ln)	175,157	191.18	0.001	0.03	0	1.95
INC Mobs (ln)	175,157	145.85	0.001	0.03	0	1.61
Wealth	2,062,660	1.05e+08	50.69	1,005.80	-2,556.24	3,186.81
BPL	2,059,595	1,105,883	0.54	0.50	0	1
	Panel B: Jharkhand					
Early Phase	4,336	2,797	0.65	0.48	0	1
Distance (km)	4,336	50,698	11.69	24.43	-62.05	84.69
Absolute Distance (km)	4,336	91,483.77	21.10	16.99	0.03	84.69
Assaults (ln)	4,337	65.62	0.02	0.16	0	2.56
Δ Assaults (ln)	4,304	49.32	0.01	0.14	0	2.20
BJP Riots (ln)	4,337	10.40	0.002	0.04	0	0.69
BJP Mobs (ln)	4,337	10.40	0.002	0.04	0	0.69
INC Riots (ln)	4,337	3.47	0.001	0.02	0	0.69
INC Mobs (ln)	4,337	3.47	0.001	0.02	0	0.69
Registration Muslim Sh.	21,242	2,684.32	0.13	0.21	0	1
NREGA Work Days (ln)	19,951,740	2.23e+07	1.12	1.72	0	5.55
Hindutva Shift (Sh.)	29,147	9,458.32	0.32	0.43	-1	1
Secular Votes (Sh.)	29,147	5,270.35	0.18	0.24	0	1
Muslim Cand. (Sh.)	29,147	71.26	0.002	0.007	0	0.27
BJP (Sh.)	29,147	14,728.67	0.51	0.29	0	1
INC (Sh.)	29,147	4,758.19	0.16	0.24	0	1
Wealth	81,467	-5.78e+07	-709.75	854.84	-2,526.35	2,764.01
BPL	81,329	53,759	0.66	0.47	0	1

Notes: This table shows summary statistics for the full sample in Panel A and Jharkhand in Panel B. Early Phase is a binary variable equal to one if a GP got internet in phase I, and zero if in phase II. Distance (km) denotes the distance (phase I: kilometer, phase II: kilometer$\times(-1)$) from the GP to the closest point on the boundary. Assaults captures (the change of) the natural logarithm of one plus the number of assaults. Riots captures the natural logarithm of one plus the number of riots by supporters of the BJP or INC. Mobs are a subset of riots. Registration Muslim Sh. captures the share of Muslims among newly registered households within a year and GP in NREGA. NREGA work days denote the natural logarithm of one plus the number of NREGA work days allocated to a household per year. Hindutva shift captures the number of BJP votes minus the number of votes for secular parties as a share of total votes. Secular votes is the share of votes to the INC, All India Trinamool Congress, Communist Party of India, Communist Party of India (Marxist, Liberation), Communist Party of India (Marxist-Leninist, Red Star). Muslim Cand. is the share of votes for Muslim candidates irrespective of the party. Wealth and BPL are based on the DHS (2019-2021) wave. Wealth denotes the rural DHS-Wealth index. BPL is a binary variable denoting whether a DHS-household has a below-poverty-line card.

Table A.2 – Balance Table, Pre-Treatment, Full Sample

	Panel A: India			*Panel B: Jharkhand*		
	Early Phase	Late Phase	Diff.	Early Phase	Late Phase	Diff.
Total Population	3316.2	2698.2	618.0***	3894.7	2525.5	1369.1*
No. of Households	672.7	558.3	114.4***	728.2	490.3	238.0
Literate Population	2171.4	1641.5	529.9***	2417.5	1358.7	1058.8*
Scheduled Caste Population	538.5	430.8	107.7***	477.1	292.0	185.1***
Scheduled Tribe Population	243.4	224.5	18.9***	568.3	660.9	-92.5
No. of Primary Schools	2.3	2.1	0.2***	1.9	1.7	0.2
No. of Middle Schools	1.2	0.9	0.3***	1.3	1.0	0.3
No. of Secondary Schools	0.6	0.4	0.2***	0.4	0.3	0.1*
No. of Senior Secondary Schools	0.3	0.2	0.1***	0.1	0.1	0.0
No. of Colleges	0.1	0.1	0.0***	0.0	0.0	0.0
Assault	0.1	0.1	-0.0	0.0	0.0	-0.0
Violent Protest	0.0	0.0	0.0	0.0	0.0	0.0
Peaceful Protest	0.0	0.0	-0.0	0.0	0.0	0.0

Notes: Columns 1 and 2 report means for phase I and phase II GPs. Column 3 reports the difference between phase I and phase II. Stars denote p-values from a t-test, where $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$. Columns 4-6 repeat the exercise for GPs in Jharkhand.

B Further Robustness

Table B.1 – Balance Test, Pre-Treatment, 5km Bandwidth

	Panel A: India			*Panel B: Jharkhand*		
	Early Phase	Late Phase	Diff.	Early Phase	Late Phase	Diff.
Total Population	2260.9	2288.7	-27.7	2538.5	2989.3	-450.8
No. of Households	435.7	438.6	-2.8	480.0	581.8	-101.7
Literate Population	1351.9	1362.2	-10.3	1421.8	1552.3	-130.5
Scheduled Caste Population	409.0	391.7	17.3	351.1	297.3	53.8
Scheduled Tribe Population	171.0	172.7	-1.7	565.1	553.5	11.6
No. of Primary Schools	1.7	1.7	0.1	1.6	1.6	-0.0
No. of Middle Schools	0.9	0.8	0.1*	0.9	0.9	-0.1
No. of Secondary Schools	0.4	0.3	0.0*	0.3	0.2	0.1
No. of Senior Secondary Schools	0.2	0.2	0.0*	0.1	0.1	-0.0
No. of Colleges	0.0	0.0	-0.0	0.0	0.0	-0.0
Assault	0.0	0.1	-0.0	0.0	0.0	0.0
Violent Protest	0.0	0.0	-0.0	0.0	0.0	0.0
Peaceful Protest	0.0	0.0	-0.0	0.0	0.0	0.0

Notes: Columns 1 and 2 report means for phase I and phase II GPs located within 5km of the boundary. Column 3 reports the difference between phase I and phase II. Stars denote p-values from a t-test, where $*p < 0.1$, $**p < 0.05$, and $***p < 0.01$. Columns 4-6 repeat the exercise for GPs in Jharkhand.

Table B.2 − Test for other Discontinuities - Other

	(1) Nightlights	(2) Avg. Nightlights	(3) Electricity	(4) Rural Consumption	(5) PMGSY Timing	(6) PMGSY Length
				Panel A: India		
Early Phase	-0.002	0.008	0.000	51.517	65.566	-0.193
	(0.021)	(0.012)	(0.006)	(59.252)	(66.312)	(0.226)
Observations	48616	48616	42296	48081	6700	5724
				Panel B: Jharkhand		
Early Phase	0.116	0.039	0.000	-424.857	-551.968	-0.965
	(0.146)	(0.091)	(.)	(355.860)	(344.055)	(1.426)
Observations	1162	1162	21	1106	309	219

Notes: This table tests for pre-existing discontinuities at the boundary. It estimates the main coefficient of interest (β_1) of model Y for a number of outcomes from SHRUG. The model is estimated at the GP level, restricted to locations less than 10km away from the boundary. Panel A reports coefficients for the full sample and includes 20km segment- and state-fixed effects. Panel B is restricted to Jharkhand and includes 20km segment-fixed effects. Observations are weighted using triangular kernel weights. *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$

Table B.3 − Internet and Conflict, Robustness

	(1) Assault	(2) Δ Assault	(3) BJP Riot	(4) BJP Mob	(5) INC Riot	(6) INC Mob
	Panel A: Change in Tribal Population, Control					
Internet	0.419*	0.358	0.120***	0.112**	-0.098	-0.111*
	(0.251)	(0.220)	(0.045)	(0.044)	(0.061)	(0.059)
Observations	54,615	54,398	54,615	54,615	54,615	54,615
	Panel B: Change in Nr. of Colleges, Control					
Internet	0.417*	0.342	0.115**	0.109**	-0.099	-0.109*
	(0.251)	(0.220)	(0.045)	(0.044)	(0.061)	(0.059)
Observations	54,555	54,341	54,555	54,555	54,555	54,555

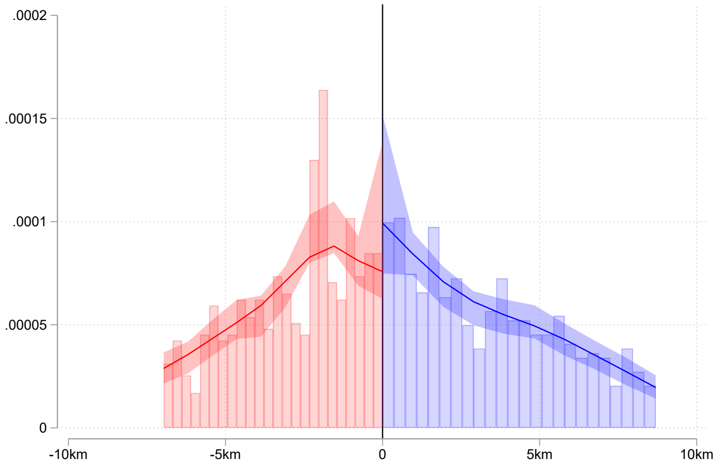
Notes: This table re-estimates Table (2.3) including the change in the tribal population (Panel A) and the change in the number of colleges (Panel B) as control variables. All coefficients are multiplied by 100. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$

Table B.4 – Internet and Conflict, Bandwidths

	(1) Assault	(2) Δ Assault	(3) BJP Riot	(4) BJP Mob	(5) INC Riot	(6) INC Mob
	Panel A: India					
Internet	0.571*	0.492*	0.098*	0.098*	-0.132*	-0.134*
	(0.330)	(0.289)	(0.053)	(0.053)	(0.075)	(0.072)
Observations	43,169	42,993	43,169	43,169	43,169	43,169
BW	7.5km	7.5km	7.5km	7.5km	7.5km	7.5km
Internet	0.499*	0.430*	0.111**	0.104**	-0.095	-0.108*
	(0.267)	(0.234)	(0.044)	(0.043)	(0.060)	(0.058)
Observations	55,707	55,487	55,707	55,707	55,707	55,707
BW	10km	10km	10km	10km	10km	10km
Internet	0.485**	0.396**	0.095**	0.087**	-0.075	-0.082*
	(0.229)	(0.202)	(0.038)	(0.037)	(0.052)	(0.050)
Observations	66,387	66,099	66,387	66,387	66,387	66,387
BW	12.5km	12.5km	12.5km	12.5km	12.5km	12.5km
	Panel B: Jharkhand					
Internet	18.758***	18.628***	3.678**	3.678**	-0.161	-0.161
	(5.703)	(5.727)	(1.453)	(1.453)	(0.100)	(0.100)
Observations	957	950	957	957	2,147	2,147
BW	7.5km	7.5km	7.5km	7.5km	17.5km	17.5km
Internet	12.785***	13.075***	3.006***	3.006***	-0.099	-0.099
	(4.093)	(4.099)	(1.049)	(1.049)	(0.069)	(0.069)
Observations	1,294	1,287	1,294	1,294	2,372	2,372
BW	10km	10km	10km	10km	20km	20km
Internet	8.892***	9.835***	2.364***	2.364***	-0.044	-0.044
	(3.183)	(3.156)	(0.829)	(0.829)	(0.047)	(0.047)
Observations	1,613	1,602	1613	1,613	2,593	2,593
BW	12.5km	12.5km	12.5km	12.5km	22.5km	22.5km
Internet	4.389**	5.386***	1.317***	1.317***	-0.099	-0.099
	(1.981)	(1.945)	(0.507)	(0.507)	(0.069)	(0.069)
Observations	2,372	2,353	2,372	2,372	2,372	2,372
BW	20km	20km	20km	20km	20km	20km

Notes: This table re-estimates Table (2.3) applying different bandwidths. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$

Figure B.1 − Density around the Cutoff



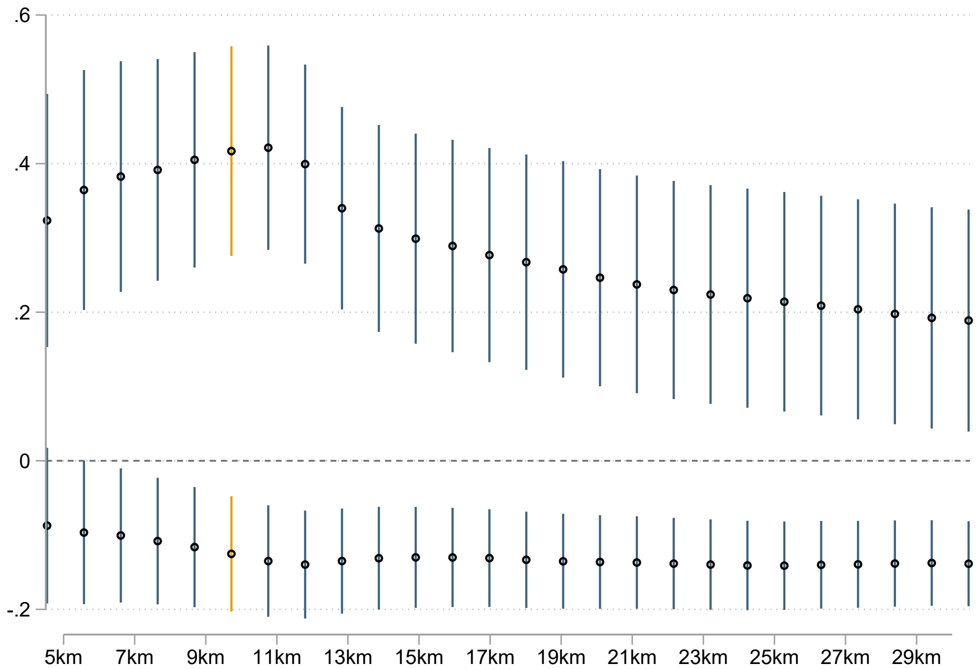
Notes: The figure shows the density of the observations (NREGA GPs in Jharkhand) with respect to their distance to the boundary. It tests whether there are significant differences in the density around the cutoff following Cattaneo et al. (2020). There is no evidence for a significant difference (p-value = 0.15).

Table B.5 – Internet and Public Works, Bandwidths

	(1) Registration: Muslim Share	(2) Registration: Muslim Share	(3) Work Days	(4) Work Days	(5) Work Days	(6) Work Days
Internet	-0.026*	-0.053	0.017	-0.057	-0.926**	-0.074
	(0.014)	(0.038)	(0.064)	(0.061)	(0.430)	(0.060)
Muslim × Internet			-0.058	-0.079	0.200***	-0.103*
			(0.052)	(0.052)	(0.069)	(0.053)
Muslim Pres. × Muslim × Internet						0.386***
						(0.092)
Observations	5,133	864	5,359,580	4,368,508	386632	4,755,140
Segment FE	Yes	Yes	Yes	Yes	Yes	Yes
BW	7.5km	7.5km	7.5km	7.5km	7.5km	7.5km
Internet	-0.028**	-0.056*	-0.013	-0.067	-0.494	-0.074
	(0.011)	(0.031)	(0.054)	(0.050)	(0.317)	(0.050)
Muslim × Internet			-0.056	-0.109**	0.221***	-0.125***
			(0.047)	(0.047)	(0.065)	(0.047)
Muslim Pres. × Muslim × Internet						0.417***
						(0.086)
Observations	6,697	1,128	6,865,104	5,589,676	487,120	6,091,712
Segment FE	Yes	Yes	Yes	Yes	Yes	Yes
BW	10km	10km	10km	10km	10km	10km
Internet	-0.021**	-0.050*	-0.034	-0.050	-0.319	-0.055
	(0.010)	(0.027)	(0.048)	(0.044)	(0.273)	(0.044)
Muslim × Internet			-0.074*	-0.125***	0.223***	-0.138***
			(0.043)	(0.043)	(0.062)	(0.044)
Muslim Pres. × Muslim × Internet						0.363***
						(0.082)
Observations	8,059	1,369	8,254,524	6,628,000	691,708	7,347,460
Segment FE	Yes	Yes	Yes	Yes	Yes	Yes
BW	12.5km	12.5km	12.5km	12.5km	12.5km	12.5km

Notes: This table re-estimates Table (2.4) for different bandwidths. $^{*}p < 0.1$, $^{**}p < 0.05$, and $^{***}p < 0.01$

Figure B.2 – Internet and Public Works, Bandwidths



Notes: This figure re-estimates Table (2.4), column 6 for different bandwidths starting at 5km and moving in 1km steps to 30km. It presents the two coefficients of interest β_2 and β_3 based on model (2.3). 90% confidence intervals are displayed.

Table B.6 – Internet and Public Works, Robustness

	(1) Registration: Muslim Share	(2) Registration: Muslim Share	(3) Work Days	(4) Work Days	(5) Work Days	(6) Work Days
			Panel A: Main Specification			
Internet	-0.028**	-0.056*	-0.013	-0.067	-0.494	-0.074
	(0.011)	(0.031)	(0.054)	(0.050)	(0.317)	(0.050)
Muslim × Internet			-0.056	-0.109**	0.221***	-0.125***
			(0.047)	(0.047)	(0.065)	(0.047)
Muslim Pres. × Muslim × Internet						0.417***
						(0.086)
Observations	6,697	1,128	6,865,104	5,589,676	487,120	6,091,712
FE	Segment	Segment	Segment	Segment	Segment	Segment
			Panel B: Uniform Weights			
Internet	-0.020*	-0.049*	-0.071	-0.059	-0.373	-0.058
	(0.010)	(0.027)	(0.050)	(0.047)	(0.302)	(0.046)
Muslim × Internet			-0.078*	-0.174***	0.255***	-0.180***
			(0.041)	(0.041)	(0.060)	(0.042)
Muslim Pres. × Muslim × Internet						0.511***
						(0.081)
Observations	6,697	1,128	6,865,104	5,589,676	487,120	6,091,712
FE	Segment	Segment	Segment	Segment	Segment	Segment
			Panel C: Second-Order Polynomial			
Internet	-0.021	-0.045	-0.105	-0.066	-0.447	-0.124
	(0.020)	(0.052)	(0.087)	(0.089)	(0.350)	(0.085)
Muslim × Internet			-0.056	-0.107**	0.219***	-0.125***
			(0.047)	(0.047)	(0.065)	(0.047)
Muslim Pres. × Muslim × Internet						0.417***
						(0.086)
Observations	6,697	1,128	6,865,104	5,589,676	487,120	6,091,712
FE	Segment	Segment	Segment	Segment	Segment	Segment

Notes: This table re-estimates Table (2.4) applying uniform weights (Panel B) and a quadratic RD polynomial (Panel C). *$p < 0.1$, **$p < 0.05$, and ***$p < 0.01$

This chapter is co-authored with Aiko Schmeißer

3

Do Unions Shape Political Ideologies at Work?

3.1 Introduction

Political leaders credit unions for shaping welfare systems and labor market policies, such as the 8-hour day, minimum wage, safety standards, sick leave, weekends, family leave, overtime compensation, and retirement plans (e.g., Biden, 2021; King, 1965; Obama, 2010). While economists typically attribute unions' influence to their impact on work contracts via collective bargaining and study their effects on wages and other benefits (Card, 1996; DiNardo and Lee, 2004; Farber et al., 2021; Frandsen, 2021; Knepper, 2020), other scholars have seen unions as essential drivers of social and political change (Baumgartner and Leech, 1998; Dahl, 2005; Lipset, 1960). Through their enduring alliance with the Democratic party (Dark, 2001), which involves supporting Democratic candidates financially with campaign contributions and lobbying legislators to introduce labor-friendly policies, unions are often viewed as one of the few vehicles that give political voice to workers and enhances the representation of their preferences in U.S. politics (Burns et al., 2000; Rosenfeld, 2014; Schlozman, 2015). However, whether unions can bring about lasting change in welfare states and policies depends on their ability to change the political preferences and beliefs of workers and the broader public. Are unions able to shape political ideologies?

Unions' greatest political leverage likely arises from their connection to more than 14 million union members and their colleagues at the unionized workplace.[1] After family and friends, the workplace is the most important arena for political discussion (Hertel-Fernandez, 2020). Interactions among employees and social experiences at work make it a particularly influential space for unions. By providing political information and training as well as facilitating communication networks between members, unions can mobilize workers and affect their ideological positions. Still, unions' aggregate political influence at the workplace is far from clear. Even if they are able to assemble unionized workers around their political positions, it is unclear whether they can persuade the firm's management. Heightened tension between workers and managers, who represent the owners' interests, might yield adverse responses to labor issues. Any backlash in the political behavior of this powerful out-group may prevent unions from achieving their political agenda.

In this paper, we examine the influence of labor unions on the political participation and political ideologies of employees in the United States. We combine establishment-level data on 6,063 union elections with transaction-level data on 357,436 campaign contributions to federal and local candidates over the 1980-2016 period. In the campaign contribution data we observe the employer, occupation and address of individual donors which allows us to match donors of different occupations with the union election results of their employing establishments. To estimate the causal effects of unionization we compare campaign contributions of employees in establishments where workers voted for unionization with establishments that voted against unionization in a difference-in-differences (DiD) framework (tests of the underlying parallel trends assumption and alternative sources of exogenous variation are described below). We assess the political effects of unions by examining political mobilization—captured by changes in employees' total contribution amounts, and ideological shifts—captured by changes in the party composition of candidates they donate to. Linking these outcomes to union elections at the establishment level offers various new opportunities for studying the political influence of unions.

To start with, it enables us to analyze the political effects of unions at the workplace, where unions directly engage with employees and where not only members but also non-members may be affected by unionization. Exploiting the occupational information in the campaign contribution data, we can differentiate the political responses of workers and managers and study within-firm dynamics that have been previously ignored. We first ask how workplace unionization alters the political behavior of workers. Kerrissey and Schofer (2013) have argued that unions provide their members with political capital—they inform, engage, and mobilize members. Unions spend substantial resources on outreach and political education of their members. Most unions have newspapers and/or websites

[1] Unions draw on significant resources to finance their political outreach. In 2010, labor unions in the United States employed over 3,000 full-time political workers and spent 700 million USD on political activities, a figure that rose to 1.8 billion USD in 2020 (WSJ, 2012; NILRR, 2021).

that seek to inform members about topics relevant for their working conditions. They frequently hold meetings and workshops in which union members learn and exchange political views (Ahlquist and Levi, 2013; Iversen and Soskice, 2015; Macdonald, 2021). Moreover, employee gatherings, voting for union officers, participation in hiring halls, and joint strike activities can improve communication networks between workers and create social experiences that transform them into more engaged citizens (Lindvall, 2013; McAdam et al., 2001; Terriquez, 2011).

Through these mechanisms workplace unionization can increase union members' participation and support for Democratic candidates, which is consistent with our results. Our DiD estimates show that total campaign contributions from workers rise by 11% in response to unionization. This effect only shows up in the cycle of the union election and suggests a short-term political mobilization of workers through a successful union campaign at the workplace. Most importantly, when we examine the party composition of contributions, we find that unionization increases the percentage difference in donations from workers to Democrats versus Republicans by 12 percentage points in the six years following a union election. This result indicates a lasting shift in workers' ideological positions towards the political left.

Focusing the analysis on union members only would ignore an important out-group—the firm's management—that can alter any conclusion regarding the overall political impact of unions. Managers do not form part of workers' bargaining unit but may be indirectly affected by unionization in different ways. On one hand, labor unions may foster the management's understanding of worker issues and lead to an alignment of ideological positions. Unionization establishes rules for the bargaining between managers and workers and may thus increase both the quantity and quality of communication between the two groups. Labor unions give workers a voice, as they enhance the formation and communication of workers' preferences and present them on an equal footing (Freeman and Medoff, 1979, 1984). Contact theory suggests that this increase in cooperative interactions can enhance perspective-taking and reduce worker stereotypes held by management (e.g., Allport, 1954). Furthermore, unions aim to establish fairer rules at the workplace, for example through introducing formal grievance systems and ensuring representation of workers in the board of directors, which can itself lower tensions between the management and workers (Verma, 2005).[2] On the other hand, labor unions might cause a backlash from the management. Representing the interests of firm owners, managers typically are profoundly hostile to unionization.[3] The increase in bargaining power for

[2]Ash et al. (2019) find that giving workers more authority through entitlements in collective bargaining agreements reduces labor conflicts, as measured through the frequency and intensity of strikes after negative wage shocks.

[3]In the run-up to union elections, employers frequently hire anti-union law firms and consultants, try to delay the election process, hold meetings in which employees are obligated to listen to the anti-unionization arguments, and—although legally restricted—threaten employees with dismissals and establishment closures (Flanagan, 2007; Freeman and Kleiner, 1990a; Kleiner, 2001; Logan, 2002; Schmitt

workers implies a loss of status and power for the management. A large psychological literature has revealed that tensions between groups can increase if one feels threatened by the other (e.g., Sherif et al., 1961; Campbell, 1965). Labor unions could thus increase the salience of labor conflicts. If true, that may increase polarization, as groups tend to adopt the stereotypes of the salient identity (Bonomi et al., 2021).

Overall, ex ante it is not clear whether labor unions are able to persuade managers or whether they enhance the management's opposition to workers' political positions. Occupational information in our campaign contribution data allows us to directly estimate that effect. Our results suggest a leftward shift in campaign contributions not only for workers but also for managers: unionization increases the relative difference in managers' donations to Democratic vs. Republican candidates by 20 percentage points, while it does not affect their total spending. These patterns are not in line with an increase in tensions between unionized workers and their management, but rather point toward a convergence of ideological positions.

Combining an establishment-level political outcome with variation in unionization after union elections also provides us with plausible identification strategies to identify the causal impact of unionization on the political behavior of employees. Given that we only consider establishments with union elections, i.e., where workers have shown an interest in unionization, our sample can be expected to be more similar than a random sample of establishments. Within that sample, we compare campaign contributions from establishments where workers voted for unionization with establishments that voted against unionization by estimating a stacked DiD model. The stacked DiD accounts for issues arising in a setting with staggered treatment timing and heterogeneous treatment effects (Goodman-Bacon, 2021).[4]

The DiD design relies on the assumption that campaign contributions in losing establishments would have developed in parallel to campaign contributions from winning establishments in the absence of unionization. The plausibility of that assumption is validated through complementing the DiD framework with tests originating from a regression discontinuity design (RDD) and with a novel instrumental variables (IV) approach. First, we test whether changes in outcomes are correlated with the pro-union vote share among the establishments that lost the union election. Since the treatment status discontinuously changes at the 50% threshold, there should be no differential trends among establishments with different vote shares below 50%. Indeed, we do not find any evidence for differential changes across different vote-shares, which helps us to rule out the possibility that any sizeable confounding factors correlated with the pro-union vote share and the timing of the election drive the results. Second, we restrict the sample to estab-

and Zipperer, 2009).

[4]We also check the robustness of our results to employing different DiD estimators introduced by the literature for a setting with staggered treatment timing (Borusyak et al., 2021; Callaway and Sant'Anna, 2021), which replicate our main results from the stacked DiD model.

lishments with increasingly close elections that are more likely to follow similar trends in contribution patterns. Our results are robust to a wide range of vote share bandwidths around the 50% cutoff, even when focusing on elections decided by only a 5-10% margin. Finally, we complement the DiD with arguably exogenous variation in union support from random shocks to the salience of workplace safety that are triggered by unexpected spikes in sector-level fatal work accidents shortly before the union election. The DiD-IV results support our main findings.

The effects of unionization on workers' and managers' campaign contributions at the establishment level could be explained by a change in the composition of the employed workforce and management. In order to differentiate between compositional and individual-level effects, we exploit the fact that we can use donor identifiers to track each donor's campaign contributions over time. We develop two specifications. First, we take out any direct effect of unionizing on contributions and focus only on compositional changes. We compare contribution patterns before the union election for donors that donated after the election in establishments where the union won relative to establishments where the union lost. We do not find any sizeable effect. Second, we study individual-level effects by restricting our sample to individuals who were employed at the establishment before and after the union election and donated before and after. We find a significant leftward shift in donations for workers as well as managers. In sum, these results are consistent with labor unions persuading members and their management to support Democratic candidates.

To study a potential mechanism underlying this result, we examine the role of Right-to-Work (RTW) laws under which employees at unionized establishments do not have to pay union fees to reap the benefits of union representation. Feigenbaum et al. (2018) provide evidence that RTW laws put pressure on union revenues, forcing unions to shift scarce resources from political activities into membership recruitment activities and have aggregate consequences in terms of reduced turnout as well as fewer votes for Democratic candidates at the county level. Building on their analysis, we study how RTW laws affect the political responses of employees to unionization at the establishment level. We find the positive effects of unionization on contributions from workers and managers to Democratic versus Republican candidates to be smaller in states with RTW legislation. This finding highlights the role of unions' mobilization activities for their ability to raise support for their political agenda.

Finally, our data enable us to move beyond party preferences by considering candidates' ideological positions and the support of interest groups. We document considerable within-party variation in the effects on contributions to different candidates. Liberal candidates gain and conservative candidates lose, while moderate candidates are not significantly impacted on average. This suggests that our findings are not only driven by an increased signal of Democratic versus Republican partisan affiliation but reflect shifts between candidates with clearly distinguished ideological positions. In addition, we show that our results extend to contributions to Political Action Committees (PACs). In particular, we find that unions are able to mobilize workers, increasing their donations to labor and membership PACs. At the same time, unions decrease managers' contributions to corporate PACs. The increased support for labor and civil society interest groups from workers and the reduced support for business interest groups from managers match with the observed pro-liberal shift in their contributions to candidates.

Our results contribute to several strands of literature. First, we complement the literature on the economic impacts of unions by providing insights on the political channel. Several studies have assessed the impact of unionization on wages and employee compensation at the establishment level (DiNardo and Lee, 2004; Frandsen, 2021; Freeman and Kleiner, 1990b; Knepper, 2020). These studies document an absence of large wage effects but some positive effects on fringe benefits. The limited establishment-level effects are difficult to reconcile with evidence on the aggregate economic effects of unions. Stansbury and Summers (2020) show that declines in worker power can explain the entire decrease in the labor share of income in the U.S. over the last decades. Moreover, Western and Rosenfeld (2011) and Farber et al. (2021) document negative effects of unions on income inequality, which they argue is hard to explain by income changes of union members alone, suggesting a potential link between unions and distributional legislation.[5]

Second, we speak to the literature on the direct political influence of unions on their members. By comparing union members to non-union members, several studies have documented a significant association with political outcomes, such as voting (Freeman, 2003; Leighley and Nagler, 2007), preferences for redistribution (Mosimann and Pontusson, 2017), and trade liberalization support (Ahlquist et al., 2014; Kim and Margalit, 2017).[6] We add to these studies by assessing the causal impact of unions on campaign contribution patterns of workers. Campaign contributions are viewed as essential for candidates to win elections. Their influence on the set of candidates who run and win elections has been documented (e.g., Bekkouche and Cagé, 2018; Schuster, 2020). Moreover, donors prefer to give to ideologically proximate candidates on average, such that campaign contribution patterns reveal the political ideology of donors (e.g., Bonica, 2014, 2018). An assessment of campaign contribution patterns can therefore highlight the influence of unions on an important input into the political process and permits conclusions

[5] Several studies point toward an important role of unions in promoting greater political representation of the working class. Sojourner (2013) shows that workers' likelihood of serving as state legislator increases with their occupation's unionization rate. Moreover, local union density is correlated with a more equal legislative responsiveness toward the poor vs. the rich (Flavin, 2018; Becher and Stegmueller, 2021). See also Ahlquist (2017) for a review on how unions affect economic and political inequalities.

[6] Union membership is also related to social attitudes more broadly, such as lower racial resentment (Frymer and Grumbach, 2021) and a stronger identification with the working class (Franko and Witko, ming).

about shifts in political ideology.[7]

Third, we shed new light on the spread of political preferences at work through combining establishment-level union election data with an individual-level political outcome. The existing literature on the political impact of unions has focused either on individual union members and their households (e.g., Freeman, 2003) or on aggregate outcomes comprising the whole county or state population (e.g., Feigenbaum et al., 2018). By focusing on the unionizing workplace, we are the first to consider within-firm dynamics and, in particular, the reaction of management—the out-group that is likely indirectly affected by unionization and a key actor when it comes to political influence. Thus, we relate to studies documenting contagion effects in political behavior in general (e.g., Nickerson, 2008), spillovers in political donations between managers and workers (Babenko et al., 2020; Stuckatz, 2022), and effects of intergroup contact at the workplace on political preferences (Andersson and Dehdari, 2021).

The paper is organized as follows. Section 3.2 describes the institutional background, while Section 3.3 introduces the data. The empirical approach is outlined in Section 3.4, after which Section 3.5 presents the results. We explore potential mechanisms and extensions in Section 3.6 and conclude in Section 3.7.

3.2 Institutional Background

3.2.1 Unionizing through NLRB Elections

Since 1935, the National Labor Relations Act (NLRA) gives most private-sector workers in the U.S. the right to organize in unions and take collective action, such as bargaining and strikes. Collective bargaining between unions and employers takes place at the establishment level. Traditionally, workers unionize through a secret ballot election at their establishment that is administered by the National Labor Relations Board (NLRB).[8] The unionization procedure involves three main steps: a petition drive, an election, and certification.[9]

[7]Thus, we also contribute to the broader literature on political preference formation. It has been shown that context is a significant determinant of political behavior (Cantoni and Pons, 2022), while individual factors like early life experience (Jennings and Niemi, 2015) and education (Cantoni et al., 2017) are also important. Our results highlight workplace unionization as one influential contextual factor that shapes political preferences.

[8]While union elections are the primary means by which private-sector workers gain union representation, there are alternative procedures for unionization. First, employers may voluntarily recognize unions without an election through neutrality agreements and "card checks". These cases are less common, however, since employers generally oppose union organization (Schmitt and Zipperer, 2009). Second, some workers' bargaining rights are not regulated by the NLRA. For example, the *Railway Labor Act* determines bargaining rights of airline and railroad workers and several federal, state, and local laws regulate the organization of public-sector employees.

[9]The description of the unionization process follows Frandsen (2021) and Wang and Young (2021).

The organizing drive can be initiated either by the workers at an establishment or by a union organization. The initiator first needs to gather the signatures of at least 30% of workers in the proposed bargaining unit who thereby express a desire for unionization. With these signatures, an election petition is filed to the NLRB. The NLRB decides whether to accept the petition by ascertaining whether workers in the proposed bargaining unit share common interests that can be adequately represented by the union. If the petition is accepted, the NLRB schedules a secret ballot election, which usually takes place at the workplace. The union wins the election if it obtains a strict majority of the votes cast. In case of union victory, the NLRB certifies the union as the sole authorized representative of employees in the bargaining unit.

Union certification requires the employer to bargain "in good faith" with the union. This bargaining generally aims at concluding a first contract between union and employer. While there is no legal obligation to reach such an agreement, evidence suggests that in 55-85% of winning elections a first contract is reached within three years of the election (CRS, 2013). When both parties cannot reach a first agreement (or when subsequently they are disputing over the terms and conditions of the first contract), they can consult a neutral third party to resolve disputes via mediation or arbitration. After one year has passed since certification, employees can also decide to hold a decertification election to vote out the union.

The NLRA also lays out which employees may form a bargaining unit. While a bargaining unit can generally include all professional and nonprofessional employees at an establishment, managers and supervisors are always excluded.[10] These employees are considered to be part of a firm's management rather than its labor force and can therefore not join a union or be part of a bargaining unit. Representing the interests of capital owners, managers and supervisors typically oppose unionization and are thus treated as the "out-group" in our analysis. All other occupations form the "in-group", as they are potentially in the bargaining unit and directly benefit from unionization.

3.2.2 Campaign Contributions in U.S. Politics

Money plays a dominant role in U.S. politics. Monetary resources are viewed as essential for political candidates in order to take part and be successful in the political process. There is indeed increasing evidence that campaign donations can influence who runs for and who wins elections (e.g., Barber, 2016b; Bekkouche and Cagé, 2018; Schuster, 2020). While much of the public debate on campaign finance regulations centers around donations from corporations and other interest groups, the large majority of campaign

[10]The NLRA uses a rather broad definition for supervisors. It includes all individuals who have the authority to assign and direct the work of other employees, as long as this involves some independent judgment. There is no restriction as to the actual share of working time that involves supervisory duties. See Appendix B.3 for details.

contributions in the U.S. actually comes from individual donors. For the 2020 congressional elections, 77% of the total money received by candidates came from individuals. This share increased over time from 55% in the 2002 elections (FEC, 2022a). While political spending is certainly concentrated among the wealthy (Bonica and Rosenthal, 2018; Hill and Huber, 2017), it is a prevalent form of political participation for a substantial share of the U.S. electorate. Bouton et al. (2022) estimate that 12.7% of the adult U.S. citizen population have made at least one campaign contribution between 2006 and 2020.

Unlike corporations, which are prohibited by U.S. federal law to support candidates directly out of treasury funds, individual donors are allowed to make direct contributions to political candidates.[11] There are, however, restrictions to the maximum amount that an individual can donate to a candidate. The limit varies by recipient type and election cycle. For the 2018 federal elections, for example, individuals were allowed to donate at most 2,700 USD to a single candidate and 5,000 USD to a PAC (Whitaker, 2018). Recipients are obligated to itemize all individual contributions greater than 200 USD and report the donor's identifying information along with the amount and date of the contribution. Donations smaller than 200 USD are not required to be itemized but are included in the total amount that the recipient reports to the Federal Election Commission.

Political scientists differentiate between two broad motivations for why individuals contribute to political candidates. First, contributions can be seen as consumption goods that give individuals consumption value from participating in politics and sponsoring candidates that are ideologically close to their own political position (Ansolabehere et al., 2003). Second, donors may view contributions as investment goods that can buy access to politicians and benefit their own material interests. There is extant evidence that individuals' donations are ideologically motivated. Individual donors self-report that candidate ideology has great importance when deciding to whom to give (Barber, 2016a). Moreover, in comparison to access-seeking PACs, who prefer donating to moderate candidates, individuals tend to support more ideologically extreme candidates (Barber, 2016b; Stone and Simas, 2010). In merged survey-administrative data, contribution-based ideology measures are also found to predict policy preferences of donors, even of donors from the same party (Bonica, 2018). While for the rank-and-file there is consistent evidence in line with ideology being the main driver of political spending, for corporate elites the motivations are more debated. Teso (2022) shows that a business leader's likelihood of donating to a Congress member increases when the politician becomes assigned to a committee that is policy-relevant to the business leader's company. Based on the estimates, however, Teso (2022) concludes that only 13% of the observed gap in donations to policy-relevant versus other politicians is driven by an influence-seeking motive in line with corporate elites lobbying on behalf of their company. Moreover, Bonica (2016) finds that donations

[11] To make campaign donations, companies must set up a PAC, which may only solicit contributions from the firm's employees. The PAC can in turn donate directly to political candidates or other recipients.

from corporate board members are ideologically quite diverse, both across and within companies. Compared to corporate PACs, business leaders also tend to support more non-incumbent candidates and less powerful legislators. In summary, the evidence suggests that individuals primarily donate to candidates for ideological reasons. A number of papers have therefore interpreted changes in campaign contribution patterns as indicators of changes in political ideology (e.g., Autor et al., 2020; Bonica et al., 2016; Dreher et al., 2020).

3.3 Data

Previous studies have been unable to assess the political impact of unions at the establishment level due to a lack of matched employer-employee data for political outcomes. Campaign contribution data are uniquely suited to overcome this constraint. To ensure transparency in politicians' campaign funds, contributors are required to disclose their name, employer, address, and occupation. The employer and location information allows us to link donors to the union election results of their employers. We are not aware of any other large-scale data on political behavior with employer information in the U.S. that would allow this link. Furthermore, we can use the occupation information to study the political effects of unionization not only on directly affected non-managerial workers but also on potentially indirectly affected managers and supervisors. In the following, we describe how we construct a new establishment-level dataset that links union elections to campaign contributions from employees.

3.3.1 Union Elections

We start with a comprehensive dataset on the universe of U.S. union representation elections between 1961 and 2018. Specifically, we combine data collected by Farber (2016) with public data from NLRB election reports.[12] Each data point represents a union election at a single establishment and contains vote counts for and against unionization, the dates of the petition filing and of the actual election, as well as the name of the union organization. Moreover, it includes the establishment's name and address, which we exploit to match campaign contributions.

Sample restrictions. Before matching elections to campaign contributions, we impose several sample restrictions.[13] First, we only consider elections held between 1985 and 2010. Given that our contribution data cover the years 1979-2016, this allows us to

[12]We obtain the dataset originally assembled by Farber (2016) from the replication package of Knepper (2020). The data contain information on elections held between 1961 and 2009. For elections between 2010 and 2018, we retrieve data from NLRB election reports available on https://www.nlrb.gov/reports/agency-performance/election-reports.

[13]See Appendix B.1 for details on the union election data and the sample restrictions we impose.

observe trends in contributions for three political election cycles (six years) before and after each union election. Second, we follow Frandsen (2021) and restrict the sample to union elections where at least 20 votes were cast. This restriction ensures that winning establishments are affected by a non-trivial rise in union representation. Moreover, it helps to exclude small establishments, which are more likely to have come into existence recently and have a lower probability of survival over our period of analysis. Third, following Knepper (2020) and Wang and Young (2021), we only keep the first union election in each establishment.[14] Excluding non-inaugural elections avoids having multiple observations for the same establishment with reversed treatment status over time, and helps alleviate election manipulation issues if managers or unions learn how to apply manipulation tactics in repeat elections. Our estimates should thus be interpreted as the effects of winning the first union election.[15] These restrictions leave us with a sample of 28,823 union elections, which we seek to match to the campaign contribution data.

Table 3.1 – Election and Contribution Descriptive Statistics

	All	Union Loss	Union Win
	[A] Election characteristics		
Number of elections	6,063	3,397	2,666
Union vote share (average)	.4950	.3204	.7175
Number of votes (average)	119.37	135.31	99.06
Number of votes (total)	723,752	459,661	264,091
	[B] Contribution characteristics		
Amount (total, in million 2010 USD)	105.82	65.38	40.43
Number of contributions (total)	357,436	204,797	152,639
Number of donors (total)	46,719	26,661	20,243
Number of recipients (total)	9,942	7,208	5,681

Notes: Data from NLRB union certification elections, which have at least one employee contribution matched in any of seven election cycles around the union election (three before, cycle of union election, three after). Contribution characteristics refer to the total numbers over all these seven election cycles.

Summary statistics. Table 3.1 shows summary statistics for characteristics of the matched union elections that are included in our final estimation sample (see details on the matching in the next subsection). 44% of the elections were won by the union, with

[14] In the election data, we identify an establishment as a unique address or a unique combination of the standardized firm name and commuting zone. For a firm that has multiple establishments within the same commuting zone, we thus only consider the first election among these establishments.

[15] This does not perfectly correspond to the effect of union representation in all post-election periods for two reasons. First, establishments may lose representation after a decertification election. Wang and Young (2021) show that 5-10% of establishments that win a first union election hold a decertification election within 5 years. Second, establishments, after losing the first election, can hold another successful election in subsequent years. According to DiNardo and Lee (2004), this is the case for around 10% of lost first elections. By focusing on the effect of winning the first election, we thus accept an attenuation of our estimates relative to the true effects of union representation over all post-election periods.

an average union vote share of 50%. On average, 119 votes were cast in each election, which yields a total of 723,752 voters who participated in all elections of our sample.

3.3.2 Campaign Contributions

To measure the political mobilization and ideology of employees, we use the Database on Ideology, Money in Politics and Elections (DIME) compiled by Bonica (2019).[16] DIME provides transaction-level data on campaign contributions registered with the Federal Election Commission and other state and local election commissions. We exploit all campaign contributions from individuals to candidates running for office at the federal and local level (specifically the House of Representatives, Senate, President, Governor, and upper and lower chambers of state legislature), as well as to all PACs (including single-party or single-candidate and interest-group PACs). The dataset covers the 1979-2016 period and includes the amount and exact date of the donation, as well as identifying information on the donor and recipient.[17]

Bonica (2019) deploys identity resolution techniques to assign unique identifiers to each donor. The identifiers allow us to track donors' contributions over time, which we exploit to study whether establishment-level effects are driven by compositional changes from leaving and newly hired employees or by individual-level effects on employees remaining in the firm. Further, the DIME includes measures for the political ideology of recipients and donors, so-called campaign finance (CF) scores, which are derived by Bonica (2014) from solving a spatial model of contributions. The model formalizes the idea that donors contribute more to candidates with a similar ideological position and estimates ideal points of both recipients and donors along a typical liberal-conservative scale. Using the ideology scores, we can go beyond previous papers that only relate unions to Democratic versus Republican party affiliation and study how unionization affects ideological contribution patterns for candidates within the same party.

Matching algorithm. We link the campaign contributions to the employing establishments with union elections by combining a spatial match with a fuzzy match of firm names. We start by restricting potential matches to the same local labor market using 1990 commuting zones. 92% of the population live and work in the same local labor market, making it very likely that a donor in our sample works at an establishment in the same local labor market (Fowler and Jensen, 2020). The restriction substantially reduces the computational requirements for the fuzzy match and ensures that for multi-

[16]Other papers have used these data to study, among others, the political consequences of import competition (Autor et al., 2020), immigration (Dreher et al., 2020), contribution limits (Barber, 2016b), advertising firms (Martin and Peskowitz, 2018), or consultant networks (Nyhan and Montgomery, 2015).

[17]Accurate reporting of the donor information (name, employer, address, occupation) is enforced by the Federal Election Commission through regular audits, as well as fines and further legal action in case of non-compliance. See FEC (2022b) for enforcement statistics.

establishment firms we do not incorrectly match employees to establishments of the same firm in other locations.[18] To match the employer name in the contribution data to the establishment name in the union election data, we use an automated record-linkage program introduced by Blasnik (2010) and Wasi and Flaaen (2015). The linkage process first standardizes employer names and then calculates bigram scores for the similarity of each string pair. Lastly, we manually review all matches with a score above a minimum threshold.[19] To arrive at an establishment-level panel of employee contributions, we sum up all matched contributions within an establishment and two-year election cycle. Our period of analysis covers three cycles before to three cycles after each union election. Out of the 28,823 elections that we started with in the matching process, we only include establishments for which we have at least one matched contribution over this period. This leaves us with an estimation sample of 6,063 (21%) matched establishments (and 42,441 establishment-cycle observations).[20] As Table 3.1 reports, our sample is built from 357,436 matched contributions that amount to 105.8 million USD spent by 46,719 different donors to 9,942 different recipients.

Classification of occupations. In order to differentiate between workers eligible for unionization and their managers and supervisors who are always excluded from the bargaining unit, we classify self-reported occupations of donors. Here, we only briefly describe the classification procedure and provide more details in Appendix B.3. We start by mapping the free-text occupation descriptions in the DIME to the 6-digit Standard Occupation Classification (SOC). For this, we combine an ensemble classifier called SOCcer (Russ et al., 2016), sub- and fuzzy string matching to an extensive crosswalk of laymen's occupation titles from O*NET, as well as manual reviews of the most common occupation titles. Appendix Figure A.1 shows the occupation distribution for the classified donations. While the largest share (44%) is given by donors in management occupations, we also see substantial shares of contributions originating from lower-tier white-collar occupations such as healthcare, education, culture and sports, or financial operations workers. Blue-collar occupations, in contrast, account for small shares of the overall number of contributions, which is not surprising given that wealth is a strong predictor of political donating.

With the classified SOC codes at hand, we categorize donors into managers and su-

[18]We accept measurement error from assigning donors to the wrong establishment if a firm has several establishments within a commuting zone. However, within-firm interactions may generate spillover effects across establishments. The results of Knepper (2020), for example, imply large spillovers in the effects of unionization on firm-level employee compensation.

[19]See Appendix B.2 for details on the matching process.

[20]Appendix Table A.1 compares characteristics of matched and non-matched establishments. Elections in our matched sample involve more voters, i.e., are likely to be larger, and tend to be held in more recent years as contribution numbers have sharply increased over time. At the same time, the matching does not strongly affect the selection of union elections in terms of voting outcome and industry composition.

pervisors versus non-managerial workers. We identify managers and supervisors by using all contributions from "Management Occupations" (SOC group 11) and adding all occupations that involve a significant amount of supervising following the NLRA definition of supervisor tasks and leveraging occupational task descriptions from O*NET. Non-managerial workers are then defined as all remaining donors to whom we were able to assign a SOC code. The occupational composition in our final sample of candidate contributions looks as follows: 42% of contributions originate from managers and supervisors (hereafter only termed "managers"), 30% from non-managerial workers (hereafter only termed "workers"), and for 28% we are unable to obtain an occupational classification. Due to the non-negligible share of unclassified occupations, we report results not only separately for managers and workers, but also for all employees together (including those without a classification).[21]

Table 3.2 – Contributions by Donor and Recipient

Recipient: \ Donor:	All employees	Workers	Managers
All	2,493.24	313.80	1,339.38
Candidates	1,181.96	173.42	594.44
Democratic candidates	575.85	112.79	261.76
Republican candidates	586.98	56.61	320.66
Political action committees	1,311.28	140.38	744.94
Party/candidate PACs	364.92	52.52	192.77
Interest-group PACs	937.22	86.37	549.31

Notes: The table reports mean values for the amount contributed in each of the 42,441 establishment-cycle combinations in the estimation sample. All amounts are in 2010 USD. Values are reported separately for contributions from all employees, from only non-managerial workers ("workers"), and from only managers and supervisors ("managers"). The difference in the amounts from all employees and the total from workers and managers is driven by contributions for which we were unable to classify the occupation.

Summary statistics. Table 3.2 reports mean values for the sum of all employees' contributions for a given establishment and election cycle. Managers donate on average 1,339 USD per cycle, while workers contribute 314 USD. Both groups support different recipients. The majority of contributions by managers are donated to Republican candidates (54%), whereas workers tend to favor Democratic candidates (65% of the average amount is donated to Democrats). Moreover, managers give a larger share of donations to committees than to candidates. In contrast, workers more often contribute directly to candidates.[22]

[21] In Appendix B.3 we also provide evidence that the likelihood of having a missing occupation classification is not affected by unionization and therefore unlikely to drive our results.

[22] To compare the contribution pattern of employees to those of unions, we also track campaign

Definition of outcome variables. In our analysis, we will consider two main outcomes of employees' political behavior at the establishment level. The first one is the total amount of campaign contributions to all political candidates which we interpret as a measure of political participation and mobilization of employees. We use the inverse hyperbolic sine (IHS) transformation to approximate log changes in contribution amounts, while retaining zero values.[23] Our second main outcome is the difference between the IHS-transformed contribution amounts to Democratic and Republican candidates. This measure approximates the percentage difference in support for Democrats versus Republicans. Given the extant evidence on ideological motivations driving individuals' donation behavior, we interpret it as a measure of employees' ideological positions.

3.4 Empirical Strategy

We aim at estimating the causal effect of unionization on the political participation and ideology of employees. A simple comparison of individuals in unionized and non-unionized workplaces will fail to account for differences between these groups along a number of dimensions. These arise because the decision to unionize is likely endogenous and correlated with many characteristics, among them potentially political behavior. Figure 3.1 depicts average campaign contribution amounts across winning and losing union elections before and after the election. Due to their shared interest in a union election at the same time, these establishments are expected to be more similar than a random sample of unionized and non-unionized establishments.[24] Pre-existing ideological differences are nevertheless visible: Workplaces that vote for unionization donate more to Democratic candidates and less to Republican candidates even before the union election.

To account for pre-existing differences, we implement a difference-in-differences approach and compare campaign contribution patterns before and after the union election in establishments where the union won versus where it lost. We complement the DiD design with methods from the RDD literature to probe the validity of the underlying parallel trends assumption. In particular, we exploit the fact that we observe the pro-

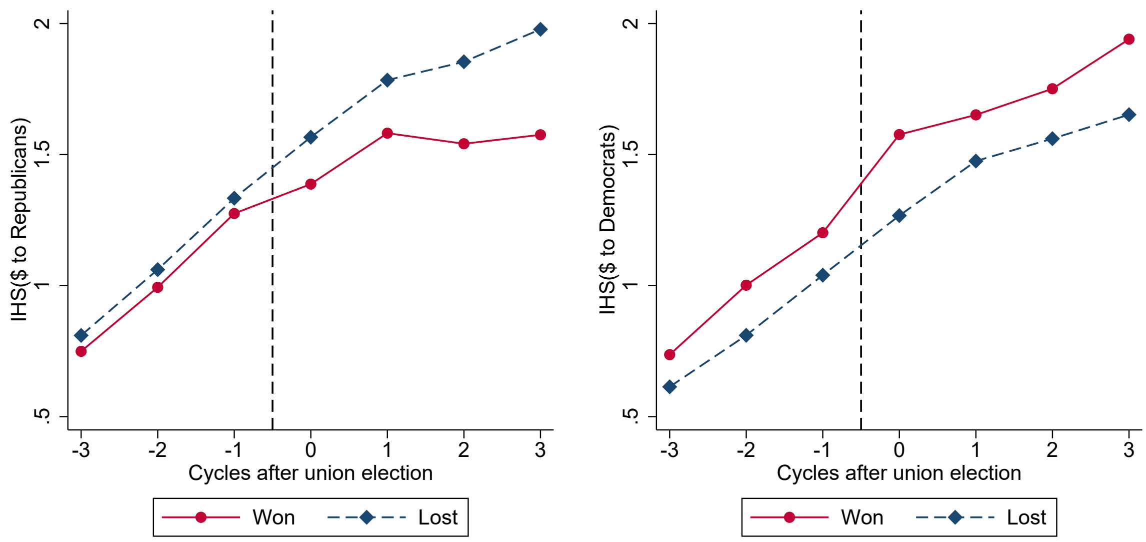
---

contributions originating from union organizations. Specifically, we consider all contributions from PACs associated with one of the unions in our matched sample, including local union branches. Appendix Table A.2 reports for each union the share of contributions to Democratic (as opposed to Republican) candidates as well as the ideology score obtained from Bonica (2014). On average, union PACs give 94% of their donations to Democrats, which demonstrates the strong link between labor unions and the Democratic Party.

[23] The inverse hyperbolic sine function is defined as $IHS(x) = ln(x + \sqrt{x^2 + 1})$. For sufficiently large x, $IHS(x) \approx ln(x) + ln(2)$. The function thus approximates the natural logarithm function for positive values but is also well defined for zero values. Applied econometrics papers frequently use it to transform non-negative variables with zeros (e.g., McKenzie, 2017; Bahar and Rapoport, 2018; Bursztyn et al., 2022).

[24] Dinlersoz et al. (2017) examine selection into union elections and find that elections are more likely to be held at younger, larger, more productive, and higher-paying establishments. Our strategy avoids such selection by comparing only establishments that hold union elections.

Figure 3.1 – Trends in Contributions for Won and Lost Union Elections



Notes: The figure depicts trends in mean contribution amounts of all employees in an establishment by union election outcome and election cycle (two years) relative to the union election. The left (right) graph shows means of IHS-transformed amounts to Republican (Democratic) candidates. $N = 42,441$ establishment-cycle observations.

union vote share, which discontinuously determines unionization at the 50% threshold. We use the vote share to estimate placebo tests for differential trends by vote shares among losing union elections as well as to examine the robustness of our DiD estimates when restricting the sample to establishments with increasingly close election results.[25] Finally, to cross-validate the causal interpretation of our results, we also develop a novel instrument which we apply in an identification strategy which combines the DiD with an IV approach. For this, we exploit variation in unionization resulting from exogenous shocks to the salience of safety at work that are triggered by unexpected fatal workplace accidents shortly before the union election.

Stacked DiD. As our main specification, we estimate the following stacked DiD model:

$$y_{ik} = \alpha_i + \beta_{kg_i} + \delta_{\text{DiD}} \times \Big(\mathbb{1}[k \geq 0] \times \mathbb{1}[V_i > .5]\Big) + \epsilon_{ik}, \tag{3.1}$$

where y_{ik} denotes a political outcome for employees in establishment i and relative event

[25]Many papers on the effects of unionization follow a RDD by comparing establishments in which the union barely won versus where it barely lost (e.g., Campello et al., 2018; DiNardo and Lee, 2004; Ghaly et al., 2021; Lee and Mas, 2012; Sojourner et al., 2015; Sojourner and Yang, 2022). This approach is complicated by the fact that unions and employers can influence election outcomes even after the election, through challenging the validity of individual ballots or filing charges of unfair labor conditions. Frandsen (2021) and Knepper (2020) provide evidence for discontinuities at the 50% threshold in the vote share distribution, as well as in pre-election establishment characteristics. Appendix Figure A.2 verifies that also in our matched sample of elections there is a significant discontinuity in the vote share density at the 50% cutoff, which indicates a manipulation of close elections.

time k. We observe each establishment from three cycles before to three cycles after the union election, i.e., $k = \{-3, -2, ..., 3\}$, where $k = 0$ refers to the cycle in which the union election takes place. Our effect of interest is captured by δ_{DiD}. It is the coefficient of an interaction term between a post-treatment dummy and a dummy indicating whether the election was won by the union, i.e., whether the pro-union vote share, V_i, is above 50%. α_i denotes establishment-fixed effects that capture all time-invariant differences between winning and losing establishments. Further, we introduce event-time $\times$ cohort-fixed effects β_{kg_i}, where cohort g_i refers to the political election cycle in which the union election was held, i.e., $g_i = \{1985/86, 1987/88, ..., 2009/10\}$. Importantly, with these fixed effects our identifying variation only comes from comparing changes across winning and losing elections within the same cohort. Thereby, it avoids "forbidden comparisons" between late and early-treated establishments that may lead to negative weights when averaging potentially heterogeneous, cohort-specific treatment effects in staggered DiD settings such as ours (de Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). Our DiD model is equivalent to the stacking approach first implemented by Cengiz et al. (2019). This approach first creates cohort-specific datasets of treated units and an appropriate set of control units that are never or not yet treated. Then, one stacks the cohort-specific datasets by time relative to treatment start in order to estimate an average treatment effect across all cohorts. By stacking and aligning cohorts in relative time, this strategy mimics a setting where all treatments occur contemporaneously, and thus avoids using already-treated units in the comparison group. Note that in our case the selection of appropriate control units for the stacking is facilitated by the possibility that we can naturally compare treated establishments to untreated establishments that have a lost election in the same cycle. Finally, we cluster standard errors at the level of treatment, the establishment.

Model (3.1) pools all periods after treatment, which yields the maximum power when estimating average treatment effects. To examine how treatment effects vary by event time, we also estimate the following stacked event-study model:

$$y_{ik} = \alpha_i + \beta_{kg_i} + \sum_{s=-3, s\neq -1}^{s=3} \delta_s \times \Big(\mathbb{1}[k = s] \times \mathbb{1}[V_i > .5] \Big) + \epsilon_{ik}, \tag{3.2}$$

where the δ_s coefficients capture dynamic treatment effects relative to the cycle before the union election was held (the interaction with $k = -1$ is omitted).

Parallel trends assumption. Our identifying assumption is that campaign contributions for winning establishments would have evolved in parallel to contributions in losing establishments had the union not won the election:

$$E[Y^0_{i,k\geq 0} - Y^0_{i,k<0} | V_i > .5] = E[Y^0_{i,k\geq 0} - Y^0_{i,k<0} | V_i \leq .5],$$

where Y_i^0 denotes the potential outcome of an establishment if the union loses the election.

We run different tests to examine the validity of this assumption. First, we analyze whether outcomes developed in parallel before the election. Figure 3.1 provides first visual evidence that pre-election changes in contribution amounts to Republican and Democratic candidates are very similar across winning and losing elections. The pre-election δ_s coefficients estimated in the event study model will provide a formal test of pre-trends.

Second, even in absence of significant pre-trends, there may still be unobserved shocks that drive union voting results at the time of the election and that may be related to changes in contribution patterns. To test whether such shocks likely violate our identifying assumption, we follow the approach of Wang and Young (2021) and analyze whether changes in outcomes are different among losing elections with different vote shares. If unobserved shocks were driving voting results that led to union victory or loss, we would also expect them to affect outcomes in losing elections with different union vote shares.[26] To implement this test, we modify the DiD model as follows:

$$y_{ik} = \alpha_i + \beta_{kg_i} + \sum_g \delta_g \times \Big(\mathbb{1}[k \geq 0] \times \mathbb{1}[V_i \in \nu^g]\Big) + \epsilon_{ik}, \tag{3.3}$$

where ν^g denotes a complete set of vote share categories. In particular, we divide the vote share distribution into the following six groups: 0-20%, 20-35%, 35-50%, 50-65%, 65-80%, 80-100%. In the model we omit the 20-35% vote share category, such that all estimated effects must be interpreted relative to that group. Significant estimates for the 0-20% or 35-50% categories would then indicate the presence of unobserved shocks that drive both voting results and campaign contribution behavior.

Third, we relax the parallel trends assumption by restricting the sample to elections where the union won or lost by an increasingly close margin. Establishments with closer election results can be expected to be more similar not only in terms of baseline characteristics but also in terms of shocks that they are exposed to over time. Specifically, we examine the robustness of the DiD estimates when restricting the sample to increasingly small vote share bandwidths around the 50% cutoff. In the limit, when comparing establishments where the union barely lost versus where it barely won, we approach the discontinuity-in-differences model estimated by Frandsen (2021) and Knepper (2020). For our baseline results from models (3.1) and (3.2), however, we follow Wang and Young (2021) and consider all elections with a pro-union vote share between 20% and 80%. This improves power and allows us to generalize effects for a broader sample of union elections.

Alternative source of variation. Lastly, we describe our DiD-IV approach, which

[26] Wang and Young (2021) formulate the identifying assumption as parallel trends across all vote shares, i.e., $E[Y^0_{i,k\geq 0} - Y^0_{i,k<0}|V_i] = E[Y^0_{i,k\geq 0} - Y^0_{i,k<0}]$, which yields the testable implication that trends should be parallel between losing elections with different vote shares.

complements the DiD strategy with arguably exogenous variation in union support driven by spikes in work-related fatalities. After the NLRB accepts a petition to hold a union election, it sets the timeline of the unionization process and fixes an election date. Any random unexpected shocks between petition and election that shift union support are then potential candidates for an instrument. We focus on sector-level fatal work accidents in the 30 days before a union election.[27] Safety at work is a fundamental concern to all workers, especially when one's life is in danger. Work-related fatalities are unfortunately still common in the United States. In 2019, the Occupational Safety and Health Administration (OSHA) reported 1,943 deaths at work, more than 5 per day on average. Unions often campaign on safety issues and are found to improve safety conditions at the workplace (e.g., AFL-CIO, 2022; Hagedorn et al., 2016; Li et al., 2022).

We implement the DiD-IV approach by estimating the following two-stage model:

$$V_i = \alpha_1 + \alpha_2 A_{st} + \alpha_3 A_{st}^2 + \alpha_4 A_{st} \times FR_s + \alpha_5 FR_s + \alpha_6 X_i + \gamma_t + \mu_m + \epsilon_i \tag{3.4}$$

$$\Delta y_i = \beta_1 + \beta_2 \mathbb{1}[\widehat{V_i} > .5] + \beta_3 FR_s + \beta_4 X_i + \gamma_t + \mu_m + \epsilon_i, \tag{3.5}$$

where Δy_i denotes the change in campaign contribution patterns in the three cycles after the union election relative to the three cycles before (excluding the cycle of the union election). By using changes as the outcome variable, the specification builds on the DiD approach and accounts for time-invariant differences between establishments that may affect the level of campaign contributions. Our main instrument is A_{st}, which represents the number of fatal accidents in 2-digit sector s in the 30 days prior to the election after accounting for seasonal variation.[28] We allow for a non-linear effect by including A_{st}^2 and for a larger impact of fatalities in sectors where fatalities are common and where workers may be more concerned about workplace safety by the interaction term $A_{st} \times FR_s$ (FR_s denotes the share of fatal work accidents occurring in a given sector out of all fatal work accidents in the sample). Importantly, instead of directly instrumenting union victory in a standard 2SLS approach, the first stage explains the continuous pro-union vote share V_i. In the second stage, we then use an indicator for predicted victory that is based on the predicted vote share in the first stage. This approach resembles the treatment assignment process and exploits the maximum available information. To account for the uncertainty from the first-stage regression, we compute standard errors using bootstrapping. In addition, we include a number of control variables. First, we account for the main effect of FR_s. Second, we include the yearly number of fatalities in a sector to ensure that sector-specific trends in fatalities do not drive our results. Third, we add the log number of employees at the sector-year level and the log number of eligible voters as precision

[27] The median time between petition and election in our sample is 47 days. Only 1% of all elections are held within 30 days after the petition.

[28] Data on fatal work accidents are obtained from OSHA in the form of Fatality and Catastrophe Investigation Summaries (OSHA form 170). Appendix Figure A.3 depicts the exploited time variation.

controls. Finally, we include year-fixed effects γ_t and month-of-the-year-fixed effects μ_m.

The exclusion restriction of the instrument relies on the notion that a shock in fatal work accidents in the same sector affects political behavior only through its impact on the likelihood that an establishment will unionize. Two points are worth highlighting in that regard. First, all individuals in our sample are potentially exposed to the information on fatal work accidents. However, only some vote on unionization in the following 30 days. That is to say, we do not exploit differences in the direct exposure to work accidents but differences in the timing of the union election relative to the information shock. Second, we are solely focusing on the medium-term impact of spikes in fatal accidents by considering campaign contributions in the three cycles after the union election. The result that common shocks in fatal work accidents influence political behavior years afterward in some but not other establishments would be difficult to explain other than through the path dependency triggered by the increase in the likelihood of unionization shortly after the accidents.

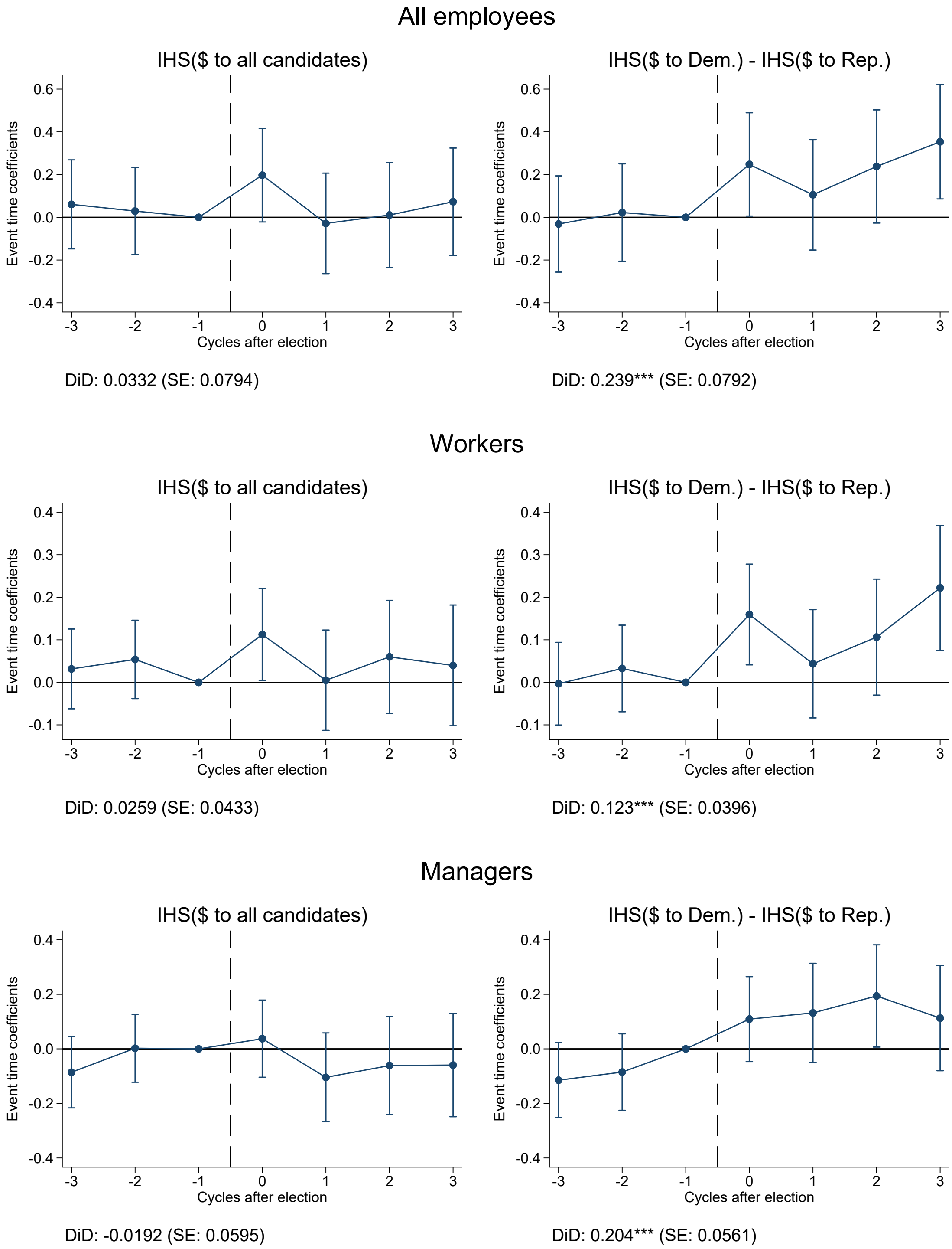
3.5 Results

3.5.1 Main Results

Figure 3.1 presents first descriptive evidence on the political impact of unionization by displaying trends in mean contribution amounts from all employees of an establishment to Republican and Democratic candidates. Before the election, contributions develop very similarly in establishments where union elections are won and where they are lost. The strong upward trend is explained by the fact that campaign contributions have strongly gained in importance in more recent election campaigns. At the time of the election, we see that contribution patterns start to diverge between winning and losing elections. The rise in donations to Republicans appears considerably smaller in unionized than in non-unionized establishments. In contrast, donations to Democrats seem to slightly increase in winning union election establishments relative to losing ones. Overall, the figure suggests a shift in contributions from Republican to Democratic candidates after successful unionization.

We now turn to our main estimation results from the DiD approach in which we estimate the effects of unionization on two outcome variables: the (IHS-transformed) total amount of campaign contributions to all political candidates, which measures the political mobilization of employees, and the difference between (IHS-transformed) contribution amounts donated to Democratic and Republican candidates, which allows to study shifts in employees' ideological positions. Figure 3.2 displays the pooled average treatment effect δ_{DiD} from the stacked DiD model (3.1) as well as the dynamic treatment effects δ_s from the stacked event-study model (3.2).

Figure 3.2 – Effect of Unionization on Candidate Contributions



Notes: The figures report the event-study coefficients δ_s estimated in model (3.2). The sample includes all establishments with a pro-union vote share between 20% and 80% and covers three election cycles (six years) before and after the union election. $N = 33,103$ establishment-cycle observations. Below each graph the DiD coefficient from model (3.1) is reported. In the graphs on the left side, the outcome is the IHS-transformed total amount contributed to all candidates. In the graphs on the right side, the outcome is the difference between the IHS-transformed amounts contributed to Democratic and Republican candidates. Results are reported for contributions from all employees (top part), from only non-managerial workers (middle part), and from only managers and supervisors (lower part). 95% confidence intervals are depicted for standard errors clustered at the establishment level.

We start with the effects on the total contribution amounts depicted on the left-hand side of the figure. The upper panel plots the results for all employees in an establishment. Note the absence of any significant differential trends between establishments winning and establishments losing the union election in the three cycles (six years) before the election. The effect of unionization on the amount of contributions is small and insignificant in all post-election periods, but we see a moderate spike in contributions in the cycle of the union election (which we are not able to estimate precisely, though). Differentiating between contributions made by workers and managers in the lower panels highlights that workers drive the increase in contributions. The event-study results indicate that unionization raises workers' contributions by 11% in the cycle of the union election (significant at the 5% level). This pattern is consistent with a short-term political mobilization of workers through a successful union campaign at the workplace. Overall, however, the DiD coefficients indicate that there is no significant average effect on the amount of contributions over the three cycles after a union election.

Next, we assess changes in the party composition of campaign contributions. If unions are able to change individuals' political views or mobilize different subgroups at the workplace, campaign contributions will shift to different candidates. The right-hand side of Figure 3.2 plots the effect of unionization on the difference between the amounts spent to Democratic versus Republican candidates. First focusing on all employees, we again see no differential trends in contribution composition before the election. After the election, however, there is a significant increase in contributions donated to Democratic relative to Republican candidates. The effect on partisan support appears to be strongest in the long term, i.e., six years after the election. The DiD coefficient indicates that, over all post-election periods, unionization increases the difference in contributions to Democrats versus Republicans by 24 percentage points (significant at the 1% level). Differentiating again between workers and management in the lower two panels reveals that the effect is driven similarly by both groups. Not only workers, but also managers significantly shift contributions from Republican to Democrat candidates in response to successful unionization. Quantitatively, the DiD estimates show that winning the union election increases donations to Democrats relative to Republicans by 12 percentage points for workers and by 20 percentage points by managers (both significant at the 1% level). These patterns are not consistent with an increase in tensions between unionized workers and their management but rather point toward an alignment of ideological positions.

3.5.2 Addressing Identification Challenges

DiD-RDD. We continue presenting results for our RDD-motivated tests to probe the validity of the underlying parallel trends assumption of the DiD model.[29] Figure 3.3 focuses on the measure of partisan contribution composition, while effects on the total amount of contributions are presented in Appendix Figure A.5. Results are always reported separately for workers and managers. We first analyze the heterogeneous effects of unionization across the vote share distribution. Panel (a) of Figure 3.3 displays the δ_g coefficients from model (3.3) on the interaction between the post-election dummy and different vote share categories. The results show that there are no significantly different trends among losing elections with a vote share of 0-20% or 35-50% relative to those with 20-35%, for contributions from both workers and managers. The post-treatment partisan contribution composition thus appears to evolve similarly across losing establishments with different vote shares. Therefore, we do not find evidence for unobserved shocks correlated with voting results that could drive our results.[30] Moreover, the results indicate whether treatment effects are heterogeneous across vote shares among winning union elections. For the composition of contributions from managers, the estimate is significant across all vote share categories above 50%. Thus, the political response of managers does not appear to depend on whether workers won the union election with large or small margins of victory. For workers, the effect on partisan support is significant only for vote shares between 50 and 80% and appears smaller for elections won by a large margin.

Panel (b) of Figure 3.3 presents coefficients from the DiD model (3.1) when restricting the sample to establishments with increasingly close election results. Establishments with more similar voting results can be expected to be more similar in other characteristics and to be exposed to more similar shocks, which makes the parallel trends assumption more plausible. Results are reported in 5% steps of the union vote share bandwidth around the 50% cutoff. Our baseline results from Figure 3.2 include only elections with a pro-union

[29]One particular concern for the parallel trends assumption would arise if union elections were endogenously timed around federal election dates. Appendix Figure A.4 investigates whether union elections follow political cycles. Across years with and without federal elections, there are no strong differences in the number of union elections held and the probability of winning a union election, in particular not around the week of federal elections. Thus, we do not see evidence that employers or unions successfully manipulate union election dates to change union support around federal election cycles.

[30]In Appendix Figure A.6, we also investigate whether pre-trends in the contribution composition are similar across the vote share distribution. For this, we estimate the following modified version of model (3.3):

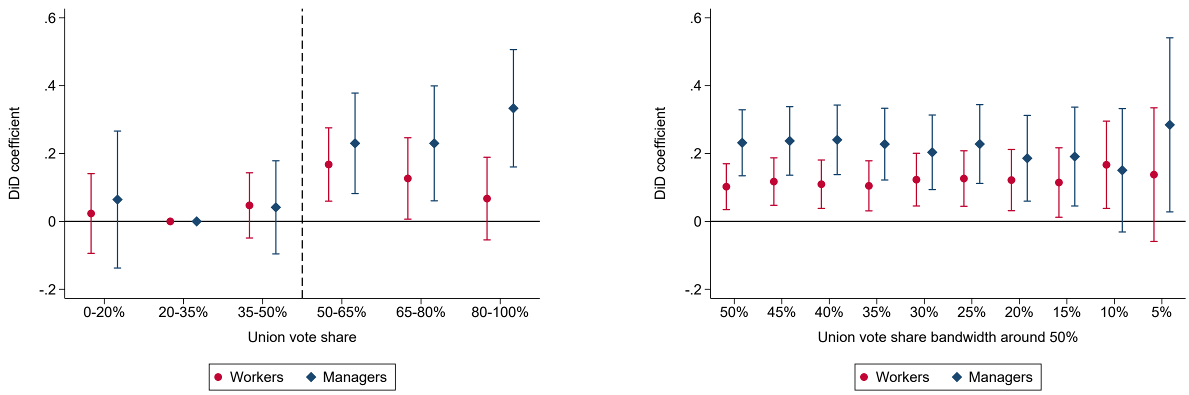
$$y_{ik} = \alpha_i + \beta_{kg_i} + \sum_g \delta_g^{PRE} \times \Big(\mathbb{1}[k < -1] \times \mathbb{1}[V_i \in \nu^g]\Big) + \sum_g \delta_g^{POST} \times \Big(\mathbb{1}[k \geq 0] \times \mathbb{1}[V_i \in \nu^g]\Big) + \epsilon_{ik} \quad (3.1)$$

The results show that none of the estimated δ_g^{PRE} coefficients are significantly different from zero, which indicates that also before the union election contribution patterns evolved similarly across establishments with different voting results.

Figure 3.3 – Effect of Unionization on Democratic versus Republican Support - DiD-RDD Results

(a) Vote Share Heterogeneity

(b) Vote Share Bandwidth Sample Restrictions



Notes: The graphs show RDD-type placebo and robustness tests for the effect of unionization on the difference between the IHS-transformed amounts contributed to Democratic and Republican candidates. Panel (a) reports the δ_g coefficients estimated in model (3.3). The vote share distribution is partitioned into six bins, indicated on the x-axis. The omitted reference group is 20-35%. Panel (b) reports DiD coefficients estimated in model (3.1). Each dot refers to a single DiD coefficient that is estimated among elections with a union vote share in a given bandwidth around the 50% cutoff. Estimates from smaller bandwidths compare changes between increasingly close elections. Results are always shown separately for contributions from non-managerial workers ("workers") and from managers and supervisors ("managers"). 95% confidence intervals are depicted for standard errors clustered at the establishment level.

vote share between 20 and 80%, i.e., a bandwidth of 30%. Figure 3.3 shows that treatment effects are very similar when instead using all elections. More importantly, the results are also very stable when focusing on closer elections. Even when restricting the sample to establishments that won with a maximum vote margin of 5%, we see a positive and significant effect on the composition of campaign contributions for managers. Similarly, for workers a maximum vote margin of 10% already yields a positive and significant effect.

DiD-IV. We also assess the sensitivity of our DiD results when exploiting arguably exogenous variation in unionization from shocks to the salience of workplace safety before the union election. Table 3.3 reports the results of our DiD-IV approach. The first-stage results show that sector-level fatal work accidents are a significant predictor of the union election outcome, with an F-statistic of 16.5. We find that the positive effect of spikes in work accidents on unionization is stronger in sectors where work accidents are more common, i.e., where workplace safety may be a greater concern for workers. The second-stage results confirm our main findings from the DiD model, highlighting a leftward shift in campaign contributions in response to unionization.[31] The magnitude of the

[31] We also report results when estimating model (3.5) by OLS. Given that the outcome is the change in outcomes before vs. after the election, the results are very similar to those obtained from our main DiD model (3.1). Small differences arise from the inclusion of additional controls in model (3.5) and the exclusion of the cycle in which the union election takes place.

coefficients is comparable but slightly larger than in the DiD model. As compliers respond to information on fatal work accidents, we deem it plausible that they also react more strongly to information provided by unions and to changes to their work environment induced by unionization. The estimates are considerably less precise, however. While the effects on the party composition of contributions from managers are still significant at the 5% level, the effects for workers are no longer significant. We thus use our DiD-IV approach to validate the main results and proceed with our main DiD model for the analysis of mechanisms in Section 3.6.[32]

Table 3.3 – DiD-IV Results

	IHS($ to all candidates)			IHS($ to Dem.) − IHS($ to Rep.)		
	All	Workers	Managers	All	Workers	Managers
	(1)	(2)	(3)	(4)	(5)	(6)
[A]: OLS						
$\mathbb{1}[V_i > .5]$	-0.092	0.038	-0.072	0.227***	0.089**	0.232***
	(0.082)	(0.044)	(0.062)	(0.079)	(0.041)	(0.056)
[B]: 2nd stage						
$\mathbb{1}[\widehat{V_i} > .5]$	0.036	0.086	-0.042	0.334*	0.115	0.260**
	(0.174)	(0.097)	(0.134)	(0.176)	(0.086)	(0.125)
[C]: 1st stage						
A_{st}	0.002	0.002	0.002	0.002	0.002	0.002
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
A_{st}^2	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$A_{st} \times FR_s$	0.223***	0.223***	0.223***	0.223***	0.223***	0.223***
	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)	(0.055)
K-P F-stat	16.50	16.50	16.50	16.50	16.50	16.50
[D]: 2nd stage falsification: pre-trend						
$\mathbb{1}[\widehat{V_i} > .5]$	-0.007	0.093	0.033	0.124	-0.020	0.046
	(0.207)	(0.094)	(0.116)	(0.230)	(0.100)	(0.129)

Notes: The table reports results from the DiD-IV approach for the effect of unionization on the IHS-transformed total amount contributed (columns (1) - (3)) and on the difference between the IHS-transformed amounts contributed to Democratic and Republican candidates (columns(4) - (6)). Panel A reports OLS coefficients, Panel B reports the second-stage coefficients from model (3.5), and Panel C reports the first-stage coefficients from model (3.4). In Panels A and B, the outcome is the difference between the average outcome in the three cycles after and the average outcome in the three cycles before the union election (excluding the cycle of the union election). In Panel D, the outcome is the change between one and two cycles before the union election. $N = 5,803$ establishments. Bootstrapped standard errors (with 500 replications), shown in parentheses, are clustered at the establishment level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[32]We also verify the DiD-IV approach with a falsification exercise. We re-estimate model (3.5) using the change in campaign contribution patterns between $t-1$ and $t-2$ as the outcome. We do not find any evidence for pre-existing differential trends related to spikes in fatal work accidents.

3.5.3 Robustness

We now discuss further robustness checks for our main DiD estimates. Results are presented in Appendix Tables A.3, A.4, and C.1.

Alternative staggered DiD estimators. The recent econometrics literature has proposed different methods to circumvent issues of treatment effect heterogeneity in staggered DiD designs. All the proposed estimation strategies have in common that they restrict the set of effective comparison units by ruling out the use of early-treated units in the estimation of treatment effects for currently-treated units. They differ, however, in terms of how exactly comparison units are identified and used in the estimation, as well as in terms of how cohort- or individual-specific treatment effect estimates are aggregated.[33] In Panels B and C of Appendix Table A.3, we present results from the imputation approach of Borusyak et al. (2021) and the estimator developed by Callaway and Sant'Anna (2021). The estimates are very similar to our stacked DiD results.

Alternative outcome transformations. Roth and Sant'Anna (2023) point out that the parallel trends assumption of a DiD design generally implies a functional form restriction on potential outcomes. Transformations of the outcome may imply different parallel trends assumptions. We therefore test the sensitivity of our results to alternative outcome transformations. First, instead of transforming contribution amounts with the IHS function, we use the log function and add one to the amounts to retain zero values. Second, we leave amounts untransformed (in 2010 USD). Results, shown in Panels D and E, yield qualitatively the same conclusions as the results for the IHS-transformed outcomes.

Alternative manager-worker classifications. In Appendix Table A.4, we check whether our results are sensitive to the exact definition of managers and supervisors versus non-managerial workers. To see whether the political response is different for lower- and upper-tier managers, we use more stringent definitions of managers/supervisors. First, we vary the cutoff for the importance of supervisor tasks (Panels B and C). Second, we only consider "Management Occupations" (SOC group 11) and treat all other occupations

[33] In our stacking approach of model (3.1), we effectively only compare winning elections to losing elections that were held in the same period, i.e., we only use never-treated units in the comparison group. The strategies by Borusyak et al. (2021) and Callaway and Sant'Anna (2021), in contrast, also allow including not-yet-treated units in the comparison group. Both approaches differ in that Borusyak et al. (2021) use the average pre-treatment outcome over all pre-treatment periods, whereas Callaway and Sant'Anna (2021) only use the outcome one period before treatment start. In terms of aggregation, Gardner (2021) shows that the stacking approach identifies a convexly weighted average of cohort-specific treatment effects where the weights are given by the number of treated units and the variance of treatment within each cohort. In comparison, Borusyak et al. (2021) and Callaway and Sant'Anna (2021) first estimate unit- or cohort-specific effects and then aggregate through a simple average across treated units. Callaway and Sant'Anna (2021) also allow other weights, but we use the default option where cohort-specific estimates are weighted by the number of treated units in each cohort.

(including those with a high importance of supervisor tasks) as workers (Panel D). The results do not change much with these alternative classifications. Even for more upper-tier managers unionization leads to an increase in the support for Democrats relative to Republicans.

Effects of losing a union election. Our DiD results measure the differential change in contributions from establishments where the union won versus establishments where the union lost the election. The observed relative shift in donations could not only be explained by the effects of unionization after winning the election, but also by an effect of holding and losing an election. Interaction with the union organization in preparation for the union election as well as a potentially increased salience of worker issues and distributional conflicts may affect the political behavior of employees, in particular in the short term, even if the union election is lost. We test this by estimating the effects of losing an election compared to holding no election. To avoid selection into which establishments hold and lose elections, we exploit only variation in the timing of union elections and use establishments that hold and lose an election in the future as control group. We implement this approach in a stacked DiD model similar to our baseline model (3.1).[34] Results are presented in Appendix Table C.1. We obtain small and insignificant estimates for our two main outcomes and for both workers and managers with a precision similar to our baseline results. This suggests that losing a union election can indeed be viewed as an untreated counterfactual and that our results are driven by the effect of unionization after winning a union election.

Overall, our estimates provide robust evidence that unionization changes the composition of employees' campaign contributions in favor of Democratic (relative to Republican) candidates. Importantly, this effect is found for both workers and managers.

3.6 Potential Mechanisms and Extensions

3.6.1 Compositional versus Individual-Level Effects

One potential explanation for the establishment-level effects may be compositional changes regarding what type of employees separate from and are newly hired into unionized establishments. Frandsen (2021) finds that unionization leads older and higher-paid workers to leave and younger workers to join union jobs. Separations and hirings may also be selective in terms of political ideology. For example, conservative union-avoiding managers may want to leave unionized workplaces and may be replaced with more liberal ones. If this is the case, our establishment-level results may be fully explained by composition

[34]See Appendix C for details of the stacking implementation. We also implement the DiD estimators by Borusyak et al. (2021) and Callaway and Sant'Anna (2021), which yield similar results.

effects rather than by individual-level changes in political behavior. To differentiate between the two, we exploit the donor identifiers in the DIME, which allow us to track donors' contributions over time.

Table 3.4 – Composition versus Individual-Level Effects

	Composition effects		Individual-level effects for stayers	
	IHS($ to all candidates)	IHS($ to Dem.) – IHS($ to Rep.)	IHS($ to all candidates)	IHS($ to Dem.) – IHS($ to Rep.)
	(1)	(2)	(3)	(4)
[A]: All employees				
δ_{DiD}	-0.0265	0.0705	0.196	0.552^{***}
	(0.0696)	(0.0636)	(0.135)	(0.188)
N	33,103	33,103	5,740	5,740
[B]: Workers				
δ_{DiD}	0.0455	0.0534^{*}	0.624^{***}	0.648^{**}
	(0.0363)	(0.0294)	(0.233)	(0.309)
N	33,103	33,103	2,052	2,052
[C]: Managers				
δ_{DiD}	-0.0666	0.0371	-0.0718	0.532^{**}
	(0.0514)	(0.0454)	(0.186)	(0.261)
N	33,103	33,103	2,890	2,890

Notes: The table reports DiD coefficients for the composition and individual-level effects of unionization on the IHS-transformed total amount contributed (columns (1) and (3)) and on the difference between the IHS-transformed amounts contributed to Democratic and Republican candidates (columns (2) and (4)). In columns (1) and (2), the establishment-level outcomes for the post-election periods are constructed from pre-election contributions from those donors matched to an establishment in the respective post-election period. Aggregates for the pre-election periods are constructed as before from the actual contributions in those periods. Columns (3) and (4) show results for individual-level regressions in a sample of donors who have a matched contribution to the same union election establishment at least once before and once after the union election. We aggregate all matched contributions into one pre- and one post-period observation and estimate a two-period DiD version of model (3.1) with individual and cohort × post-election-fixed effects. All samples include establishments / individuals from establishments with a pro-union vote share between 20 and 80%. Standard errors, shown in parentheses, are clustered at the establishment level in columns (1) and (2) and at the individual level in columns (3) and (4). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

First, we seek to examine pure composition effects. In other words, we take out any direct effect on individuals in unionized workplaces. For this, we modify the construction of our establishment-level aggregates of employee donations in the following way. For each post-election event time $k \geq 0$, we still consider the set of donors that have at least one contribution matched to the respective establishment in that period. Then, instead of using these donors' contributions in that period, we trace their contributions before the election (in the three pre-election cycles) and use them in the establishment-level aggregation. As a result, the post-election aggregates only reflect pre-existing contribution patterns. We use them along with the actual pre-election aggregates (constructed as before from the actual matched contributions in those periods) in our DiD model. Results, presented in Table 3.4, columns (1) and (2), show very small and almost always insignificant DiD estimates, indicating that the set of post-election employees does not differentially change in unionized versus non-unionized establishments in terms of pre-existing contribution amounts. Only for workers do we see a marginally significant estimate in line with more Democratic workers entering union jobs (or fewer Democratic workers leaving union jobs). The effect size, however, is much smaller than in our main estimates, which suggests that composition effects are unlikely to fully explain the results.[35]

Second, we aim at directly studying employee-level effects of unionization, i.e., we consider the direct effect of unionization on individuals. For this, we focus on a sample of individuals who are employed in the same establishment before and after the union election, which we identify as having at least one matched contribution to the same union election establishment at least once before and once after the union election. We then aggregate all matched contributions from these individuals over our 7-cycle window into one pre- and one post-election observation and estimate a two-period DiD (with individual and cohort $\times$ post-election-fixed effects).[36] Estimates are reported in columns (3) and (4) of Table 3.4. For all employees jointly, we find no significant effect on the total contribution amounts but a significant increase in the amount donated to Democratic relative to Republican candidates.[37] When restricting the sample to workers, we see a significant rise in total donations, which is, however, entirely driven by an increase in support for Democrats. For managers, the results indicate a significant shift from Republicans to Democrats without a change in total amounts. Overall, the results point to the conclusion that our establishment-level effects are driven by individual-level changes in donation patterns rather than by compositional effects.[38]

[35]Note that the compositional analysis is complicated by the fact that we only observe employees if they contribute. In principle, our compositional test may thus also pick up changes in the extensive margin in terms of which employees stop donating after the union election. As regards candidates' party affiliation, we would expect that unionization decreases [increases] the likelihood that employees stop donating to Democrats [Republicans]. Then, the extensive margin channel would yield a positive effect on contributions to Democrats relative to Republicans that post-election employees donated before the election, in line with what we expect for the actual compositional effect. Our results show that the sum of both effects is small, suggesting that both play a minor role.

[36]We refrain from aggregating contributions for each relative cycle k separately. Since we do not know an individuals' employing establishment if the individual does not donate in a given cycle, we are not able to construct a balanced panel over all cycles that includes observations with zero amounts.

[37]Note that the substantially larger magnitude of the estimates in comparison to the establishment-level results is likely because we have aggregated all pre- and post-election cycles for the individual-level analysis.

[38]Another composition effect potentially explaining our establishment-level result may arise from transitions of individuals across occupational groups. To rule out that the promotion of workers to management positions is driving our results for managers, in columns (3) and (4) of Table 3.4 we have classified individuals as *managers* only if they held a manager position both before and after the election. Individuals who have some matched contributions with an occupation categorized as *manager* and some categorized as *worker* are all included in the *worker* subsample.

3.6.2 Political Involvement and Ideology of Union Organizations

Unions are not a uniform political force, but can rather be understood as heterogeneous and evolving organizations that vary in their internal governance and institutional environments. Ahlquist and Levi (2013) and Kim and Margalit (2017) show that unions differ in the importance they place on political activities, in the intensity and form of communication with members, and in their policy views. We therefore seek to study the role that varying political activities and positions of union organizations play in moderating our results.

Table 3.5 – Heterogeneous Effects by Political Involvement and Ideology of Union Organizations

	IHS($ to all candidates)			IHS($ to Dem.) − IHS($ to Rep.)		
	All	Workers	Managers	All	Workers	Managers
	(1)	(2)	(3)	(4)	(5)	(6)
[A.1]: State without right-to-work law						
	0.0453	0.0663	-0.0394	0.284***	0.131***	0.218***
	(0.0896)	(0.0499)	(0.0672)	(0.0884)	(0.0456)	(0.0635)
N	26,208	26,208	26,208	26,208	26,208	26,208
[A.2]: State with right-to-work law						
δ_{DiD}	-0.0548	-0.119	0.00832	0.0164	0.0700	0.142
	(0.170)	(0.0820)	(0.125)	(0.177)	(0.0769)	(0.117)
N	6,895	6,895	6,895	6,895	6,895	6,895
[B.1]: More liberal union organization (below median CF score)						
δ_{DiD}	0.0250	0.00389	-0.0780	0.251**	0.0826	0.197**
	(0.115)	(0.0642)	(0.0857)	(0.116)	(0.0596)	(0.0837)
N	14,875	14,875	14,875	14,875	14,875	14,875
[B.2]: Less liberal union organization (above median CF score)						
δ_{DiD}	0.0864	0.0406	0.0416	0.240**	0.119**	0.186**
	(0.124)	(0.0615)	(0.0912)	(0.120)	(0.0543)	(0.0834)
N	14,882	14,882	14,882	14,882	14,882	14,882

Notes: The table presents DiD coefficients, estimated in model (3.1), for the effect of unionization on the IHS-transformed total amount contributed (columns (1) - (3)) and on the difference between the IHS-transformed amounts contributed to Democratic and Republican candidates (columns (4) - (6)). Panels A.1 and A.2 distinguish between establishments in states with versus without right-to-work laws in the union election year. Panels B.1 and B.2 report results for elections of union organizations with an ideology score below vs. above the median ideology score of all elections in our estimation sample. Unions' ideology scores are derived from Bonica (2014) and based on the campaign contributions that union organizations donate themselves (see Table A.2). All samples include establishments with a pro-union vote share between 20% and 80%. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

First, we examine how effects differ by whether or not unionization takes place under a state-level Right-to-Work law. RTW laws allow employees to enjoy the benefits of collective bargaining and union representation without having to become a union member and pay fees. Feigenbaum et al. (2018) study in detail the political consequences of RTW legislation. They provide evidence that RTW laws put pressure on union revenues and force unions to reallocate scarce resources from political activities (such as lobbying, voter mobilization, candidate recruitment, or donating campaign contributions) into membership recruitment activities. The reduced political involvement of unions following the passage of RTW laws is found to have aggregate consequences in terms of lower turnout and reduced vote shares for Democratic candidates at the regional level. We complement this analysis by studying how RTW laws moderate the effect of unionization on campaign contributions from employees at the establishment level. It is the unionized workplace where unions are directly connected to employees and where RTW laws may thus have a large impact on unions' political influence. To analyze this, we split our estimation sample based on whether or not the union election takes place in a state that has a RTW law in force at the time of the election. Results are presented in Table 3.5, Panels A.1 and A.2. In states without RTW laws, we see significantly positive effects of unionization on support for Democratic (relative to Republican) candidates, while for RTW states the coefficients are smaller and not significant. This is true for all employees as well as for workers and managers separately. Thus, fewer political mobilization efforts under RTW legislation seem to decrease unions' ability to channel campaign contributions from employees at unionized workplaces.

Second, we investigate whether results vary across union organizations with different ideological positions. We exploit union-level differences in ideology scores, which are derived from Bonica (2014) and are based on the campaign contributions donated by union organizations' political action committees. Note that all unions in our sample have ideology scores substantially below zero and can be clearly viewed as liberal donors (see Appendix Table A.2). Nevertheless, we partition the sample of union elections into union organizations with an ideology score below vs. above the sample median. The mean ideology scores in the two subsamples are -.807 and -.654, meaning that we only compare somewhat more and less liberal unions. The estimated effect sizes, shown in Panels B.1 and B.2 of Table 3.5, are similar across the two groups. We thus do not find evidence for differential effects by union ideology.[39] In concert, the results suggest that unions' political activities matter more for their political impact on employees than their ideological position (in which we observe little variation).

3.6.3 Differentiating Recipients

So far, we have distinguished recipients of campaign contributions only with respect to their party affiliation. We now examine candidate heterogeneity more closely by consider-

[39] Results are similar when we split by terciles of ideology scores.

ing ideological differences among candidates within the Democratic and the Republican party and by differentiating between candidates in federal and local elections. Moreover, we study whether the observed changes in contributions to candidates extend to contributions to political action committees.

Table 3.6 – Differentiating Candidates by Within-Party Ideology

	Democrats			Republicans		
	All	Moderate	Liberal	All	Moderate	Conservative
	(1)	(2)	(3)	(4)	(5)	(6)
[A]: All employees						
δ_{DiD}	0.0920	-0.0182	0.121***	-0.147**	-0.0686	-0.153***
	(0.0634)	(0.0544)	(0.0462)	(0.0654)	(0.0547)	(0.0494)
[B]: Workers						
δ_{DiD}	0.0728**	0.0308	0.0550*	-0.0502	-0.0155	-0.0309
	(0.0352)	(0.0237)	(0.0298)	(0.0317)	(0.0225)	(0.0257)
[C]: Managers						
δ_{DiD}	0.0735	0.0129	0.0896***	-0.130***	-0.0563	-0.123***
	(0.0467)	(0.0391)	(0.0347)	(0.0490)	(0.0397)	(0.0369)

Notes: The table reports DiD coefficients estimated in model (3.1) for the effect of unionization on IHS-transformed amounts contributed to different candidate groups. Moderate (liberal) Democrats refer to Democratic candidates with a CF score above (below) the median CF score of all Democratic candidates observed in our sample of matched contributions. Moderate and conservative Republicans are differentiated accordingly using the median Republican CF score. The sample includes establishments with a pro-union vote share between 20% and 80%. $N = 33,103$ establishment-cycle observations. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Within-party ideological differences. Our results show that unionization increases support for Democratic relative to Republican candidates. The change in party composition may reflect a change in employees' ideological position or merely an increased signaling of party affiliation. To further examine the ideological patterns in campaign contributions, we study ideological differences among candidates within the same party. For this, we make use of Bonica's (2014) CF scores that assign each recipient an ideal point along a liberal-conservative scale. Democratic candidates are categorized as "moderate" versus "liberal" if their CF score lies above the median CF of all Democrats observed in our sample of matched contributions. Similarly, we distinguish between "moderate" and "conservative" Republicans using the median Republican CF score. Table 3.6 shows results from our DiD model, where the outcome is the amount contributed to each of the candidate types. Considering first all employees jointly, we see strong differences in the effects of unionization by the within-party ideological positions of candidates. Unionization significantly increases employees' support for the most liberal Democrats and decreases support for the most conservative Republicans. In contrast, contributions to moderate Democrats or Republicans are not significantly affected. These results are similar when we focus on donations from managers only, and also for workers the increased support for Democrats is more pronounced for more liberal Democrats. Overall, our effects appear to be driven by a shift in contributions between clearly distinguishable conservative and liberal candidates (instead of a shift at the margin from moderate Republicans to moderate Democrats).

Federal versus local candidates. We continue by examining whether our effects are limited to contributions to candidates in either federal or local (i.e., state) elections. U.S. legislation on labor issues, which unions may particularly focus on when endorsing candidates and policies at the unionized workplace, is enacted not only at the federal level, but also at the state-level (e.g., state-specific minimum wages, right-to-work laws). In line with this, Panels F and G of Appendix Table A.3 show that our estimates are driven by contributions to both federal and local candidates. Effect sizes are a bit larger for contributions to candidates running for federal offices, but at both levels we see a significant shift in donations from Republicans to Democrats in response to unionization (and no effect on total amounts).

Table 3.7 − Contributions to Political Action Committees

	Party/candidate PACs		Interest-group PACs					
	All	Dem − Rep	All	Corporation	Trade assoc.	Member orga.	Labor orga.	Dem − Rep
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
[A]: All employees								
δ_{DiD}	-0.0255	0.0968**	-0.0824	-0.0929**	-0.0261	-0.00886	0.0168	0.0599
	(0.0522)	(0.0478)	(0.0635)	(0.0409)	(0.0440)	(0.0311)	(0.0109)	(0.0407)
[B]: Workers								
δ_{DiD}	0.0624*	0.00991	0.0876**	-0.0199	0.0211	0.0461**	0.0188***	0.0239
	(0.0320)	(0.0275)	(0.0347)	(0.0205)	(0.0158)	(0.0190)	(0.00709)	(0.0266)
[C]: Managers								
δ_{DiD}	-0.000602	0.102***	-0.0931*	-0.0821**	-0.0259	0.000722	0.00369	0.0810**
	(0.0344)	(0.0315)	(0.0488)	(0.0340)	(0.0331)	(0.0179)	(0.00684)	(0.0324)

Notes: The table presents DiD coefficients estimated in model (3.1) for the effect of unionization on IHS-transformed amounts contributed to different committee groups. In columns (2) and (7) the dependent variable is the difference between the IHS-transformed amounts contributed to Democratic and Republican committees. Interest-group PACs are categorized as "Democratic" ("Republican") if more (less) than 50% of their own campaign contributions goes to Democratic candidates. The sample includes establishments with a pro-union vote share between 20% and 80%. $N = 33,103$ establishment-cycle observations. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Contributions to political action committees. In Table 3.2, we have shown that contributions to PACs account for a large share of political contributions from employees. If unions particularly encourage workers to donate to candidates, this may come at the detriment of workers' contributions to committees. On the other hand, if unions mobilize workers to participate by donating to labor PACs, then we will underestimate the total effect of unionization on political donations. Table 3.7 reports DiD estimates from model (3.1) for PAC contributions. We distinguish between single-party/candidate PACs and interest-group PACs, where the latter are further disaggregated into corporate, trade association, membership organization, and labor organization PACs. Besides considering the total amount given to these committees, we also measure partisan support by the difference in contribution amounts to Democratic versus Republican PACs. For interest-group PACs, party affiliation is determined from the recipients of the PAC's own campaign contributions.[40] Considering first the contributions from all employees of an establishment to party/candidate PACs, the results mimic those for candidate contributions. While there is no effect on total amounts, unionization leads to a significant shift from Republican to Democratic committees. Among interest-group PACs, there is a significant decrease in donations to corporate PACs. When distinguishing between donations from workers and managers, results differ somewhat. For workers, we see a significant increase in the total amounts donated to both party/candidate committees and interest-group PACs, which implies that unions are successful in mobilizing PAC contributions from workers. The increase in donations appears to be driven by membership and labor organizations, pointing toward an increased support for civil society and labor interest groups. In contrast to our results on candidates, however, we do not see a significant shift across party affiliations. For managers, the results are very similar to those on candidate contributions. While there is no effect on overall PAC spending, managers increasingly donate to Democratic rather than Republican PACs. In particular, donations to corporate PACs drop, which highlights that unionization can decrease managers' support for business interest groups. Overall, these results match with the observed pro-liberal shift in workers' and managers' contributions to political candidates.

3.7 Conclusion

Labor unions employ vast resources to shape labor policies and welfare regulations through political activities such as lobbying legislators or supporting candidates financially. Lasting change, however, requires changes in preferences and beliefs. Do unions influence political ideologies? To understand the political power of labor unions, it is important to understand their effect on millions of individuals at the unionized workplace. At work, unions can provide information and shape social interactions among employees that affect their political behavior. Importantly, unions' aggregate political impact does not only

[40] To track contributions that PACs donate themselves, we exploit that Bonica (2019) has matched recipient identifiers to contributor identifiers for recipients' own contributions. Based on the matched outgoing contributions from PACs, we define an interest-group PAC as "Democratic" ("Republican") if more (less) than 50% of its campaign contributions goes to Democratic candidates in a given election cycle.

depend on their effect on the in-group that benefits from unionization, but also hinges on the reaction of potential out-groups, in particular the firm's management. Managers' power at the workplace and in politics makes their response to unionization particularly relevant for the assessment of the overall impact of unionization.

This paper analyzes the political effects of workplace unionization, building on an establishment-level dataset that combines union elections with campaign contributions from employees spanning the 1980-2016 period in the United States. Comparing establishments with an interest in unionization that won and lost the union election in a stacked DiD model, we find that unionization increases contributions to Democratic candidates relative to Republican candidates by 12 percentage points for workers and 20 percentage points for managers, while we do not find a permanent impact on the overall amount of contributions. These effects do not seem to be driven by a change in the composition of donors but by changes of political behavior at the individual level. Overall, we show that labor unions influence the political preferences not only of union members but also of their firms' management.

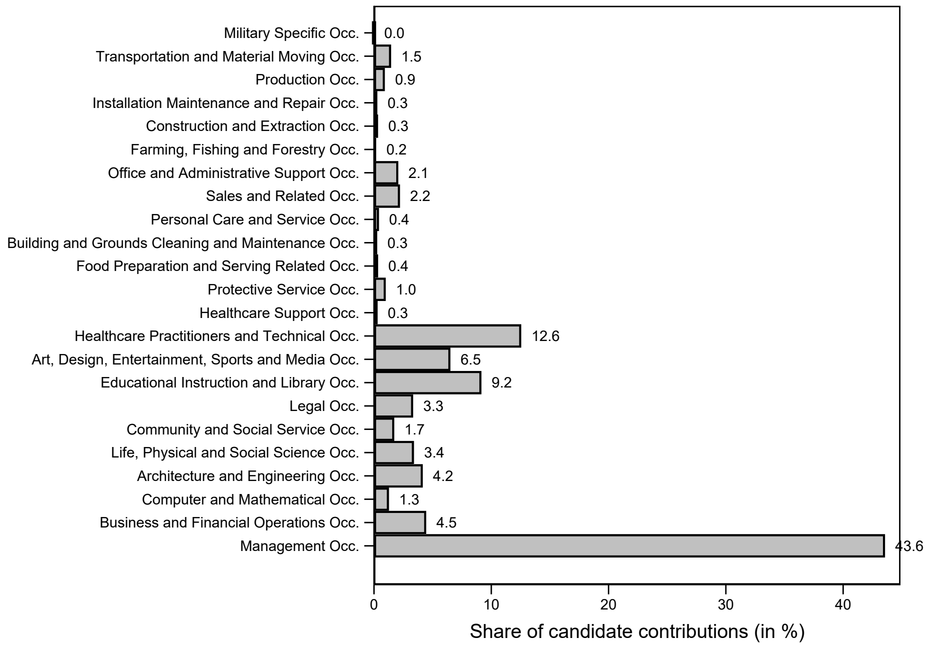
The results are indicative of a reduction of worker-manager cleavages in ideological positions, which is consistent with an improvement in workplace labor relations. If unionization fosters bargaining and communication between workers and managers on a more equal playing field, contact theory suggests an enhancement in managers' understanding of workers' political preferences. While the results may appear surprising in light of the strong opposition of employers toward unions in the United States, a distinction between ex-ante beliefs and ex-post effects of unionization seems crucial. The literature has found little evidence that unionization leads to higher wages (DiNardo and Lee, 2004; Frandsen, 2021; Freeman and Kleiner, 1990b) or reduced productivity (Dube et al., 2016; Sojourner et al., 2015), which could lower firms' profitability. We welcome future work that studies more closely how managers form beliefs about unionization.

Our findings may have implications for broader developments in U.S. politics. The longstanding decline in private-sector union density from 24.2% in 1973 to 6.1% in 2021 (Hirsch and Macpherson, 2022) implies that millions of individuals have forfeited the engagement with unions, which has led to lasting shifts in political preferences. The erosion of unionization can be an important contribution to the increased alignment of workers with the political right that has been observed over the last decades (Gethin et al., 2022). More recently, labor shortages during the COVID-19 pandemic have led to a renewed interest in labor activism. Prominent examples of strikes and union petition drives in Starbucks shops, Amazon warehouses, and healthcare facilities suggest a moment of resurgence for labor organization. Whether this trend persists may be consequential for the balance of political power and support for pro-labor politics in the United States.

Appendices

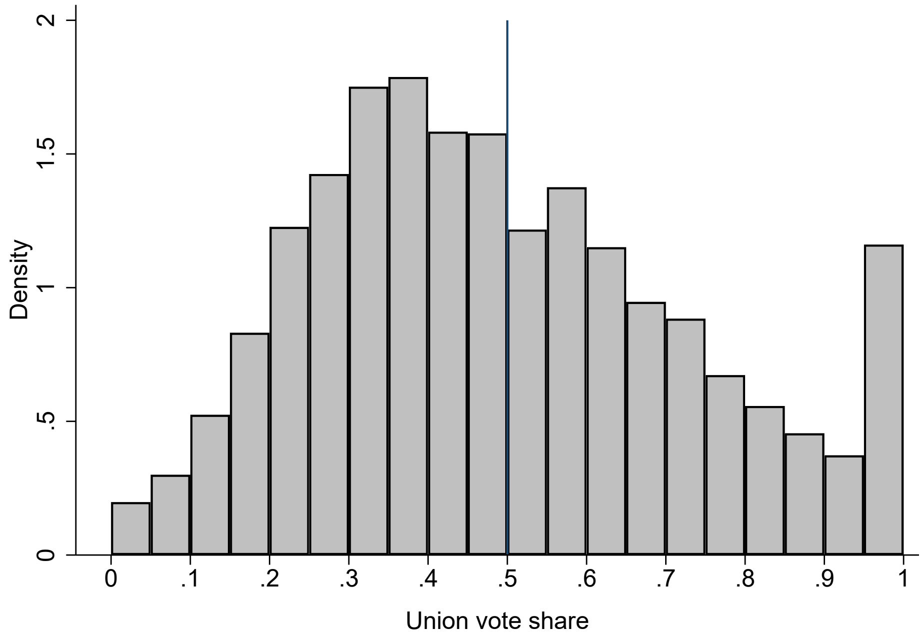
A Additional Figures and Tables

Figure A.1 – Donor Occupations



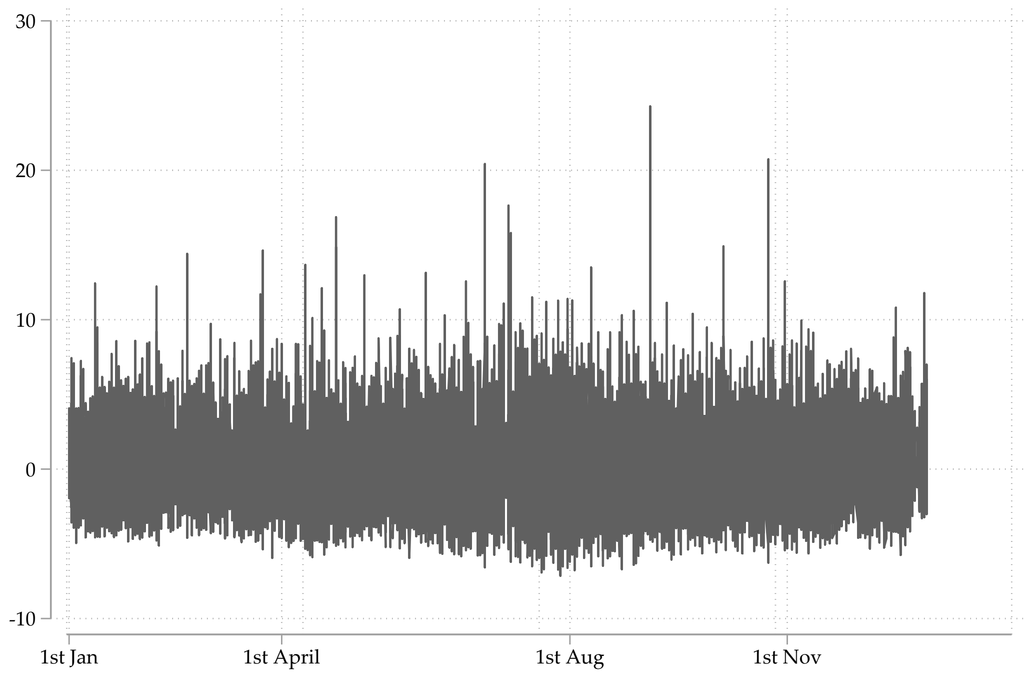
Notes: The figure shows the distribution of occupations for all candidate contributions that are included in our matched estimation sample and have a classified occupation. For 28.1% of the contributions we were not able to assign an occupation code. Occupation groups are 2-digit codes of the 2018 Standard Occupational Classification (SOC). See Appendix B.3 for details on the occupation classification procedure.

Figure A.2 − Vote Share Distribution



Notes: The figure plots the density of union vote shares for all 6,063 union elections included in our matched estimation sample. The Frandsen (2017) test strongly rejects continuity in the union vote share density at the 50% cutoff (p-value = .002 for $k = 0$ and p-value = .003 for $k = .02$).

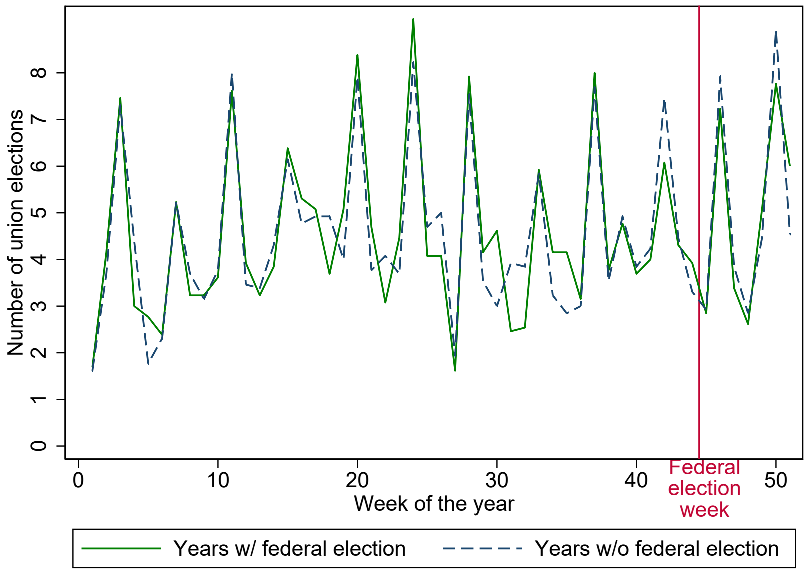
Figure A.3 − Seasonally Adjusted Fatal Work Accidents, 1984-2012



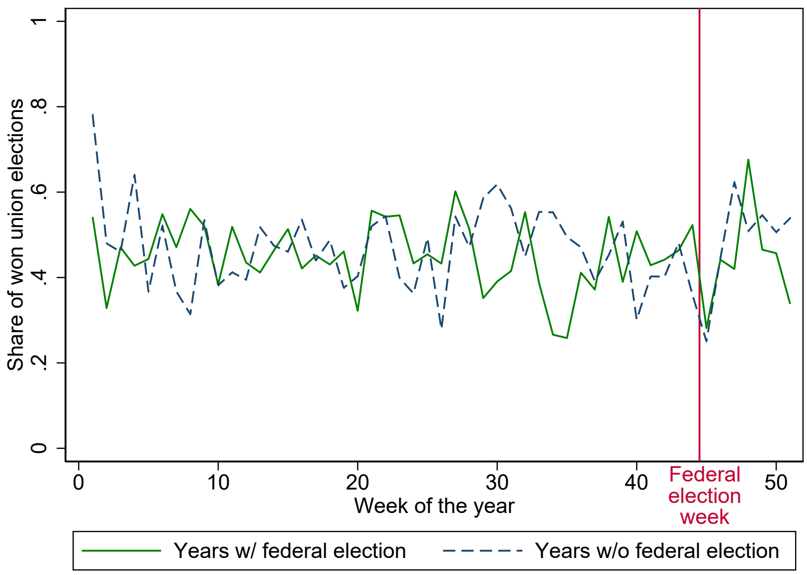
Notes: The graph shows the number of fatalities caused by work accidents on a given day of a year (e.g., January 1st) for all years in our sample period after the mean number of fatalities on that given day over our sample period (e.g., mean number of fatalities on January 1st between 1984 and 2012) is subtracted.

Figure A.4 – Cyclicality of Union Elections

(a) Number of Union Elections per Week of the Year



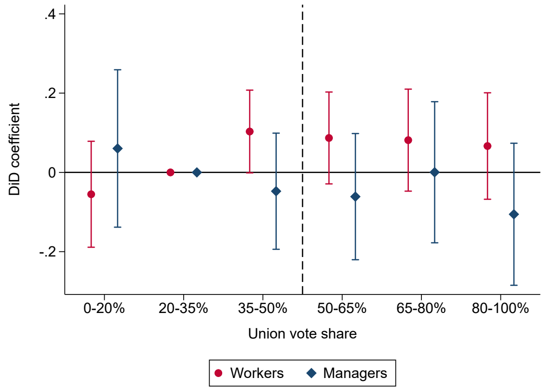
(b) Share of Won Union Elections per Week of the Year



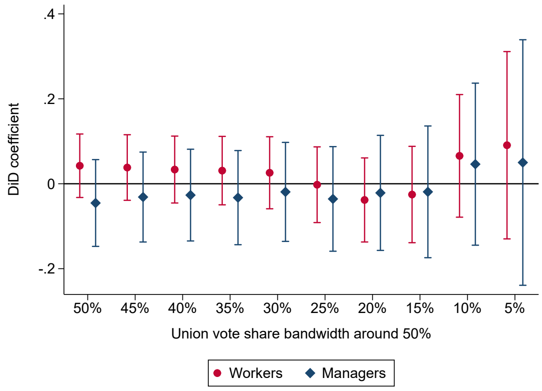
Notes: The graphs show the mean number of elections (Panel (a)) and mean share of won union elections (Panel (b)) per week of the year across all years in our period of analysis, i.e., between 1985 and 2010. The means are based on our matched estimation sample. We distinguish between years with and without federal elections. The red line highlights the week of federal elections, which is calendar week 44 or 45.

Figure A.5 – Effect of Unionization on Total Contribution Amounts - DiD-RDD Results

(a) Vote Share Heterogeneity



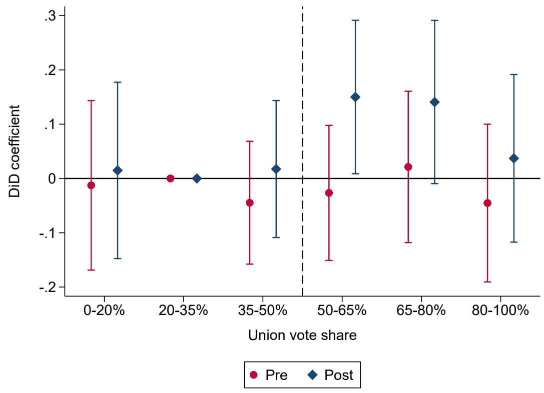
(b) Vote Share Bandwidth Sample Restrictions



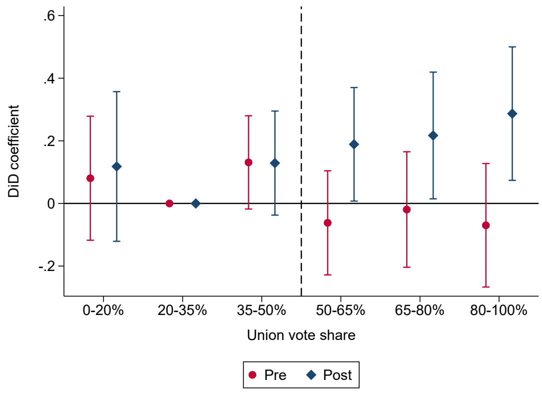
Notes: The graphs show RDD-type placebo and robustness tests for the effect of unionization on the IHS-transformed total amount contributed. Panel (a) reports the δ_g coefficients estimated in model (3.3). The vote share distribution is partitioned into six bins, indicated on the x-axis. The omitted reference group is 20-35%. Panel (b) reports DiD coefficients estimated in model (3.1). Each dot refers to a single DiD coefficient that is estimated among elections with a union vote share in a given bandwidth around the 50% cutoff. Estimates from smaller bandwidths compare changes between increasingly close elections. Results are always shown separately for contributions from non-managerial workers ("workers") and from managers and supervisors ("managers"). 95% confidence intervals are depicted for standard errors clustered at the establishment level.

Figure A.6 – Effect of Unionization on Democratic versus Republican Support - Vote Share Heterogeneity in Pre- versus Post-Effects

(a) Workers



(b) Managers



Notes: The graphs report coefficients for interactions between union win, six vote share categories, and two dummies for pre- versus post-union election periods. The regressions modify model (3.3) by including an additional interaction with a pre-period dummy (three and two cycles before the union election). The reference event time is the cycle before the union election and the reference vote share category is 20-35%. The outcome variable is the difference between the IHS-transformed amounts contributed to Democratic and Republican candidates. Results are shown separately for contributions from non-managerial workers ("workers") and from managers and supervisors ("managers"). 95% confidence intervals are depicted for standard errors clustered at the establishment level.

Table A.1 − Characteristics of Matched and Non-Matched Union Elections

	Matched	Not matched
Number of elections	6,063	22,760
Union win (dummy)	.4397	.4405
Union vote share	.4950	.4955
Number of votes	119.37	81.92
Number of eligible voters	139.27	94.01
Industry: mining	.0397	.0388
Industry: manufacturing	.3338	.3731
Industry: transport	.1785	.1731
Industry: trade	.1397	.1251
Industry: finance	.1008	.0584
Industry: services	.1834	.2192
Years 1985-89	.1618	.2795
Years 1990-94	.1908	.2529
Years 1995-99	.2319	.2261
Years 2000-04	.2547	.1617
Years 2005-10	.1608	.0798

Notes: The table reports mean characteristics of matched and non-matched union elections. Matched elections form our estimation sample and are defined as those for whom we were able to match at least one employee contribution in any of the seven election cycles around the union election (three before, cycle of union election, three after).

Table A.2 – Contributions of Union Organizations

Union organization	# of elections	% of contr. to Dem.	CF score
Teamsters Union	1605	91.0	-.655
United Steelworkers	481	98.0	-.770
United Food & Commercial Workers Union	434	97.7	-.800
Service Employees International Union	407	93.6	-.795
International Brotherhood of Electrical Workers	320	94.4	-.731
United Auto Workers	249	98.0	-.958
Machinists/Aerospace Workers Union	217	98.5	-.779
Operating Engineers Union	208	86.5	-.549
Communications Workers of America	170	95.8	-.761
UNITE HERE	136	94.0	-.706
Laborers Union	119	93.3	-.707
Carpenters & Joiners Union	110	89.6	-.650
American Federation of State/Cnty/Munic Employees	91	79.9	-.747
Office and Professional Employees International Union	51	99.3	-.816
Plumbers/Pipefitters Union	51	91.8	-.662
Amalgamated Transit Union	50	92.8	-.727
National Union of Hospital and Health Care Employees	47	96.7	-.567
Security, Police and Fire Professionals of America	44	100.0	-.793
International Longshore/Warehouse Union	43	94.2	-.920
Bakery, Confectionery, Tobacco & Grain Union	40	99.6	-.822
International Alliance Theatrical Stage Employees	40	95.0	-.742
American Nurses Association	38	83.7	-.561
Sheet Metal, Air, Rail & Transportation Union	35	92.9	-.635
United Mine Workers	33	92.2	-.640
Utility Workers Union of America	33	96.8	-.821
Transport Workers Union	27	94.1	-.663
Bridge, Structural, Ornamental and Reinforcing Iron Workers	26	94.0	-.719
Boilermakers Union	25	94.6	-.703
Painters & Allied Trades Union	25	89.1	-.714
United Electrical, Radio and Machine Workers of America	22	100.0	-1.115
American Federation of Teachers	19	96.3	-.748
Glass, Molders, Pottery, Plastics and Allied Workers	18	99.2	-.826
International Union of Journeymen and Allied Trades	18	96.8	-.698
Operative Plasterers and Cement Masons	17	91.2	-.703
Seafarers International Union	16	71.2	-.206
National Nurses United	15	98.3	-1.060
Roofers Union	14	92.7	-.765
International Guards Union of America	13	82.9	-.637
American Federation of Government Employees	12	95.9	-.791
SAG-AFTRA	9	100.0	-.933
American Postal Workers Union	9	96.5	-.735
International Union of Allied Novelty and Production Workers	8	-	-.524
Marine Engineers Beneficial Assn	7	85.2	-.606
International Association of Firefighters	6	84.2	-.504
American Federation of Musicians	6	91.8	-.562
Bricklayers Union	5	95.7	-.694
Insulators Union	4	94.3	-.815
Intl Fedn of Prof & Technical Engineers	2	87.4	-.824
International Longshoremens Assn	1	91.5	-.524
National Education Assn	1	86.3	-.519
Actors' Equity Assn	1	-	-.880
Total	5,378	93.5	-.726

Notes: The table reports characteristics of campaign contributions donated by union organizations in our sample of union elections. We consider all contributions from PACs associated with a union, including local union branches.'% of contr. to Dem.' refers to the share of contributions going from a union to Democratic (as opposed to Republican) candidates. 'CF score' is the ideology score obtained from Bonica (2014) (when we match several PACs to one union organization, we average the ideology score of the different PACs, weighting each score by the number of donations). For 685 out of the 6,063 elections in our estimation sample, we are not able to match any PAC contribution. Totals in the last row give the weighted average over all union organizations, where the weights are the number of elections in our sample.

Table A.3 − Robustness of Main Results

	\$ to all candidates			\$ to Dem. − \$ to Rep.		
	All	Workers	Managers	All	Workers	Managers
	(1)	(2)	(3)	(4)	(5)	(6)
[A]: Baseline						
δ_{DiD}	0.0332	0.0259	-0.0192	0.239***	0.123***	0.204***
	(0.0794)	(0.0433)	(0.0595)	(0.0792)	(0.0396)	(0.0561)
[B]: Borusyak, Jaravel, and Spiess (2021)						
δ_{DiD}	0.0900	0.0420	0.00861	0.236***	0.130***	0.183***
	(0.0747)	(0.0422)	(0.0576)	(0.0742)	(0.0390)	(0.0545)
[C]: Callaway and Sant'Anna (2021)						
δ_{DiD}	0.0152	0.0416	-0.0378	0.243***	0.137***	0.135**
	(0.0827)	(0.0444)	(0.0606)	(0.0871)	(0.0453)	(0.0619)
[D]: Log(Amount+1)						
δ_{DiD}	0.0273	0.0236	-0.0190	0.220***	0.111***	0.186***
	(0.0727)	(0.0393)	(0.0544)	(0.0721)	(0.0358)	(0.0511)
[E]: Untransformed amounts						
δ_{DiD}	-27.62	2.414	-22.95	116.7***	15.58**	65.38***
	(60.18)	(10.34)	(33.02)	(36.88)	(6.223)	(20.13)
[F]: Only federal candidates						
δ_{DiD}	0.0476	0.0257	-0.0177	0.207***	0.0982***	0.182***
	(0.0751)	(0.0390)	(0.0535)	(0.0764)	(0.0364)	(0.0519)
[G]: Only local candidates						
δ_{DiD}	-0.0472	0.0241	-0.0337	0.158***	0.0454*	0.130***
	(0.0500)	(0.0285)	(0.0427)	(0.0440)	(0.0245)	(0.0384)

Notes: The table presents robustness checks for our DiD estimates of the effect of unionization on the total amount contributed (columns (1) - (3)) and on the difference between the amounts contributed to Democratic and Republican candidates (columns (4) - (6)). $N = 33,103$ establishment-cycle observations. Panel A shows the baseline results from the stacked DiD model (3.1) with IHS-transformed amounts. Panel B presents results from the imputation approach introduced by Borusyak et al. (2021). Panel C implements the DiD estimator of Callaway and Sant'Anna (2021), where we use both never-treated establishments (i.e., lost elections) and not-yet-treated establishments (i.e., won elections in later cycles) as comparison units. In Panel D, outcomes are transformed as log(amount +1), while in Panel E we use untransformed amounts. In Panels F and G only contributions to candidates in federal (congressional and presidential) or state elections are considered, respectively. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4 − Robustness to Alternative Worker-Manager Classifications

	IHS($ to all candidates)		IHS($ to Dem.) − IHS($ to Rep.)	
	Workers	Managers	Workers	Managers
	(1)	(2)	(3)	(4)
[A]: Baseline (80^th percentile of supervisor tasks)				
δ_{DiD}	0.0259	-0.0192	0.123***	0.204***
	(0.0433)	(0.0595)	(0.0396)	(0.0561)
[B]: 90^th percentile of supervisor tasks				
δ_{DiD}	0.0430	-0.0409	0.140***	0.201***
	(0.0458)	(0.0585)	(0.0421)	(0.0546)
[C]: Supervisor tasks "very important" (4 out of 5 in ranking)				
δ_{DiD}	0.0271	-0.0218	0.131***	0.203***
	(0.0432)	(0.0597)	(0.0394)	(0.0561)
[D]: Non-managerial supervisors as workers				
δ_{DiD}	0.0400	-0.0506	0.163***	0.183***
	(0.0481)	(0.0570)	(0.0448)	(0.0529)

Notes: The table presents robustness checks for alternative worker-manager classifications. Reported are the DiD coefficients estimated in model (3.1) for the effect of unionization on the IHS-transformed total amount contributed (columns (1) and (2)) and on the difference between the amounts contributed to Democratic and Republican candidates (columns (3) and (4)). $N = 33,103$ establishment-cycle observations. Panel A shows the baseline results in which "managers" are defined as donors in "Management occupations" (SOC group 11) or in occupations above the 80^th percentile of supervisor tasks and independent judgment. "Workers" are all remaining donors with a classified occupation. In Panel B, we increase the cutoff for supervisor tasks and independent judgment to the 90^th percentile. Panel C, instead, uses an absolute cutoff for the importance of supervisor tasks and independent judgment (both need to be "very important", i.e., have a score of 4 or above in the 5-score ranking). In Panel D, we only consider "Management occupations" (SOC group 11) as "managers" and treat all other classified occupations as "workers" (including those with high importance in supervisor tasks and independent judgment). See Appendix B.3 for more details on the classifications. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5 – Alternative Sample Restrictions

	\$ to all candidates			\$ to Dem. – \$ to Rep.		
	All (1)	Workers (2)	Managers (3)	All (4)	Workers (5)	Managers (6)
[A]: Baseline (at least one matched contribution in any cycle						
δ_{DiD}	0.0332	0.0259	-0.0192	0.239***	0.123***	0.204***
	(0.0794)	(0.0433)	(0.0595)	(0.0792)	(0.0396)	(0.0561)
N	33,103	33,103	33,103	33,103	33,103	33,103
[B]: No restriction						
δ_{DiD}	-0.00780	0.00403	-0.0110	0.0514***	0.0252***	0.0436***
	(0.0175)	(0.00940)	(0.0132)	(0.0165)	(0.00827)	(0.0116)
N	159,026	159,026	159,026	159,026	159,026	159,026
[C]: At least one matched contribution in pre-period						
δ_{DiD}	-0.186	0.0411	-0.222*	0.235	0.170*	0.261**
	(0.150)	(0.0970)	(0.124)	(0.166)	(0.0892)	(0.123)
N	11,193	11,193	11,193	11,193	11,193	11,193
[D]: At least one matched contribution in $k = -1$						
δ_{DiD}	0.0524	0.105	-0.167	0.490***	0.303***	0.362***
	(0.163)	(0.104)	(0.135)	(0.169)	(0.0958)	(0.134)
N	10,997	10,997	10,997	10,997	10,997	10,997
[E]: Period between first and last matched contribution						
δ_{DiD}	0.0303	0.117	-0.126	0.469**	0.299***	0.432***
	(0.156)	(0.115)	(0.142)	(0.206)	(0.115)	(0.161)
N	17,911	17,911	17,911	17,911	17,911	17,911

Notes: Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6 – Effect of Unionization on Any Contribution in Cycle

	All (1)	Workers (2)	Managers (3)
[A]: Contributions to all recipients			
δ_{DiD}	-0.00217	0.00222	-0.00261
	(0.00259)	(0.00159)	(0.00196)
[B]: Contributions to candidates			
δ_{DiD}	-0.000133	0.000456	-0.000940
	(0.00223)	(0.00130)	(0.00168)

Notes: $N = 159,026$ establishment-cycle observations. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Data Appendix

B.1 Union Election Data

Data sources. We start by accessing data on NLRB union representation elections between 1961 and 2009 from the replication package of Knepper (2020). The data were originally compiled by Farber (2016). Then, we add data on elections between 2010 and 2018 from NLRB election reports available on https://www.nlrb.gov/reports/agency-performance/election-reports. In concert, our data cover the universe of union elections between 1961 and 2018 and includes information on vote counts, voting outcome, petition filing and election date, establishment name, address, and industry, as well as the name of the union organization.

Sample restrictions. Before matching campaign contributions, we impose the following restrictions on the sample of union elections:

- We only consider elections where a union seeks to be certified and drop elections that stem from petitions of either employers or employees seeking to remove an existing union.
- We delete duplicate entries (multiple records of the same election).
- For multiple entries that reflect elections where more than one union were on the ballot or where different worker groups formed different bargaining units, we follow Frandsen (2021) and retain only the entry with the largest union vote share.
- We further drop a few elections where the voting outcome (won or lost) is not consistent with the vote counts.
- Following the RDD literature on union elections, we restrict the sample to union elections where at least 20 votes were cast.
- We only keep the first union election in each establishment. For this, we identify an establishment as a unique address or a unique combination of the standardized firm name and commuting zone. For a firm that has multiple establishments within the same commuting zone, we thus only consider the first election among these establishments.
- Finally, we only use elections held between 1985 and 2010 to be able to observe employee contributions for three election cycles before and after each union election.

After these restrictions, we are left with 28,823 union elections.

B.2 Details on the Matching of Elections and Campaign Contributions

We link the campaign contributions from employees to union elections in their employing establishment by combining a spatial match with a fuzzy match of firm names.

Geocode commuting zones. In preparation for the spatial match, we first geocode all union election establishments based on their city and state (using the Open Street Map and Google Maps APIs) and assign the 1990 commuting zone. For the employees' campaign contributions, we rely on donor addresses geocoded by Bonica (2019) up to 2016.[41] We use these geocodes to match to them the 1990 commuting zones.

Firm name cleaning. Firm names in both the union election and the contribution data are cleaned and harmonized using the `stnd_compname` Stata command developed by Wasi and Flaaen (2015). The algorithm removes non-standard characters and whitespaces, doing-as-business and FKA names, as well as business entity types (e.g., CORP, INC, LLC). Moreover, it abbreviates common strings in firm names (e.g., Manufacturing → MFG, Professional → PROF).

Linkage algorithm. For each commuting zone, we create lists of all cleaned firm names from the union election and the contribution data. Then, we use the `reclink2` Stata command from Wasi and Flaaen (2015) to compare the string similarity of firm names.[42] For each possible pair of firm names within the commuting zone, the command computes modified bigram scores. We keep potential matches with a score of at least .98 and manually review all of them. We identify roughly 70% of them as correct matches.[43] In our review, we generally took a conservative approach and were more tolerant of possibly rejecting a true match than retaining an incorrect match. This means that we measure a lower bound for the sum of contributions from all employees of an establishment. To demonstrate the spatial dimension of the matching procedure, Figure B.1 shows an example for the location of a union election establishment and all campaign contributions matched to it.

Establishment-level aggregation. As a last step, we use all contributions with a matched establishment name and sum them up at the establishment-election cycle level. Our period of analysis covers three cycles before to three cycles after each union election, i.e., we observe each establishment over a period of seven cycles (14 years). While we generally keep establishment-cycle observations without any matched contribution and code them as zero, we retain only establishments for which we observe at least one matched contribution over the 14-year period. Out of the initial 28,823 union election establishments, we thereby keep 6,063 matched establishments which form our final estimation sample. Table A.1 compares the characteristics of matched and non-matched

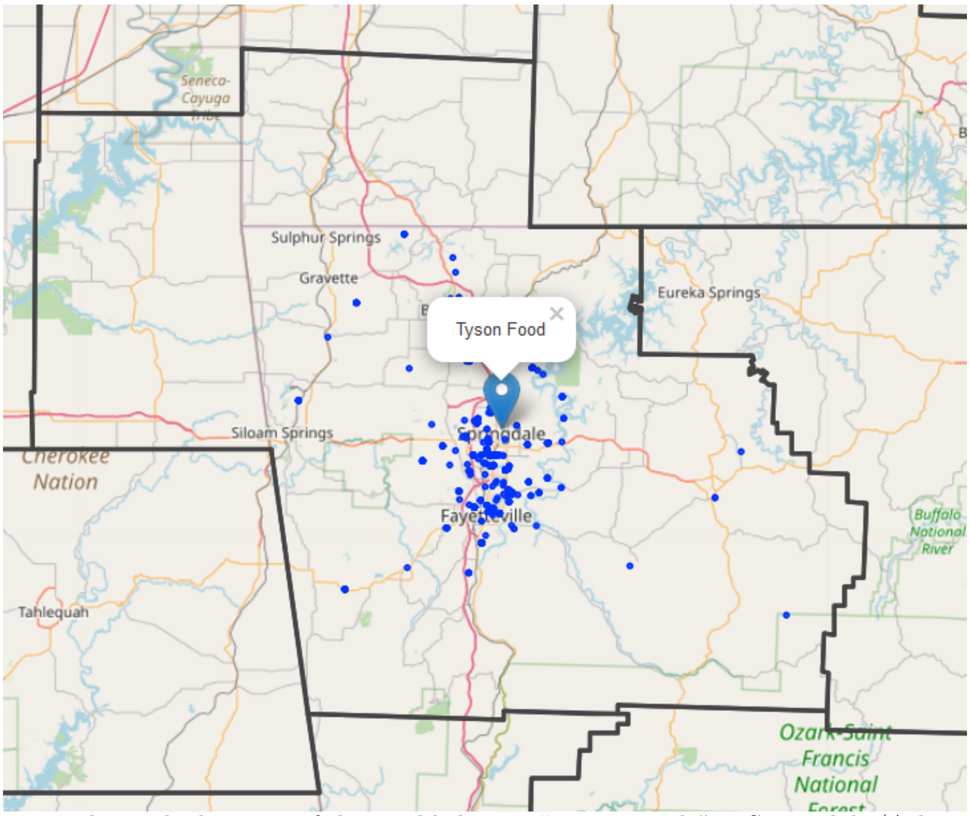
[41] Bonica (2019) contains campaign contributions until 2018 but geocodes are only provided until 2016.

[42] `reclink2` builds on `reclink` written by Blasnik (2010).

[43] The share of matches identified as correct is strongly increasing in the bigram score. For scores between .995 and 1, we keep 90% of the potential matches, while for scores between .98 and .985 this share is only 34%. We also tried keeping potential matches with a lower score (.95), but a manual review of a subsample of those revealed that a very low share of them represented correct matches.

establishments.

Figure B.1 – Example of Spatial Matching Procedure



Notes: The map shows the location of the establishment "Tyson Foods" in Springdale (Arkansas), which held a union election on 22/06/2006. Blue dots represent the location of all campaign contributions matched to the establishment. Black lines are 1990 commuting zone borders.

B.3 Occupation Classification

NRLA definitions. We rely on the definition of the National Labor Relations Act (NLRA) to differentiate between employees eligible for unionization and employees banned from unionizing. The NLRA passed by Congress in 1935 sets rules for the unionization of private sector employees. It establishes who can and who cannot join a union. Section 7 describes the right of employees to join a union:

> "Employees shall have the right to self-organization, to form, join, or assist labor organizations, to bargain collectively through representatives of their own choosing [...] and shall also have the right to refrain from any or all of such activities [...]." (29 U.S.C. § 157)

The NRLA explicitly restricts the right to unionize to employees. It does not extend it to individuals with management and supervisory responsibilities, as they are part of the company's management: The term 'employee' "shall include any employee [...] but shall not include any individual [...] employed as a supervisor" (29 U.S.C. § 152(3)). The distinction between supervisors and employees, however, is not clear-cut, and the NLRA goes on to define supervisors as follows:

> "The term 'supervisor' means any individual having authority, in the interest of the employer, to hire, transfer, suspend, lay off, recall, promote, discharge, assign, reward, or discipline other employees, or responsibly to direct them, or to adjust their grievances, or effectively to recommend such action, if in connection with the foregoing the exercise of such authority is not of a merely routine or clerical nature, but requires the use of independent judgment." (29 U.S.C. § 152(11))

To differentiate between the labor force eligible for unionization and the company's management, we follow two steps: First, we harmonize occupations, and second, we calculate the supervisory element of each occupation based on the NLRA definition.

Occupation harmonization. The free-text occupations reported in DIME are not standardized. Thus, we map them to the 6-digit Standard Occupation Classification. For this, we combine an ensemble classifier called SOCcer (Russ et al., 2016), fuzzy string matching to an extensive crosswalk of laymen's occupation titles from O*NET, as well as manual reviews from Dreher et al. (2020) and manual reviews of the most common occupation titles. In particular, we implement the following steps to identify good matches between a free-text occupation and a SOC code. First, we keep a match determined by SOCcer if the score of the first best match is higher than 0.3 and the difference to the second best match is larger than 0.1. Second, we search for exact matches of any substring of the free-text occupations and a list of laymen's occupation titles, abbreviations and reported titles by experts obtained from O*NET. Third, we fuzzy match the lists from O*NET with the free-text occupations and keep matches with a score above 0.99. Fourth, we add matches from Dreher et al. (2020), which are based on a manual review. Finally, we manually review the free-text occupations that appear more than 50 times in our database of candidate contributions. With that procedure, we are able to assign a SOC code to 72% of all candidate contributions in our matched sample.

Since the share of non-classified occupations is not negligible, we seek to understand whether non-classification can impact our results on the effects of unionization. For this, we use the contribution-level dataset and estimate our baseline model (3.1) with an indicator for missing occupation classification as the dependent variable. The model yields an insignificant DiD coefficient of .0058 (p-value = 0.76). Thus, the likelihood of occupation non-classification does not appear to be related to unionization.

Manager/supervisor versus worker classification. We follow the NLRA and classify an individual as a supervisor if *independent judgment* and a *supervisor task* are important for her occupation. In order to identify occupations with these characteristics, we merge the Occupational Information Network database (O*NET, version 26.3) containing task- and skill-content of 6-digit SOC occupations to our DIME occupations. The information in O*NET is supported by the U.S. Department of Labor and based on surveys of workers working in the respective occupation. Only the importance of specific skills and abilities for an occupation is determined by occupational analysts. We select six variables that closely resemble at least one work activity of a supervisor as defined in the NLRA to identify occupations with *supervisor tasks*. The variables are listed in Table B.1 and measure the importance of the activity in each occupation. We classify an occupation as containing *supervisor tasks* if the importance of at least one listed task is equal or above the 80$^{\text{th}}$ percentile of all 6-digit SOC occupations.[44] We then go on to evaluate whether the occupation requires *independent judgment*, the second condition that we identify in the NLRA definition of a supervisor. We evaluate whether an occupation requires *independent judgment* based on the following four variables: Independence (Work Styles), Leadership (Work Styles), Structured versus Unstructured Work (Work Context), and Freedom to Make Decisions (Work Context). Again, we classify an occupation as requiring *independent judgment* if the importance of at least one of the listed variables is equal or above the 80$^{\text{th}}$ percentile.[45] Finally, we classify individuals as managers or supervisors if their occupation is classified as "Management Occupation" in SOC (SOC group 11) or contains a *supervisor task* and *independent judgment* as defined above.[46] Examples of occupations in the top 95$^{\text{th}}$ percentile of both the *independent judgment* and *supervisor task* score are *Chief Executives*, *Human Resource Managers* and *First-Line Supervisors of Retail Sales Workers*. Non-managerial workers are then identified as all remaining donors to whom we were able to assign a SOC code. With these definitions, we obtain the following occupational composition in our sample of candidate contributions: 42% of contributions originate from managers and supervisors, 30% from non-managerial workers, and for 28% we are unable to obtain a classification.

[44] In our robustness checks, we also use the 90$^{\text{th}}$ percentile as cutoff and an absolute scale classifying any occupation as supervisor where a supervisor task is at least "very important" (a score of 4 or above in the 5-score ranking).

[45] Again, in our robustness checks we also use the 90$^{\text{th}}$ percentile as the cutoff and an absolute scale classifying any occupation as supervisor where independence is at least "very important" (a score of 4 or above in the 5-score ranking).

[46] We were not able to assign a 6-digit SOC code for some of the individuals in our data in cases where the free-text occupation was vague. Instead, we assigned 4-, 3- or 2-digit SOC codes. We classify a 2-digit SOC code occupation as supervisor if all 6-digit SOC code occupations have been classified as supervisors. We proceed accordingly for 3- and 4-digit SOC code occupations. We are thereby conservative and allow for some attenuation bias if supervisors are consequently incorrectly coded as workers.

Table B.1 – Supervisor Tasks in NLRA and O*NET Occupations

Tasks of a *supervisor* defined in NLRA	Corresponding O*NET work activity / skill / context
Hire / transfer / suspend / lay off / discharge	Staffing organizational units
Recall / assign	Management of personnel resources Coordinating the work and activities of others
Promote / reward / discipline	Guiding, directing, and motivating subordinates Resolving conflicts and negotiating with others
Direct employees / adjust their grievances	Management of personnel resources Guiding, directing, and motivating subordinates Coordinating the Work and Activities of Others Coordinate or Lead Others

C Effects of Losing a Union Election

We estimate the effects of losing a union election compared to holding no election by using establishments who hold and lose an election in the future as a control group. Consider the treatment cohort of elections that were held and lost in the cycle 1985/86. Given that we observe each establishment only up to three cycles before the union election, we can use elections held and lost in the next two cycles as control cohorts. The untreated pre-election observations of the 1987/88 control cohort refer to the cycles 1981/82, 1983/84, and 1985/86 (event times $k = \{-2, -1, 0\}$ of the treated cohort), and those of the 1989/90 control cohort refer to the cycles 1983/84, 1985/86, and 1987/1988 (event times $k = \{-1, 0, 1\}$ of the treated cohort). Note that later cohorts are not observed before the treated cohort hold their election and can therefore not be used in a DiD comparison. Consequently, we only have untreated observations that we can compare to the treated cohort's observations in cycles 1981/82, 1983/84, 1985/86, and 1987/88 (event times $k = \{-2, -1, 0, 1\}$). This means we can only identify short-term effects.

Given these considerations, we implement a stacked DiD model as follows. For each cohort of lost elections in cycle g, we create a cohort-specific dataset that is built from cycles in event times $k = \{-2, -1, 0, 1\}$ of the treated cohort $g_i = g$ and from the three pre-election cycles of lost elections in the control cohorts $g_i = \{g + 1, g + 2\}$. Then, the stacked DiD model is estimated as:

$$y_{ik} = \alpha_{ig} + \beta_{kg} + \delta_{\text{DiD}} \times \Big(\mathbb{1}[k \geq 0] \times \mathbb{1}[g_i = g]\Big) + \epsilon_{ik}, \tag{C.1}$$

where k now denotes the number of cycles relative to the cycle when the treated cohort held its union election. Establishment-fixed effects are now saturated with indicators for the cohort-specific dataset g to account for the fact that establishments enter several datasets. The DiD coefficient δ_{DiD} is given by the interaction between a dummy for post-election cycles of the treated cohort ($k \geq 0$) and a dummy for the treated cohort ($g_i = g$). Results are reported in Panel A of Table C.1.

In Panels B and C of Table C.1, we also show results for the alternative staggered DiD estimators by Borusyak et al. (2021) and Callaway and Sant'Anna (2021). In line with our stacking implementation, in settings with no never-treated units, both estimators use not-yet-treated observations as controls. The methods differ from the stacked DiD model in the number of pre-treatment periods used and the aggregation of unit- or cohort-specific effects. In our results, however, the estimates are very similar to those of the stacked DiD model.

Table C.1 − Effects of Losing a Union Election

	IHS($ to all candidates)			IHS($ to Dem.) − IHS($ to Rep).		
	All	Workers	Managers	All	Workers	Managers
	(1)	(2)	(3)	(4)	(5)	(6)
[A]: Stacking						
δ_{DiD}	-0.0491	-0.0263	0.0705	0.0568	-0.0131	0.0366
	(0.0881)	(0.0396)	(0.0529)	(0.0966)	(0.0429)	(0.0575)
N	31,501	31,501	31,501	31,501	31,501	31,501
[B]: Borusyak, Jaravel, and Spiess (2021)						
δ_{DiD}	-0.0481	-0.0285	0.0745	0.0796	-0.00682	0.0485
	(0.0901)	(0.0447)	(0.0590)	(0.100)	(0.0490)	(0.0641)
N	16,658	16,658	16,658	16,658	16,658	16,658
[C]: Callaway and Sant'Anna (2021)						
δ_{DiD}	-0.0434	-0.0381	0.0615	0.0761	-0.00688	0.0534
	(0.0947)	(0.0469)	(0.0637)	(0.105)	(0.0515)	(0.0698)
N	16,658	16,658	16,658	16,658	16,658	16,658

Notes: The table presents DiD estimates for the effect of losing a union election versus holding no election. We compare establishments with a lost union election in a given cycle (treated cohort) with establishments with a lost union election in one of the next two cycles (control cohorts) in a DiD design. Thereby, we estimate short-term effects of losing an election (i.e., for event times $k = \{0, 1\}$). Panel A shows results from a stacked DiD model, and Panels B and C implement the staggered DiD estimators of Borusyak et al. (2021) and Callaway and Sant'Anna (2021). See Appendix C for details of the implementation. Standard errors clustered at the establishment level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References

Abramowitz, A. I. (2010). *The Disappearing Center.* Yale University Press.

Abramowitz, A. I. (2012). *The Polarized Public.* Pearson Higher Ed.

Acemoglu, D., T. A. Hassan, and A. Tahoun (2018). The Power of the Street: Evidence from Egypt's Arab Spring. *The Review of Financial Studies 31*(1), 1–42.

Adena, M., R. Enikolopov, M. Petrova, V. Santarosa, and E. Zhuravskaya (2015). Radio and the Rise of the Nazis in Prewar Germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Adão, R., M. Kolesár, and E. Morales (2018). Shift-Share Designs: Theory and Inference. *NBER Working Paper No. 24944*.

AFL-CIO (2022). Death on the Job. The Toll of Neglect. A National and State-By-State Profile of Worker Safety and Health in the United States. Technical Report 31st Edition, AFL-CIO.

Ahlquist, J. S. (2017). Labor Unions, Political Representation, and Economic Inequality. *Annual Review of Political Science 20*, 409–432.

Ahlquist, J. S., A. B. Clayton, and M. Levi (2014). Provoking Preferences: Unionization, Trade Policy, and the ILWU Puzzle. *International Organization 68*(1), 33–75.

Ahlquist, J. S. and M. Levi (2013). *In the Interest of Others: Organizations and Social Activism.* Princeton, NJ: Princenton University Press.

Aker, J. C. and I. M. Mbiti (2010). Mobile Phones and Economic Development in Africa. *Journal of Economic Perspectives 24*(3), 207–232.

Allcott, H., L. Braghieri, S. Eichmeyer, and M. Gentzkow (2020). The Welfare Effects of Social Media. *American Economic Review 110*(3), 629–676.

Allport, G. (1954). *The Nature of Prejudice.* Addison-Wesley, Oxford.

Anderson, B. (1983). *Imagined Communities.* London: Verso.

Andersson, H. and S. H. Dehdari (2021). Workplace Contact and Support for Anti-Immigration Parties. *American Political Science Review 115*(4), 1159–1174.

Anelli, M., I. Colantone, and P. Stanig (2019). We Were the Robots: Automation and Voting Behavior in Western Europe. *BAFFI CAREFIN Centre Research Paper* (2019-115).

Ang, D. (2023). The Birth of a Nation: Media and Racial Hate. *American Economic Review 113*(6), 1424–1460.

Angrist, J. D. and J.-S. Pischke (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

Ansolabehere, S., J. M. D. Figueiredo, and J. M. Snyder Jr (2003). Why is There so Little Money in U.S. Politics? *Journal of Economic Perspectives 17*(1), 105–130.

Armand, A., P. Atwell, J. F. Gomes, and Y. Schenk (2023). It's a Bird, it's a Plane, it's Superman! Using Mass Media to Fight Intolerance. *Université Catholique de Louvain, Institut de Recherches Economiques, No. 2023012*.

Ash, E., S. Asher, A. Bhowmick, S. Bhupatiraju, D. Chen, T. Devi, C. Goessmann, P. Novosad, and B. Siddiqi (2021). In-group Bias in the Indian Judiciary: Evidence from 5 Million Criminal Cases. *Mimeo.*

Ash, E., W. B. McLeod, and S. Naidu (2019). The Language of Contract: Promises and Power in Union Collective Bargaining Agreements. *Mimeo.*

Asher, S., T. Lunt, R. Matsuura, and P. Novosad (2021). Development Research at High Geographic Resolution: an Analysis of Night-lights, Firms, and Poverty in India using the Shrug Open Data Platform. *The World Bank Economic Review 35*(4), 845–871.

Autor, D., D. Dorn, and G. Hanson (2016). The China Shock: Learning from Labor-Market Adjustment to Large Changes in Trade. *Annual Review of Economics 8*, 205–240.

Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2020). Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure. *American Economic Review 110*(10), 3139–3183.

Babenko, I., V. Fedaseyeu, and S. Zhang (2020). Do CEOs Affect Employees' Political Choices? *Review of Financial Studies 33*(4), 1781–1817.

Bahar, D., C. Parsons, and P.-L. Vézina (2022). Refugees, Trade and FDI. *Oxford Review of Economic Policy*, forthcoming.

Bahar, D. and H. Rapoport (2018). Migration, Knowledge Diffusion and the Comparative Advantage of Nations. *The Economic Journal 128*(612), F273–F305.

Bahia, K., P. Castells, G. Cruz, X. Pedros, T. Pfutze, C. Rodriguez Castelan, and H. Winkler (2020). The Welfare Effects of Mobile Broadband Internet: Evidence from Nigeria. *World Bank Policy Research Working Paper* (9230).

Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science 348*(6239), 1130–1132.

Banerjee, A. V. and R. Pande (2007). Parochial Politics: Ethnic Preferences and Politician Corruption. *CEPR Discussion Paper No. DP6381.*

Bansak, K., J. Hainmueller, and D. Hangartner (2016). How Economic, Humanitarian, and Religious Concerns Shape European Attitudes Toward Asylum Seekers. *Science 354*(6309), 217–222.

Barber, M. (2016a). Ideological Donors, Contribution Limits, and the Polarization of American Legislatures. *The Journal of Politics 78*(1), 296–310.

Barber, M. J. (2016b). Ideological Donors, Contribution Limits, and the Polarization of American Legislatures. *The Journal of Politics 78*(1), 296–310.

Barber, M. J., B. Canes-Wrone, and S. Thrower (2017). Ideologically Sophisticated Donors: Which Candidates do Individual Contributors Finance? *American Journal of Political Science 61*(2), 271–288.

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis 23*(1), 76–91.

Barone, G., A. D'Ignazio, G. De Blasio, and P. Naticchioni (2016). Mr. Rossi, Mr. Hu and Politics. The Role of Immigration in Shaping Natives' Voting Behavior. *Journal of Public Economics 136*, 1–13.

Baumgartner, F. R. and B. L. Leech (1998). *Basic Interests: The Importance of Groups in Politics and in Political Science.* Princeton University Press.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2019). Unity in Diversity? How Intergroup Contact can Foster Nation Building. *American Economic Review 109*(11), 3978–4025.

Bazzi, S. and M. Gudgeon (2021). The Political Boundaries of Ethnic Divisions. *American Economic Journal: Applied Economics 13*(1), 235–266.

Becher, M. and D. Stegmueller (2021). Reducing Unequal Representation: The Impact of Labor Unions on Legislative Responsiveness in the U.S. Congress. *Perspectives on Politics 19*(1), 92–109.

Bekkouche, Y. and J. Cagé (2018). The Price of a Vote: Evidence from France, 1993-2014. *CEPR Discussion Papers 12614*.

Berger, J. and K. L. Milkman (2012). What Makes Online Content Viral? *Journal of Marketing Research 49*(2), 192–205.

Bessone, P., F. R. Campante, C. Ferraz, and P. Souza (2022). Social Media and the Behavior of Politicians: Evidence from Facebook in Brazil. *NBER Working Paper No. 30306*.

Biden, J. (2021). Remarks by President Biden in Honor of Labor Unions. Accessed via https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/09/08/remarks-by-president-biden-in-honor-of-labor-unions/. Speeches and Remarks, The White House.

Blasnik, M. (2010). Reclink: Stata Module to Probabilistically Match Records. *Mimeo*.

Blouin, A. and S. W. Mukand (2019). Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda. *Journal of Political Economy 127*(3), 1008–1062.

Bluhm, R. and M. Pinkovskiy (2021). The Spread of COVID-19 and the BCG Vaccine: A Natural Experiment in Reunified Germany. *The Econometrics Journal 24*(3), 353–376.

Bond, J. R. and R. Fleisher (2000). Congress and the President in a Partisan Era. In *Polarized Politics: Congress and the President in a Partisan Era*, pp. 1–8.

Bonica, A. (2013). Ideology and Interests in the Political Marketplace. *American Journal of Political Science 57*(2), 294–311.

Bonica, A. (2014). Mapping the Ideological Marketplace. *American Journal of Political Science 58*(2), 367–386.

Bonica, A. (2016). *Database on Ideology, Money in Politics, and Elections: Public Version 2.0*. Stanford, CA: Stanford University Libraries.

Bonica, A. (2018). Are Donation-Based Measures of Ideology Valid Predictors of Individual-Level Policy Preferences? *The Journal of Politics 81*(1), 327–333.

Bonica, A. (2019). *Database on Ideology, Money in Politics, and Elections: Public Version 3.0*. Stanford, CA: Stanford University Libraries.

Bonica, A., A. S. Chilton, and M. Sen (2016). The Political Ideologies of American Lawyers. *Journal of Legal Analysis 8*(2), 277–335.

Bonica, A. and H. Rosenthal (2018). The Wealth Elasticity of Political Contributions by the Forbes 400. *Mimeo.*

Bonomi, G., N. Gennaioli, and G. Tabellini (2021). Identity, Beliefs, and Political Conflict. *The Quarterly Journal of Economics 136*(4), 2371–2411.

Borusyak, K., P. Hull, and X. Jaravel (2022). Quasi-experimental Shift-Share Research Designs. *The Review of Economic Studies 89*(1), 181–213.

Borusyak, K., X. Jaravel, and J. Spiess (2021). Revisiting Event Study Designs: Robust and Efficient Estimation. *arXiv 2108.12419.*

Bouton, L., J. Cagé, E. Dewitte, and V. Pons (2022). Small Campaign Donors. *NBER Working Paper No. 30050.*

Boxell, L., M. Gentzkow, and J. M. Shapiro (2022). Cross-country Trends in Affective Polarization. *The Review of Economics and Statistics*, 1–60.

Brown, C., R. Hedges, and L. Powell (1980). Modes of Elite Political Participation: Contributors to the 1972 Presidential Candidates. *American Journal of Political Science 24*(2), 259–290.

Brown, R. (2000). Social Identity Theory: Past Achievements, Current Problems and Future Challenges. *European Journal of Social Psychology 30*(6), 745–778.

Bruno, A. (2017). Reception and Placement of Refugees in the United States. *Congressional Research Service.*

Burgess, R., R. Jedwab, E. Miguel, A. Morjaria, and G. Padró i Miquel (2015). The Value of Democracy: Evidence from Road Building in Kenya. *American Economic Review 105*(6), 1817–1851.

Burns, P. F., P. L. Francia, and P. S. Herrnson (2000). Labor at Work: Union Campaign Activities and Legislative Payoffs in the US House of Representatives. *Social Science Quarterly 81*(2), 507–522.

Bursztyn, L., T. Chaney, T. A. Hassan, and A. Rao (2022). The Immigrant Next Door: Long-Term Contact, Generosity, and Prejudice. *NBER Working Paper No. 28448.*

Bursztyn, L., G. Egorov, R. Enikolopov, and M. Petrova (2019). Social Media and Xenophobia: Evidence from Russia. *NBER Working Paper No. 26567.*

Busby, J. (2020). *Coming Together or Coming Apart? Attitudes of Foreign Policy Opinion Leaders and the Public in the Trump Era.* Chicago Council on Global Affairs.

Callaway, B. and P. H. C. Sant'Anna (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics 225*(2), 200–230.

Campante, F., R. Durante, and F. Sobbrio (2018). Politics 2.0: The Multifaceted Effect of Broadband Internet on Political Participation. *Journal of the European Economic Association 16*(4), 1094–1136.

Campbell, D. T. (1965). Ethnocentric and Other Altruistic Motives. In D. Levine (Ed.), *Nebraska Symposium on Motivation.* Lincoln: University of Nebraska Press.

Campello, M., J. Gao, J. Qiu, and Y. Zhang (2018). Bankruptcy and the Cost of Organized Labor: Evidence from Union Elections. *Review of Financial Studies 31*(3), 980–1013.

Campo, F., S. Giunti, and M. Mendola (2020). The Political Impact of Refugee Migration: Evidence from the Italian Dispersal Policy. CefES Working Paper 456.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017). Curriculum and Ideology. *Journal of Political Economy 125*(2), 338–392.

Cantoni, E. and V. Pons (2022). Does Context Outweigh Individual Characteristics in Driving Voting Behavior? Evidence from Relocations within the United States. *American Economic Review 112*(4), 1226–72.

Card, D. (1996). The Effect of Unions on the Structure of Wages: A Longitudinal Analysis. *Econometrica 64*(4), 957–979.

Card, D., C. Dustmann, and I. Preston (2012). Immigration, Wages and Compositional Amenities. *Journal of International Economics 10*(1), 78–119.

Carlson, E. (2015). Ethnic Voting and Accountability in Africa: A Choice Experiment in Uganda. *World Politics 67*(2), 353–385.

Carothers, T. and A. O'Donohue (2019). *Democracies Divided: The Global Challenge of Political Polarization.* Brookings Institution Press.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019). *A Practical Introduction to Regression Discontinuity Designs: Foundations.* Cambridge University Press.

Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association 115*(531), 1449–1455.

Cavaille, C. and J. Ferwerda (2020). How Distributional Conflict over In-Kind Benefits Generates Support for Far-Right Parties. *The Journal of Politics*, forthcoming.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics 134*(3), 1405–1454.

Chen, M. K. and R. Rohla (2018). The Effect of Partisanship and Political Advertising on Close Family Ties. *Science 360*(6392), 1020–1024.

Chin, A. and K. Cortes (2015). The Refugee/Asylum Seeker. In B. Chiswick and P. Miller (Eds.), *Handbook of the Economics of International Migration*. Volume 1. North-Holland, 585-658.

Cho, W. and J. Gimpel (2010). Rough Terrain: Spatial Variation in Campaign Contributing and Volunteerism. *American Journal of Political Science 54*(1), 74–89.

Christensen, R. A. and A. Ebrahim (2006). How Does Accountability Affect Mission? The Case of a Nonprofit Serving Immigrants and Refugees. *Nonprofit Management and Leadership 17*(2), 195–209.

Christian, P. and C. Barrett (2017). Revisiting the Effect of Food Aid on Conflict: A Methodological Caution. *Policy Research Working Paper 8171*.

Colantone, I. and P. Stanig (2018). The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe. *American Journal of Political Science 62*(4), 936–953.

Conconi, P., G. Facchini, M. F. Steinhardt, and M. Zanardi (2020). The Political Economy of Trade and Migration: Evidence from the US Congress. *Economics & Politics 32*(2), 250–278.

Coppock, A. (2023). *Persuasion in Parallel: How Information Changes Minds about Politics*. University of Chicago Press.

Corno, L., E. La Ferrara, and J. Burns (2022). Interaction, Stereotypes, and Performance: Evidence from South Africa. *American Economic Review 112*(12), 3848–3875.

Correa, A. F., K. J. Mchombu, D. Ndiaye, G. M. Rodriguez, D. Rose, and N. U. Yapa (1997). Rural Information Provision in Developing Countries - Measuring Performance and Impact. Technical report, Report prepared for UNESCO on behalf of IFLA - Paris: UNESCO, 1997.- ix, 116.

CRS (2013). The National Labor Relations Act (NLRA): Union Representation Procedures and Dispute Resolution. *Congressional Research Service Report RL32930*.

CSDS (2022). Media in India: Access, Practices, Concerns and Effect. Technical report, Lokniti, New Delhi.

Dahl, R. A. (2005). *Who Governs?: Democracy and Power in an American City*. Yale University Press.

Dark, T. E. (2001). *The Unions and the Democrats: An Enduring Alliance*. Cornell University Press.

de Benedictis-Kessner, J. and C. Warshaw (2016). Mayoral Partisanship and Municipal Fiscal Policy. *The Journal of Politics 78*(4), 1124–1138.

de Chaisemartin, C. and X. D'Haultfœuille (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review 110*(9), 2964–2996.

Dell, M., N. Lane, and P. Querubin (2018). The Historical State, Local Collective Action, and Economic Development in Vietnam. *Econometrica 86*(6), 2083–2121.

Dell, M. and B. A. Olken (2020). The Development Effects of the Extractive Colonial Economy: The Dutch Cultivation System in Java. *The Review of Economic Studies 87*(1), 164–203.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border Media and Nationalism: Evidence from Serbian Radio in Croatia. *American Economic Journal: Applied Economics 6*(3), 103–132.

Depetris-Chauvin, E., R. Durante, and F. Campante (2020). Building Nations through Shared Experiences: Evidence from African Football. *American Economic Review 110*(5), 1572–1602.

Dimock, M. and R. Wike (2020). America is Exceptional in the Nature of its Political Divide. *Pew Research Center 13*.

DiNardo, J. and D. S. Lee (2004). Economic Impacts of New Unionization on Private Sector Employers: 1984-2001. *The Quarterly Journal of Economics 119*(4), 1383–1441.

Dinlersoz, E., J. Greenwood, and H. Hyatt (2017). What Businesses Attract Unions? Unionization over the Life Cycle of U.S. Establishments. *ILR Review 70*(3), 733–766.

Dreher, A., S. Langlotz, J. Matzat, and C. Parsons (2020). Immigration, Political Ideologies and the Polarization of American Politics. *CESifo Working Paper 8789*.

Druckman, J. N., E. Peterson, and R. Slothuus (2013). How Elite Partisan Polarization Affects Public Opinion Formation. *American Political Science Review 107*(1), 57–79.

Dube, A., E. Kaplan, and O. Thompson (2016). Nurse Unions and Patient Outcomes. *ILR Review 69*(4), 803–833.

Dustmann, C. and I. Preston (2007). Racial and Economic Factors in Attitudes to Immigration. *The B.E. Journal of Economic Analysis & Policy 7*(1), 1–39.

Dustmann, C., K. Vasilijeva, and A. Piil Damm (2020). Refugee Migration and Electoral Outcomes. *Review of Economic Studies 86*(5), 2035–2091.

Dutta, P., R. Murgai, M. Ravallion, and D. Van de Walle (2012). Does India's Employment Guarantee Scheme Guarantee Employment? *Economic and Political Weekly*, 55–64.

Economic Times (2017). Jio, Airtel Take lead in Signing up for BharatNet Connectivity. *Accessed last on July 12 2022 via* $https://economictimes.indiatimes.com/tech/internet/reliance-jio-airtel-vodafone-idea-take-first-dive-into-bharatnet/articleshow/61626019.cms?utm_source = contentofinterestutm_medium = textutm_campaign = cppst.$

Edo, A., Y. Giesing, J. Öztunc, and P. Poutvaara (2019). Immigration and Electoral Support for the Far-Left and the Far-Right. *European Economic Review 115*, 99–143.

Einav, L. and J. Levin (2014). Economics in the Age of Big Data. *Science 346*(6210), 1243089.

Ensley, M. J. (2009). Individual Campaign Contributions and Candidate Ideology. *Public Choice 138*(1–2), 221–238.

Esposito, E., T. Rotesi, A. Saia, and M. Thoenig (2023). Reconciliation Narratives: The Birth of a Nation after the US Civil War. *American Economic Review 113*(6), 1461–1504.

Facchini, G. and A. Mayda (2009). Does the Welfare State Affect Individual Attitudes Toward Immigrants? Evidence Across Countries. *The Review of Economics and Statistics 91*(2), 295–314.

Facchini, G., A. Mayda, and P. Mishra (2011). Do Interest Groups Affect US Immigration Policy? *Journal of International Economics 85*(1), 114–128.

Facchini, G. and M. F. Steinhardt (2011). What Drives US Immigration Policy? Evidence from Congressional Roll Call Votes. *Journal of Public Economics 95*(7-8), 734–743.

Farber, H. (2016). Union Organizing Decisions in a Deteriorating Environment: The Composition of Representation Elections and the Decline in Turnout. *ILR Review 68*(5), 1126–1156.

Farber, H., D. Herbst, I. Kuziemko, and S. Naidu (2021). Unions and Inequality Over the Twentieth Century: New Evidence from Survey Data. *The Quarterly Journal of Economics 136*(3), 1325–1385.

Farina, C. R. (2015). Congressional Polarization: Terminal Constitutional Dysfuction. *Columbia Law Review 115*, 1689.

FEC (2022a). Campaign Finance Statistics.

FEC (2022b). Enforcement Statistics for Fiscal Years 1977- 2022.

Feigenbaum, J., A. Hertel-Fernandez, and V. Williamson (2018). From the Bargaining Table to the Ballot Box: Political Effects of Right to Work Laws. *NBER Working Paper No. 24259*.

Fetzer, T. (2019). Did Austerity Cause Brexit? *American Economic Review 109*(11), 3849–3886.

Financial Times (2014). Narendra Modi to be India's First Social Media Prime Minister. *Accessed last on March 3 2023 via https://www.ft.com/content/e347de5c-e088-11e3-9534-00144feabdc0*.

Financial Times (2019). India: The WhatsApp Election. *Accessed last on July 19 2023 via https://www.ft.com/content/9fe88fba-6c0d-11e9-a9a5-351eeaef6d84*.

Fiorina, M. P. and S. J. Abrams (2008). Political Polarization in the American Public. *Annual Review of Political Science 11*, 563–588.

Flanagan, R. (2007). *What do Unions Do? A Twenty Year Perspective*, Chapter Has Management Strangled U.S. Unions? New York: Routledge.

Flavin, P. (2018). Labor Union Strength and the Equality of Political Representation. *British Journal of Political Science 48*(4), 1075–1091.

Fleisher, R. and J. R. Bond (2004). The Shrinking Middle in the US Congress. *British Journal of Political Science 34*(3), 429–451.

Ford, R. and W. Jennings (2020). The Changing Cleavage Politics of Western Europe. *Annual Review of Political Science 23*, 295–314.

Fowler, C. S. and L. Jensen (2020). Bridging the Gap between Geographic Concept and the Data we Have: The Case of Labor Markets in the USA. *Environment and Planning A: Economy and Space 52*(7), 1395–1414.

Frandsen, B. R. (2017). Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design when the Running Variable is Discrete. *Regression Discontinuity Designs (Advances in Econometrics) 38*, 281–315.

Frandsen, B. R. (2021). The Surprising Impacts of Unionization: Evidence from Matched Employer-Employee Data. *Journal of Labor Economics 39*(4), 861–894.

Franko, W. W. and C. M. Witko (forthcoming). Unions, Class Identification and Policy Attitudes. *The Journal of Politics.*

Freeman, R. B. (2003). What Do Unions Do ... to Voting? *NBER Working Paper No. 9992*.

Freeman, R. B. and M. M. Kleiner (1990a). Employer Behavior in the Face of Union Organizing Drives. *ILR Review 43*(4), 351–365.

Freeman, R. B. and M. M. Kleiner (1990b). The Impact of New Unionization on Wages and Working Conditions. *Journal of Labor Economics 8*, S8–S25.

Freeman, R. B. and J. L. Medoff (1979). The Two Faces of Unionism. *The Public Interest 57*, 69–93.

Freeman, R. B. and J. L. Medoff (1984). *What Do Unions Do?* New York: Basic Books.

Frymer, P. and J. M. Grumbach (2021). Labor Unions and White Racial Politics. *American Journal of Political Science 65*(1), 225–240.

Gardner, J. (2021). Two-Stage Differences in Differences. *arXiv 2207.05943*.

Gehring, K. (2022). Can External Threats Foster a European Union Identity? Evidence from Russia's Invasion of Ukraine. *The Economic Journal 132*, 1489–1516.

Gelman, A. and G. Imbens (2019). Why High-order Polynomials should not be used in Regression Discontinuity Designs. *Journal of Business & Economic Statistics 37*(3), 447–456.

Gennaioli, N. and G. Tabellini (2023). Identity Politics. *CEPR Discussion Papers 18055*.

Gentzkow, M. and J. M. Shapiro (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics 126*(4), 1799–1839.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Econometrica 87*(4), 1307–1340.

Gerring, J. (2001). *Party Ideologies in America, 1828-1996.* Cambridge University Press.

Gethin, A., C. Martínez-Toledano, and T. Piketty (2022). Brahmin Left versus Merchant Right: Changing Political Cleavages in 21 Western Democracies, 1948–2020. *The Quarterly Journal of Economics 137*(1), 1–48.

Ghaly, M., A. Kostakis, and K. Stathopoulos (2021). The (Non-) effect of Labor Unionization on Firm Risk: Evidence from the Options Market. *Journal of Corporate Finance 66*, 101816.

Gimpel, J. and J. Glenn (2019). Racial Proximity and Campaign Contributing. *Electoral Studies 57*, 79–89.

Gimpel, J., F. Lee, and J. Kaminski (2006). The Political Geography of Campaign Contributions in American Politics. *The Journal of Politics 68*(3), 626–639.

Glaeser, E., G. Ponzetto, and J. Shapiro (2005). Strategic Extremism: Why Republicans and Democrats Divide on Religious Values. *The Quarterly Journal of Economics 120*(4), 1283–1330.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020). Bartik Instruments: What, When, Why, and How. *American Economic Review 110*(8), 2586–2624.

Goncalves, F. and S. Mello (2021). A Few Bad Apples? Racial Bias in Policing. *American Economic Review 111*(5), 1406–1441.

Goodman-Bacon, A. (2021). Difference-in-Differences with Variation in Treatment Timing. *Journal of Econometrics 225*(2), 254–277.

Grembi, V., T. Nannicini, and U. Troiano (2016). Do Fiscal Rules Matter? *American Economic Journal: Applied Economics 8*, 1–30.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer (2019). Fake News on Twitter during the 2016 US Presidential Election. *Science 363*(6425), 374–378.

Grosjean, P., F. Masera, and H. Yousaf (2023). Inflammatory Political Campaigns and Racial Bias in Policing. *The Quarterly Journal of Economics 138*(1), 413–463.

Guess, A., J. Nagler, and J. Tucker (2019). Less than you Think: Prevalence and Predictors of Fake News Dissemination on Facebook. *Science Advances 5*(1), eaau4586.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2021). 3G Internet and Confidence in Government. *The Quarterly Journal of Economics 136*(4), 2533–2613.

Hadzic, D., D. Carlson, and M. Tavits (2020). How Exposure to Violence affects Ethnic Voting. *British Journal of Political Science 50*(1), 345–362.

Hagedorn, J., C. A. Paras, H. Greenwich, and A. Hagopian (2016). The Role of Labor Unions in Creating Working Conditions That Promote Public Health. *American Journal of Public Health 106*(6), 989–995.

Haines, D. W. (2012). *Safe Haven? A History of Refugees in America.* Kumarian Press.

Halberstam, Y. and B. Knight (2016). Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter. *Journal of Public Economics 143*, 73–88.

Halla, M., A. Wagner, and J. Zweimüller (2017). Immigration and Voting for the Far Right. *Journal of the European Economic Association 15*(6), 1341–1385.

Hanson, G. H. (2006). Illegal Migration from Mexico to the United States. *Journal of Economic Literature 44*(4), 869–924.

Hare, C. and K. T. Poole (2014). The Polarization of Contemporary American Politics. *Polity 46*(3), 411–429.

Hertel-Fernandez, A. (2020). Power and Politics in the U.S. Workplace. Technical report, Economic Policy Institute.

Hill, S. J. and G. A. Huber (2017). Representativeness and Motivations of the Contemporary Donorate: Results from Merged Survey and Administrative Records. *Political Behavior 39*(1), 3–29.

Hindustan Times (2019). BJP Outspends Congress, Others in Social Media Advertising. *Accessed last on July 19 2023 via www.hindustantimes.com/lok-sabha-elections/bjp-outspends-congress-others-in-social-media-advertising/story-FHByCC5vUfs7xCvD9kDY5L.html*.

Hirsch, B. T. and D. A. Macpherson (2022). Union Membership and Coverage Database from the Current Population Survey. Accessed via https://www.unionstats.com.

Hjort, J. and J. Poulsen (2019). The Arrival of Fast Internet and Employment in Africa. *American Economic Review 109*(3), 1032–79.

Hodler, R. and P. A. Raschky (2014). Regional Favoritism. *The Quarterly Journal of Economics 129*(2), 995–1033.

Hollibaugh Jr, G. E. and L. S. Rothenberg (2018). The Who, When, and Where of Executive Nominations: Integrating Agency Independence and Appointee Ideology. *American Journal of Political Science 62*(2), 296–311.

Hopkins, D. A. (2017). *Red Fighting Blue: How Geography and Electoral Rules Polarize American Politics*. Cambridge University Press.

Hübscher, E., T. Sattler, and M. Wagner (2020). Does Austerity Cause Polarization? *British Journal of Political Science*, 1–19.

Human Rights Watch (2019). Violent Cow Protection in India. Technical report, Delhi.

ITU (2020). Measuring Digital Development Facts and Figures. Technical report, Geneva, CH.

Iversen, T. and D. Soskice (2015). Information, Inequality, and Mass Polarization: Ideology in Advanced Democracies. *Comparative Political Studies 48*(13), 1781–1813.

Iyengar, S., G. Sood, and Y. Lelkes (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly 76*(3), 405–431.

Jaeger, D., J. Ruist, and J. Stuhler (2018). Shift-Share Instruments and the Impact of Immigration. *NBER Working Paper No. 24285*.

Jennings, M. K. and R. G. Niemi (2015). *Political Character of Adolescence: The Influence of Families and Schools*. Princeton University Press.

Jeong, D., A. Shenoy, and L. V. Zimmermann (2023). De Jure versus De Facto Transparency: Corruption in Local Public Office in India. *Journal of Public Economics 221*, 104855.

Kantar (2021). Internet Adoption in India. Technical report, Accessed July 19th 2023 via https://images.assettype.com/afaqs/2021-06/b9a3220f-ae2f-43db-a0b4-36a372b243c4/KANTAR$_{I}CUBE_{2}020_{R}eport_{C}1.pdf$.

Kantar (2023). Internet Adoption in India. Technical report, Accessed July 19th 2023 via https://www.iamai.in/sites/default/files/research/Internet%20in%20India%202022-_Print%20version.pdf.

Kerrissey, J. and E. Schofer (2013). Union Membership and Political Participation in the United States. *Social Forces 91*(3), 895–928.

Kim, S. E. and Y. Margalit (2017). Informed Preferences? The Impact of Unions on Workers' Policy Views. *American Journal of Political Science 61*(3), 728–743.

King, M. L. (1965). Dr. King Understood the Power of Unions. Accessed via https://ctaflcio.org/news/dr-king-understood-power-unions. Blog by AFSCME president Lee Saunders.

Kleiner, M. M. (2001). Intensity of Management Resistance: Understanding the Decline of Unionization in the Private Sector. *Journal of Labor Research 22*, 519–540.

Knepper, M. (2020). From the Fringe to the Fore: Labor Unions and Employee Compensation. *The Review of Economics and Statistics 102*(1), 98–112.

Krishnan, V. (2018). Standing Committee Report Summary: Progress of Implementation of BharatNet. *RS Legislative Research*.

Laitin, D. D. (1998). *Identity in Formation: The Russian-speaking Populations in the Near Abroad*. Cornell University Press.

Lee, D. S. and A. Mas (2012). Long-run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961-1999. *The Quarterly Journal of Economics 127*, 333–378.

Leetaru, K. and P. A. Schrodt (2013). GDELT: Global Data on Events, Location, and Tone, 1979–2012. In *ISA Annual Convention*, Volume 2, pp. 1–49. Citeseer.

Leighley, J. E. and J. Nagler (2007). Unions, Voter Turnout, and Class Bias in the U.S. Electorate, 1964-2004. *The Journal of Politics 69*(2), 430–441.

Levy, R. (2021). Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review 111*(3), 831–870.

Li, L., S. Rohlin, and P. Singleton (2022). Labor Unions and Workplace Safety. *ILR Review 75*(2), 402–426.

Lindvall, J. (2013). Union Density and Political Strikes. *World Politics 65*(3), 539–569.

Lipset, S. M. (1960). *Political Man. The Social Bases of Politics*, Volume 492. London.

Lipset, S. M. (1983). Radicalism or Reformism: The Sources of Working-class Politics. *American Political Science Review 77*(1), 1–18.

Logan, J. (2002). Consultants, Lawyers, and the 'Union Free' Movement in the USA since the 1970s. *Industrial Relations Journal 33*, 197–214.

Lonsky, J. (2020). Does Immigration Decrease Far-Right Popularity? Evidence from Finnish Municipalities. GLO Discussion Paper Series 540, Global Labor Organization (GLO).

Lowe, M. (2021). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. *American Economic Review 111*(6), 1807–1844.

Lowes, S. and E. Montero (2021). The Legacy of Colonial Medicine in Central Africa. *American Economic Review 111*(4), 1284–1314.

Macdonald, D. (2021). How Labor Unions Increase Political Knowledge: Evidence from the United States. *Political Behavior 43*, 1–24.

Manacorda, M., G. Tabellini, and A. Tesei (2023). Mobile Internet and the Rise of Communitarian Politics. *CEPR Discussion Papers 18063*.

Manacorda, M. and A. Tesei (2020). Liberation Technology: Mobile Phones and Political Mobilization in Africa. *Econometrica 88*(2), 533–567.

Margalit, Y. (2019). Economic Insecurity and the Causes of Populism, Reconsidered. *Journal of Economic Perspectives 33*(4), 152–170.

Martin, G. and Z. Peskowitz (2018). Agency Problems in Political Campaigns: Media Buying and Consulting. *American Political Science Review 112*(2), 231–248.

Mason, L. (2015). "I Disrespectfully Agree": The Differential Effects of Partisan Sorting on Social and Issue Polarization. *American Journal of Political Science 59*(1), 128–145.

Matzat, J. (2023). Fueling Divisions? The Arrival of Fast Internet in Indian Villages. *Mimeo*.

Matzat, J. and A. Schmeißer (2023). Do Unions Shape Political Ideologies at Work? *CESifo Working Paper 10301*.

Mayda, A. (2006). Who Is Against Immigration? A Cross-Country Investigation Of Individual Attitudes Toward Immigrants. *The Review of Economics and Statistics 88*(3), 510–530.

Mayda, A.-M., C. Parsons, H. Pham, and P.-L. Vézina (2022). Refugees and Foreign Direct Investment: Quasi-experimental Evidence from US Resettlements. *Journal of Development Economics 156*, 102818.

Mayda, A. M., G. Peri, and W. Steingress (2022). The Political Impact of Immigration: Evidence from the United States. *American Economic Journal: Applied Economics 14*(1), 358–89.

McAdam, D., S. Tarrow, and C. Tilly (2001). *Dynamics of Contention*. New York: Cambridge University Press.

McCarty, N. and L. S. Rothenberg (1996). Commitment and the Campaign Contribution Contract. *American Journal of Political Science 40*(3), 872–904.

McGuirk, E. and M. Burke (2020). The Economic Origins of Conflict in Africa. *Journal of Political Economy 128*(10), 3940–3997.

McKenzie, D. (2017). Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review 107*(8), 2278–2307.

McKinsey Global Institute (2019). *Digital India: Technology to Transform a Connected Nation*. Delhi.

McLuhan, M. and B. R. Powers (1989). *The Global Village: Transformations in World Life and Media in the 21st century*. Communication and society.

Méndez, E. and D. Van Patten (2022). Multinationals, Monopsony, and Local Development: Evidence from the United Fruit Company. *Econometrica 90*(6), 2685–2721.

Ministry of Communications (2021). Induction Note. Technical report, As on 02.09.2021, accessed July 14th 2022 via https://dot.gov.in/sites/default/files/FINAL%20INDUCTION-NOTE.pdf.

Ministry of Rural Development (2014). Frequently Asked Questions on MGNREGA Operational Guidelines. Technical report.

Mitra, A. and D. Ray (2014). Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India. *Journal of Political Economy 122*(4), 719–765.

Mocanu, D., L. Rossi, Q. Zhang, M. Karsai, and W. Quattrociocchi (2015). Collective Attention in the Age of (Mis-) information. *Computers in Human Behavior 51*, 1198–1204.

Mosimann, N. and J. Pontusson (2017). Solidaristic Unionism and Support for Redistribution in Contemporary Europe. *World Politics 69*(3), 448–492.

Mousa, S. (2020). Building Social Cohesion between Christians and Muslims through Soccer in Post-ISIS Iraq. *Science 369*(6505), 866–870.

Müller, K. and C. Schwarz (2021). Fanning the Flames of Hate: Social Media and Hate Crime. *Journal of the European Economic Association 19*(4), 2131–2167.

Müller, K., C. Schwarz, et al. (2022). From Hashtag to Hate Crime: Twitter and Anti-minority Sentiment. *American Economic Journal: Applied Economics 15*(3), 270–312.

Mutz, D. (1995). Effects of Horse-Race Coverage on Campaign Coffers: Strategic Contributing in Presidential Primaries. *The Journal of Politics 57*(4), 1015–1042.

New York Times (2019). Under Modi, a Hindu Nationalist Surge Has Further Divided India. *Accessed last on July 19 2023 via https://www.nytimes.com/2019/04/11/world/asia/modi-india-elections.html.*

New York Times (2021). In India, Facebook Grapples With an Amplified Version of its Problems. *Accessed last on July 19 2023 via https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html.*

Nickerson, D. W. (2008). Is Voting Contagious? Evidence from Two Field Experiments. *American Political Science Review 102*(1), 49–57.

Nikolka, T. and P. Poutvaara (2016). Brexit—Theory and Empirics. *CESifo Forum 17*(4), 68–75.

NILRR (2021). Big Labor Reported Spending $1.8 Billion on Politics in the 2020 Election Cycle. *Special Report by the National Institute for Labor Relations Research.*

Nokia (2022). India Mobile Broadband Index. Technical report, Accessed July 19th 2023 via https://www.nokia.com/sites/default/files/2022-03/nokia-mbit-2022.pdf.

Norris, P. and R. Inglehart (2019). *Cultural Backlash: Trump, Brexit, and Authoritarian Populism.* Cambridge University Press.

Nyhan, B. and J. Montgomery (2015). Connecting the Candidates: Consultant Networks and the Diffusion of Campaign Strategy in American Congressional Elections. *American Journal of Political Science 59*(2), 292–308.

Obama, B. (2010). President Obama on Labor Day: The Fight for America's Workers Continues. Accessed via https://obamawhitehouse.archives.gov/blog/2010/09/06/president-obamalabor-day-fight-americas-workers-continues. Blog by Jesse Lee, The White House.

Olea, J. L. M. and C. Pflueger (2013). A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics 31*(3), 358–369.

Otto, A. and M. Steinhardt (2014). Immigration and Election Outcomes: Evidence from City Districts in Hamburg. *Regional Science and Urban Economics 45*, 67–79.

Özden, Ç., C. R. Parsons, M. Schiff, and T. L. Walmsley (2011). Where on earth is Everybody? The Evolution of Global Bilateral Migration 1960–2000. *The World Bank Economic Review 25*(1), 12–56.

Panunzi, F., N. Pavoni, and G. Tabellini (2020). Economic Shocks and Populism: The Political Implications of Reference-Dependent Preferences. *CESifo Working Paper No. 8539.*

Parsons, C. and P. Vézina (2018). Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment. *The Economic Journal 128*(612), 210–234.

Pettigrew, T. and L. Tropp (2006). A Meta-Analytic Test of Intergroup Contact Theory. *Journal of Personality and Social Psychology 90*(5), 751–783.

PEW (2014). Political Polarization in the American Public.

PEW (2016). Partisanship and Political Animosity in 2016.

PEW (2017). The Shift in the American Public's Political Values.

PEW (2021). Religion in India: Tolerance and Segregation. Technical report.

Poole, K. T. and H. Rosenthal (1985). A Spatial Model For Legislative Roll Call Analysis. *American Journal of Political Science 29*(2), 357–384.

Poole, K. T. and H. Rosenthal (2000). *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press on Demand.

Poole, K. T. and H. Rosenthal (2001). D-NOMINATE After 10 Years: A Comparative Update to Congress: A Political-Economic History of Roll-Call Voting. *Legislative Studies Quarterly 26*(1), 5–29.

Poole, K. T. and H. Rosenthal (2017). *Ideology & Congress: A Political Economic History of Roll Call Voting*. Routledge.

Portes, A. (2011). America and its Immigrants: A Game of Mirrors. *Proceedings of the American Philosophical Society 155*(4), 418–32.

Putnam, R. D. (2000). Bowling Alone: America's Declining Social Capital. In *Culture and Politics: A Reader*, pp. 223–234. Springer.

Putnam, R. D. (2007). E Pluribus Unum: Diversity and Community in the Twenty-first Century the 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies 30*(2), 137–174.

Raleigh, C., R. Linke, H. Hegre, and J. Karlsen (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research 47*(5), 651–660.

Rodrik, D. (2021). Why Does Globalization Fuel Populism? Economics, Culture, and the Rise of Right-wing Populism. *Annual Review of Economics 13*, 133–170.

Rosenfeld, J. (2014). *What Unions No Longer Do*. Cambridge, MA: Harvard University Press.

Rosenzweig, M. R. and B. Xu (2023). Classroom Competition, Student Effort, and Peer Effects.

Roth, J. and P. H. C. Sant'Anna (2023). When Is Parallel Trends Sensitive to Functional Form? *Econometrica*.

Ruggles, S., S. Flood, R. Goeken, J. Grover, E. Meyer, J. Pacas, and M. Sobek (2020). IPUMS USA: Version 10.0 [dataset]. Minneapolis, MN: IPUMS, https://doi.org/10.18128/D010.V10.0 .

Russ, D. E., K.-Y. Ho, J. S. Colt, K. R. Armenti, D. Baris, W.-H. Chow, F. Davis, A. Johnson, M. P. Purdue, M. R. Karagas, K. Schwartz, M. Schwenn, D. T. Silverman, C. A. Johnson, and M. C. Friesen (2016). Computer-based Coding of Free-text Job Descriptions to Efficiently Identify Occupations in Epidemiological Studies. *Occupational and Environmental Medicine 73*(6), 417–424.

Sahoo, N. (2020). Mounting Majoritarianism and Political Polarization in India. Technical report, Carnegie Endowment for International Peace.

Sani, G. and G. Sartori (1983). Polarization, Fragmentation and Competition in Western Democracies. *Western European Party Systems: Continuity and Change*, 307–340.

Satyanarayana, J., K. Karnik, S. Sadagopan, A. K. Bhargava, S. Mittal, T. Ray, A. Sundararajan, R. Moona, and V. Umashankar (2015). Report of the Committee on National Optical Fibre Network (NOFN). Technical report, Delhi.

Schlozman, D. (2015). *When Movements Anchor Parties: Electoral Alignments in American History*. Princeton, NJ: Princeton University Press.

Schmitt, J. and B. Zipperer (2009). Dropping the Ax: Illegal Firings During Union Election Campaigns, 1951-2007. *CEPR Working Papers*.

Schuster, S. S. (2020). Does Campaign Spending Affect Election Outcomes? New Evidence from Transaction-Level Disbursement Data. *The Journal of Politics 82*(4), 1502–1515.

Sherif, M., O. Harvey, B. White, W. Hood, and C. Sherif (1961). *Intergroup Conflict and Cooperation: The Robbers Cave Experiment. Vol 10*. Norman, OK: University Book Exchange.

Sojourner, A. and J. Yang (2022). Effects of Union Certification on Workplace-Safety Enforcement: Regression-Discontinuity Evidence. *ILR Review 75*(2), 373–401.

Sojourner, A. J. (2013). Do Unions Promote Members' Electoral Office Holding? Evidence from Correlates of State Legislatures' Occupational Shares. *ILR Review 66*(2), 467–486.

Sojourner, A. J., B. R. Frandsen, R. J. Town, D. C. Grabowski, and M. M. Chen (2015). Impacts of Unionization on Quality and Productivity: Regression Discontinuity Evidence From Nursing Homes. *ILR Review 68*(4), 771–806.

Spenkuch, J. L. and P. Tillmann (2018). Elite Influence? Religion and the Electoral Success of the Nazis. *American Journal of Political Science 62*(1), 19–36.

Stansbury, A. and L. H. Summers (2020). The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy. *NBER Working Paper No. 27193*.

Steinmayr, A. (2021). Contact Versus Exposure: Refugee Presence and Voting for the Far Right. *Review of Economics and Statistics 103*(2), 310–327.

Stone, W. J. and E. N. Simas (2010). Candidate Valence and Ideological Positioning in the U.S. House Elections. *American Journal of Political Science 54*(2), 371–388.

Stonecash, J. M., M. D. Brewer, and M. D. Mariani (2018). *Diverging Parties: Social Change, Realignment, and Party Polarization*. Routledge.

Stuckatz, J. (2022). How the Workplace Affects Employee Political Contributions. *American Political Science Review 116*(1), 54–69.

Sun, L. and S. Abraham (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics 225*(2), 175–199.

Terriquez, V. (2011). Schools for Democracy: Labor Union Participation and Latino Immigrant Parents' School-Based Civic Engagement. *American Sociological Review 76*(4), 581–601.

Teso, E. (2022). What Drives U.S. Corporate Elites' Campaign Contribution Behavior? *CEPR Discussion Paper 16966*.

The Wire (2019). 100+ Instances of Hate Speech, Religious Polarisation, Hindutva Supremacy in Adityanath's Poll Speeches. *Accessed last on July 5 2023 via https://thewire.in/communalism/100-instances-of-hate-speech-religious-polarisation-hindutva-supremacy-in-adityanaths-poll-speeches*.

Thomsen, D. (2014). Ideological Moderates Won't Run: How Party Fit Matters for Partisan Polarization in Congress. *The Journal of Politics 76*(3), 786–797.

Tichenor, D. J. (2009). *Dividing Lines: The Politics of Immigration Control in America*, Volume 104. Princeton University Press.

Time (2019). How Volunteers for India's Ruling Party Are Using WhatsApp to Fuel Fake News Ahead of Elections. *Accessed last on July 19 2023 via https://time.com/5512032/whatsapp-india-election-2019/.*

TRAI (2018). Telecom at a Glance. Technical report, Delhi.

TRAI (2019). The Indian Telecom Services. Technical report, Delhi.

Verma, A. (2005). What Do Unions Do to the Workplace? Union Effects on Management and HRM Policies. *Journal of Labor Research 26*(3), 415–449.

Vosoughi, S., D. Roy, and S. Aral (2018). The Spread of True and False News Online. *Science 359*(6380), 1146–1151.

Wang, S. and S. Young (2021). Unionization, Employer Opposition, and Establishment Closure. *Mimeo*.

Wang, T. (2021). Media, Pulpit, and Populist Persuasion: Evidence from Father Coughlin. *American Economic Review 111*(9), 3064–3092.

Wasi, N. and A. Flaaen (2015). Record Linkage Using Stata: Preprocessing, Linking, and Reviewing Utilities. *Stata Journal 15*(3), 672–697.

Weidmann, N. B. (2016). A Closer Look at Reporting Bias in Conflict Event Data. *American Journal of Political Science 60*(1), 206–218.

Western, B. and J. Rosenfeld (2011). Unions, Norms, and the Rise in U.S. Wage Inequality. *American Sociological Review 76*(4), 513–537.

Whitaker, L. P. (2018). Campaign Finance Law: An Analysis of Key Issues, Recent Developments, and Constitutional Considerations for Legislation. *Congressional Research Service Report R45320*.

Wooldridge, J. (2015). Control Function Methods in Applied Econometrics. *Journal of Human Resources 2*, 420–425.

World Bank (2021). World Development Indicators 2021. *The World Bank, Washington DC*.

WSJ (2012). Political Spending by Unions Far Exceeds Direct Donations. *Wall Street Journal*. July 10, 2012, Accessed via https://www.wsj.com/articles/SB10001424052702304782404577488584031850026.

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Evidence from the Rwandan Genocide. *The Quarterly Journal of Economics 129*(4), 1947–1994.

Zhuravskaya, E., M. Petrova, and R. Enikolopov (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics 12*, 415–438.

Zimmermann, L. (2014). Public Works Programs in Developing Countries have the Potential to Reduce Poverty. *IZA World of Labor*.

Zimmermann, L. (2021). The Dynamic Electoral Returns of a Large Antipoverty Program. *Review of Economics and Statistics 103*(5), 803–817.

Acknowledgments

This thesis has benefited immensely from the guidance and advice of my supervisor Axel Dreher. Thank you for always being ready to provide feedback, for distilling my thoughts down to the essential aspects, for creating such a great team, for introducing me to the world of research, and for consistently asking the right questions.

I would also like to thank my second supervisor Krisztina Kis-Katos. Your enthusiasm and curiosity are contagious, and your thoughtful leadership of our graduate program has contributed to the wonderful experience I had. Lastly, I want to thank my third supervisor Andreas Fuchs, who has always effortlessly integrated me into his chair whenever I visited Göttingen.

My Ph.D. journey began in a small student dorm in Delhi, where I received a nudge at just the right moment. Thank you, Marcello, for the email that led me to GLAD. Without it, I would have missed out on so much. The fantastic peers I had throughout this journey made it easier to overcome the small and big challenges. To have such a great group spirit every step of the way made a big difference: Thank you, Laura, Anna G., Lukas, Tobi, Andrea, Henry, Anna R., Lisa, Richard, Coco, Tatiana, Claudi, and Friederike. I also would like to thank you Benni for always checking on me.

I was fortunate to spend my Ph.D. years in two locations. I learned a lot from you Valentin, Gerda, Angelika, Lenny, John, and Sven. I would like to thank especially Sarah for always providing useful advice. You made me feel at home in Heidelberg and I connect great memories to all of you: Anna H. (volleyball, roses, hikes), Cristina (77A life), Teresa (rowing, singing), Marina (talks, rat), Martin (coffee breaks), Hilal (Istanbul, fortune telling), Ferit (work-buddy), Robin (what are the odds?, biking in Dossenheim), Jingke (first beer), Paula (concerts), Charlotte (good parts of office life), Annika (friendly chaos), Christoph (Austrian Schmäh), Matthias and David (spikeball), Laura (Chicago), Tobi and Regina (football), Oliver (RHCP), and Zain (my joker for South Asian contextual knowledge). I also want to thank you, Svenja, for being so supportive and for joining me in Heidelberg.

It would not have been the same without the two amazing office mates I had. Lukas, I will always keep the memories of late work sessions with good music. Charlotte, you are such a kind and thoughtful companion. The wonderful sense of humor that both of you

share is a true joy.
I would also like to say thank you to my brilliant co-authors. It has been a pleasure to work with you and I learned a lot. Thank you, Aiko, Chris, Sarah, and Axel!
Finally, I would like to thank my family. Mum and Dad, you always knew how to keep my curiosity alive. I'm grateful for the perspective on life you provided and the constant support you have given me. Brigitte, thank you for always encouraging me on my journey. Julian, growing up with you was fun, I learn much from you.

I am grateful to the following colleagues for their essential feedback on my chapters.

Chapter 1: We are grateful for helpful comments from Samuel Bazzi, Claudio Ferraz, Kai Gehring, Ilona Lahdelma, Gianmarco Ottaviano, Anna-Maria Mayda, Panu Poutvaara, Hillel Rapoport, Yang-Yang Zhou and participants at the following conferences and seminars: Georg-August University Goettingen (Development Economics Seminar and CeMig Research Colloquium), Heidelberg University, International Political Economy Society (IPES, 2020), Migration Politics Workshop (King's College, 2021), European Public Choice Society (EPCS, 2021), Spring Meeting of Young Economists (SMYE, 2021), University of Milano Bicocca (Hosting Refugees: National Policies and Local Communities Workshop, 2022), University of Bern (Research Seminar, 2022), CEMIR Junior Economist Workshop on Migration Research (Ifo, 2022), and University of Bayreuth (2023). We thank Vasil Yasenov for sharing the PRM refugee data with us. We thank Clara Coelho, Tobias Hellmundt, Lucero Carballo Madrigal, Kaja Rupieper, Aiko Schmeisser and Mareike Weiss for excellent research assistance.

Chapter 2: I thank Jenny Aker, Theodore Alysandratos, Filipe Campante, Zain Chaudry, Cristina Cibin, Clément Imbert, Rajesh Ramachandran, Charlotte Robert, Claudia Schupp and participants of the seminar at the University of Göttingen for helpful comments. I am grateful to Riyoko Shibe, Aiko Schmeißer, Paul Vogel and Gesa Woyke for excellent research assistance.

Chapter 3: We thank Marco Caliendo, Filipe Campante, David Card, Axel Dreher, Henry Farber, Katrin Huber, Simon Luechinger, Dennis Quinn, Sulin Sardoschau, Thomas Siedler, and Samuel Young, as well as participants at the Workshop on Political Economy (Dresden, 2022), Berlin School of Economics Summer Workshop (Berlin, 2022), Research Seminar for Applied Microeconomics (FU Berlin, 2022), Beyond Basic Questions Workshop (Bern, 2022), GlaD-DENeB Workshop (Goettingen, 2022), Development Economics Seminar (Goettingen, 2022), Potsdam Research Seminar in Economics (Potsdam, 2023), and the IAAEU Colloquium on Economics (Trier, 2023) for helpful comments. We are grateful to Sunna Hügemann, Luise Koch, and Yasmin Zysk for excellent research assistance.

4
Declaration for Admission

I, Johannes Matzat, confirm

1. that the dissertation ”Essays on Political Economy: Drivers of Polarization” that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,

2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,

3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,

   - No remuneration or equivalent compensation were provided
   - No services were engaged that may contradict the purpose of producing a doctoral thesis

4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Date, Signature:

5
Author Contributions

This thesis consists of three independent chapters. The contributions of each co-author to the respective chapter are summarized in the following.

Chapter 1: Immigration, Political Ideologies and the Polarization of American Politics
This chaper is co-authored with Axel Dreher, Sarah Langlotz and Christopher Parsons. All of us contributed to the conceptualization. Most of the data preparation and analysis were carried out by myself. The writing was mainly done by Axel Dreher and Christopher Parsons with some contributions by myself.

Chapter 2: Fueling Divisions? The Arrival of Fast Internet in Indian Villages
This chapter is single-authored. The conceptualization, data preparation and analysis, and the writing were carried out by myself.

Chapter 3: Do Unions Shape Political Ideologies At Work?
This chapter is co-authored with Aiko Schmeißer. We equally contributed to the conceptualization, data preparation and analysis, as well as the writing.

Date, Signature: