Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

ECONOMICS

Ciclo 33

Settore Concorsuale: 13/A1 - ECONOMIA POLITICA

Settore Scientifico Disciplinare: SECS-P/01 - ECONOMIA POLITICA

ESSAYS IN EMPIRICAL POLITICAL ECONOMICS

Presentata da: Christoph Michael Pfeufer

Coordinatore Dottorato

Maria Bigoni

Supervisore

Paolo Masella

Co-supervisore

Maria Bigoni

Esame finale anno 2022

Contents

List of Figures

List of Tables

Chapter 1

Trust and Feudal Institutions - Evidence from Germany[1]

Abstract

Does historical exposure to feudalism impair interpersonal trust today? This paper examines the long-run consequences of feudalism, a set of labor coercion and migration restrictions that deprived the peasantry from access to markets and fostered social closure. I combine survey-data from modern Germany with administrative data measuring the intensity of feudalism at the county-level in the Prussian Empire (1816 - 1849) to test whether feudal repercussions continue to depress interpersonal trust today. I find that a greater legacy of feudalism is associated with lower interpersonal trust; a result that is robust to taking into account rich sets of historical and socio-demographic controls. Moreover, I adopt a mover-design to isolate the culturally embodied part of feudalism from place-based confounders and demonstrate that distrust against strangers persists. Last, I propose a new instrument for feudalism based on the interaction of naturally occurring differences in the share of sandy soil and population losses in the Thirty Years War (1618 - 1648). Overall, the results suggest that the shadow of the coercive feudal restrictions continues to depress cultural openness towards foreigners until today.

[1] I wish to thank seminar participants at the University of Bologna for their comments and suggestions. I am also grateful to the DIW in Berlin for granting me access to the data and supporting me on site. This paper profited also greatly from advice of Matteo Cervellati, Margherita Fort and Luigi Pascalli. Mostly, I am indebted to Paolo Masella for discussions, guiding and support.

1.1 Introduction

Any social interaction and economic transaction contains the risk of defection by the counterpart. While ex-post mechanisms to punish actual defection are rife, a belief of trustworthiness is necessary to enter and then cooperate in these interactions. Indeed, interpersonal trust is conducive to financial development (Guiso, Sapienza and Zingales, 2004), trade (Guiso, Sapienza and Zingales, 2009), firm management practices (Bloom, Sadun and Van Reenen, 2012), the creation of political institutions (Putnam, 2000) and, as a consequence, economic development (Tabellini (2008*a*), Algan and Cahuc (2010)).

As trust attitudes across and, even more so, within countries are distributed widely heterogeneous (Falk et al., 2018), I ask whether different historical experiences with feudalism account for modern trust values. Feudalism was a pivotal medieval institution that set the eastern and western parts of Europe on diverging paths of development (Alesina, Glaeser and Glaeser, 2004). The main features of feudalism were severe labor coercion and migration restrictions, that fostered *social closure* by preventing outmigration from feudal areas while curtailing in-migration due to unfavorable economic conditions to its peasants (Rosenberg (1944); Ogilvie (2014)). Additionally, feudal institutions deprived peasants from participation in goods markets (Ogilvie, 2014) and thus led to a 'localization of economic activity' (Tabellini, 2008*b*). Hence, feudalism promoted interactions within members of a kinship, while reducing (the value of) cooperation with members of the out-group. Albeit the historical importance of feudal institutions, evidence on its long-term consequences, especially with regard to cultural values, is scarce.

This paper studies whether historical exposure to feudalism deteriorates contemporaneous trust beyond the kinship. To test this conjecture, I combine administrative data on the intensity of feudalism at the county-level in the Prussian Empire (1816 - 1849) with survey data on today's trust from the German Socio-Economic Panel (GSOEP). To estimate the causal effect of feudalism on trust, I employ OLS regressions enriched by a large set of historical and socio-demographic controls, the epidemiological approach that uses variation from inner-german movers to control for place-based confounders, and I develop an instrumental-variable that shields the estimates against the endogenous adoption of feudalism.

My setup offers several conceptual advantages. First, a within-country design alleviates (to a certain extent) a major concern in this type of study, namely omitted variable bias, by keeping cultural, historical and institutional features at the country level fixed. Second, I utilize the meticulousness of the Prussian administration that provide high-quality measures for historical feudalism in its last stage (i.e., 1816 - 1849) at great detail, i.e. the level of Prussian counties. Third, historians are unified in their notion, that ancient Germany offered the full range of severity in feudal contracts within a single territory (e.g., Rosenberg (1944)). While restricted to the spatial overlap between Prussia as of 1849 and modern Germany, this does not deprive me of the large variation in feudal arrangements, as Prussia covered areas constituting of low, medium and strong feudal ties.

For the analysis I combine digitized Prussian censuses, available through the Prussian Economic History Database, with survey data from the German Socio Economic Panel (GSOEP), a well-established longitudinal survey for 20,000 individuals for 1984 - 2016.[2] To measure feudalism, I construct a county-level index that uses principal component analysis on historical indicators at the extensive margin, for example landownership concentration, and the intensive margin, for example the level of redeemed labor services. I assign the feudalism index to modern day individuals based on the spatial overlap of their residence with the Prussian counties and measure their trust attitudes as in (Dohmen et al., 2011).

The main challenge for an unconfounded estimation of the causal effect of feudalism on trust attitudes is the presence of omitted variables, in particular, omitted factors that led to the historical adoption of feudalism and influence trust attitudes today. The analysis tackles this identification problem in three steps: (i) I conduct an individual-level analysis and directly account for the confounding effects of a large number of geographical, historical and socio-demographic controls as well as fixed effects at the different administrative levels; (ii) I restrict my sample to movers within Germany and perform a within-county of residence analysis, that absorbs any potential bias arising from *time-invariant* unobservables; and (iii) I use an instrumental variable strategy, that exploits a combination of the counties' share of sand paired with population losses in the Thirty Years War (1618 - 1648) as a determinant of feudalism and thereby cushions against concerns of unobserved *time-variant* factors, biasing the results.

First, I run OLS-regressions and find a robust and negative relation between historical feudalism and contemporary trust. This association is stable upon the inclusion of a wide range of potentially confounding factors that might have directly and independently affected the formation of trust in others. The analysis accounts for a large set of geographical, cultural, historical and economic characteristics that may explain the adoption of both, feudal

[2]For more information on the Prussian Economic History Database, see https://www.ifo.de/en/iPEHD. For more information on the SOEP, see https://paneldata.org/soep-core/.

institutions and generalized morality today. Specifically, I pay regard to several historical narratives that reflect upon the emergence of feudalism and can plausibly be tied to norms of generalized morality. Additionally, the analysis is performed from within variation of either historical provinces or federal states and thereby shields the estimates against influences from unobserved geographical, cultural or historical heterogeneity at those levels of aggregation.

Second, I apply the epidemiological approach, to separate the culturally embodied and portable element of feudalism from time-invariant unobservables, that exhibit finer variation than historical provinces (Fernández, 2011). I now use the municipality of birth as a proxy for their feudal ancestry and perform an OLS-estimation that includes a county of residence fixed effect, to absorbing all time-invariant confounders. The effect is identified by comparing the trust values from inner-german movers and non-movers. In this step, results are mixed. While the effects vanish on the measures of generalized trust, they remain significant on the measure of distrust against strangers. As the survey question for generalized trust asks for 'trust in others', it is well possible that respondents refer to 'others' as family, friends and neighbours, so agents whose trust has not been impaired by feudal restrictions.

Last, I pursue an instrumental variable strategy. I propose a new instrument for feudalism that combines naturally occurring variation in the geological composition of the soil, specifically the share of sand, with population losses in the Thirty Years War (1618 - 1648) to predict feudalism among Prussian counties (1816 - 1849). Historically, landlords reacted to jumps in the land-labor ratio, as induced by the Thirty Years War, by increasing labor coercion and mobility restrictions to ensure that their fields were tilled (Domar (1970); Klein and Ogilvie (2017)). However, in the aftermath of the Thirty Years War absolutist monarchs largely extended their state enterprises and armies, recruiting exclusively from the pool of noblemen. Thereby, landlords that owned inferior soil had relatively more incentives to lease the labor-intensive farming to their tenants and pursue military or governmental careers. In contrast, the desertion of high-quality land gave a stimulus to landlords to extend their demesne farming and to impose the necessary repressions to cultivate it Brenner (1976). Given that soil texture is a natural feature out of the scope of human intervention, the instrument distributed as good as random.

To create the instrument, I use data on the share of sandy soil from the Prussian Census and I digitized and geo-referenced historical maps on population losses in the Thirty Years War from de La Blache (1894), Franz (1979) and Vasold (1993) to get a complete picture of the losses. The first stage results support this notion and show that the combination of sandy soil and high population losses is a pre-historical determinant in the adoption of feudal institutions.

IV-Results are pending as the host-institution, DIW Berlin, does not accept researchers from foreign countries due to Covid-19 measures since March 2020.

Related Literature. This paper refers to the literature on cultural persistence, that demonstrates that social norms may display longevity over hundreds of years (Voigtländer and Voth, 2012*a*) and are determined by factors such as geography (Galor and Özak, 2016), agricultural practices (Alesina, Giuliano and Nunn, 2013), climate (Buggle and Durante, 2017) or institutions (Becker et al., 2016). Dohmen et al. (2011) provide evidence, that trust attitudes are strongly correlated between parents and children. The evidence in this paper most directly complements studies that assess the traces of historical shocks on interpersonal trust (Nunn and Wantchekon (2011), Tabellini (2008*a*)). My results contribute to that literature by highlighting the role of historical labor coercion and migration restrictions as a determinant of modern day trust values. Thereby, my findings provide empirical support for the prediction in Tabellini (2008*b*) that a 'localization of economic activity' fosters limited morality above kinship boundaries.

Additionally, this paper speaks to a literature on the consequences of coercive labor institutions. Serfdom has been shown to have detrimental effects on development through discouraging labour mobility (Nafziger, 2012), lower agricultural productivity (Markevich and Zhuravskaya, 2018), dampening the acquisition of human capital (Cinnirella and Hornung, 2016), preventing structural change and spatial misallocation of firms (Buggle and Nafziger, 2018).[3] This work complements previous work and emphasizes that feudalism discourages economic and social interactions through a distrust in strangers. Moreover, it is the first study to show the long-lasting impact of labor coercion in the context of Europe.

The paper proceeds as follows. Section 2 describes the historical background and explains how historical feudalism can be tied to nowadays trust conceptually. Section 3 lists the sources of data and discusses the measurement and determinants of the geographic distribution in feudalism. Section 4 presents the empirical strategy and the estimation of the impact of feudalism on trust. Section 5 introduces the instrumental variable approach. Section 6 concludes.

[3] See Ashraf et al. (2018) for a paper on the determinants of the demise of serfdom. They show that capital-skill complementarity incentivized landlords to abandon feudal institutions in order to reap the benefits of the emerging industrialization.

1.2 Historical Background & Conceptual Framework

Historical Background

Feudalism was known in many lands and lived in some places in Europe for over a thousand years with a different meaning at different times in different places (Blum, 1957). This makes it an inherently difficult object to define comprehensively. This paper denotes it, rather broadly, as the various forms of unfreedom the rural population experienced, through an institutionalized form of economic and legal subjection by their landlords. This definition covers a plethora of variants of feudalism but is confined by two modes, that describe the ends of the spectrum of feudal restrictions. In its most severe form feudalism is hardly distinguishable from serfdom. While serfdom commonly refers to a form of *personal* bondage, i.e. a serf is bond to the person of a landlord, the most unfree peasant in the feudal world was bound to the soil rather than the person. This was called estate subjection. Yet, the difference between personal and soil bondage was subtle in most practical respects (Ogilvie, 2014). The estate subjection gave the seigneurs the opportunity to interfere in the tenants and their descendants choices concerning labour allocation, geographical mobility, marriage and inheritance. On the other extreme, there were peasants whose only node to a seigneur was a monetary rent to compensate the landlord for the lease of the soil.

In the course of time most german lands experienced the severe form of feudalism. Yet, areas differ in their length of exposure and the timing in the downfall of feudal restrictions. At the beginning of the twelfth century the vast majority of Western Europe was enslaved and in Western Germany this situation was static until the beginning of the thirteenth century. It was only the grand depression between the fourteenth and fifteenth century that slowly set them on the road to freedom. A steep drop in the European population, caused by plagues, crop failures and various wars, was followed by a secular decline in grain prices (Abel et al., 1966). Seigneurial revenues were deprived of their main contributors of income, in-kind obligations and monetary rents. Upcoming urbanisation exacerbated this tendency. In an effort to counterbalance these developments, landlords lightened peasants' burdens and advanced their emancipation to hold present - and attract new peasants (Carsten, 1947). This gradual process resulted in a rural society in Western Europa relieved from the most severe feudal obligations by the sixteenth century (Whaley, 2012).

At the onset of the twelfth century the eastern part of Germany just started to be colonized. Population pressure in the West paired with greater freedom and autonomy than in their native lands, led many Germans to follow the calls of the colonist seigneurs and cross the Elbe. The east-elbian settlers were granted lower obligations, favorable inheritance rights of the land (i.e. hereditary contracts) and communal self-governing unparalleled at that time in the West (Carsten, 1941). Yet, the economic contraction of fourteenth and fifteenth century flipped this situation. By 1427, unlike their western counterparts, the large landowners in the East started to restrict freedom of movement of peasants and agricultural workers and increasingly exploited obligatory labor services to ensure that the fields were tilled (Rosenberg (1944), p. 231-232). By the end of the sixteenth century serfdom was blossoming in the lands east of the Elbe.[4]

The end of the period of the feudal system in the Eastern land falls in the reign of the Prussian Empire. In 1807 the *October edict* de jure abolished feudalism, that is it gave peasants the *right* to redeem their feudal obligations. However as this law didn't specify even a corridor of compensation payments, peasants relied on the landlords goodwill, which did not exist (Melton, 2000). In contrast, the *Ablöseordnung* (Redemption Ordinance) of 1821 corrected this shortcoming and specified that tenants could free themselves from feudal ties upon compensating the landlord by 15-25 times the annual value of obligations. As only peasants with sufficient landholdings were eligible for redeeming their services, the amount was prohibitively high for many of them and didn't release them from all obligations, the de-facto date of the abolition of feudalism in Prussia is considered 1849 (Harnisch (1984), Bowman (1993)).

Feudal Restrictions

The historical literature usually calls the manorial regime in which the lord coerced his tenants into compulsory labor working on his demesne lands, *'demesne lordship'*. The estate owners of the demesne lordships coalesced jurisdictional authority, landlordship and sovereign powers, such as the collection of taxes. Frequently they were even called 'sovereign within his estate'. This coalition of institutional powers, enabled the demesne lords to widely regulate their tenants lives imposing forced labour, restrictions on spatial mobility, marriage controls, insecure property rights and their access to the judicial system.[5]

The owed labour services under the demesne lordship were at a minimum 2-3 days a

[4] This divergent reaction between the East and West is attributed primarily to four reasons: first, a rise in political power of the lordship; second, the amalgamation of jurisdictional and economic powers over the tenants on the demesnes; third, a shift of the lordship from receiving monetary rents to producing themselves for the markets and demanding labor services; four, the decline of the recently established and emerging cities as a political counterweight and outside option (Blum, 1957).

[5] See, (Jordan-Rozwadowski, 1900) or (Bloch, 2015) for a detailed account on the duties and obligations of medieval peasants.

week, which typically required an extra draught team, composed of servants and resident offspring to handle the forced labour for the overlord.[6] Often, seigneural demands exceeded this norm and could not effectively be resisted due to the lord's jurisdictional authority. For some areas, Knapp (1887) famously coined the term of the *'moonlight-farmer'*, whose dues were so numerous, that he had to till his fields during the night. While the remaining time could in theory be freely allocated, lords shielded the most productive sectors from entry and prevented the tenants to move temporarily to the most productive locations through dues and licence fees and denied permissions. Limitations on peasants' permanent freedom of movement were stipulated throughout the fifteenth century, first in customary practice and then written in sovereign edicts on the lordships behalf in the sixteenth century (Rosenberg, 1944). These ordinances prohibited the peasant and his family from leaving the estate directly through requiring the lord's permission and indirectly through obligatory servanthood of the peasant's children at the lords demesne and limitations on the peasant's choice of a spouse (Carsten, 1947). Another way of restricting peasants geographical mobility was to drive them into indebtedness, through increased obligations. Peasants then became debtors of their lords and refunded the credits through labor services (Blum, 1957).

Harnisch (1972) characterizes the constraints as binding and burdensome and enforced by 'naked force, indeed terrorization'. The penalties for non-compliance were as severe as corporal punishment, imprisonment and ejection from the holdings. Judicial ways to protest the lordships despotism were cut off, as the jurisdictional authority for the tenants was mostly the seigneur himself. Moreover, the scope for deviating through fleeing was limited through a lack of outside options, as landlord's formed cartellistic agreement between themselves and near-by cities (Rosenberg, 1943). The circumvention of the seigneurial restrictions was only possible though 'informal sectors', black markets, in which transactions were illegitimate, whereby peasants incurred enormous costs and risks and were left open to exploitation (Ogilvie, 2014). Belloc and Bowles (2013) argue that under these harsh conditions, feudalism and norms of subjugation were actually mutually best responses to each other and this dynamic complementarity between preferences and institutions allowed the inferior equilibrium to be sustained over time.

Conceptual Framework

The cultural learning model is the cornerstone to understand how feudal institutions pro-

[6] Additionally, demesne landlords demanded serfs to supply 'forced wage labour', at an arbitrary 'wage' chosen by the landlord.

duced historically a notion of trust confined to the nuclear family. This model denotes culture as a heuristic for decision-making in environments facing imperfect - or costly acquisition of information (Boyd and Richerson, 2005). Over time, the set of social norms that maximizes the pay-off (or the chances of survival) increases its prevalence through selection, adaptation and learning processes. Within this framework, the working hypothesis is that feudal restrictions increased the returns to rules-of-thumb based on mistrust relative to rules-of-thumb based on trust and led to the emergence of a culture of limited morality.

For one, we can rationalize the impact on feudal restrictions on generalized trust through the lens of the Tabellini (2008*b*) model. In the model the spatial pattern of external enforcement of cooperation determines the diffusion of generalized versus limited trust. A legal or informal enforcement of cooperation that is designed to foster relations between unrelated individuals fosters the dissemination of trust in strangers. In contrast, a clan-based organisation of societies spreads values of limited morality. Outside of the model, but inherent in the historical narrative is the fact that the mobility restrictions created de-facto socially closed entities. Emigration from feudal areas was curtailed and severely fought by landlords, immigration to feudal areas was the inferior option for any peasant (Ogilvie, 2014). Thereby, feudal constraints intensified within-group action but reduced the value of out-group cooperation. In other words, cooperation within the group was enforced *internally*, as deviations from the cooperation equilibrium became increasingly costly, the more closed the society was. The opposite holds, for rare encounters with members of the out-group however.

A second way, through which manorial constraints interfered with the evolution of norms of cooperation is by depriving peasants from market access, explicitly through extra levies and dues that had to be paid in order to leave the manor but more so implicitly through overburdening of obligations and labor services (Harnisch, 1986). The expansion of both the frequency and intensity of market exchange in turn, facilitates trust among strangers, as they share increasingly motivations and reputational concerns (Bowles (1998); Henrich et al. (2010)). As tenants in oppressed areas were largely cut off the experiences of goods markets and, as the occasional market participants in feudal societies, had no legal protection, these transactions were fraught with danger and exploitation. Thereby, the emergence of trust in transactions beyond the local group was thwarted.

Third, the feudal system superseded inter-community exchange for geographically diversified mutual insurance arrangements against weather shocks. Buggle and Durante (2017) show that demand for insurance against weather-induced crop failures historically co-evolved with measures of generalized trust. These mutual agreements between distant communities were redundant in areas with strong feudalist ties, as here the entire risk of a bad harvest was beared by the landlord, not the peasant community (Eddie, 2013). These regions operated mainly on a system of labor coercion, where the seigneurs demanded but certainly as well required the labor force of their tenants. In cases of crop failures, the main income risk was on the side of landlords, that even had to provide their tenants with food and equipment. In lieu thereof, areas with weak feudal ties were organized around fixed monetary rents between landlords and peasants (Jordan-Rozwadowski, 1900). Here, the risk of weather shocks was on the side of peasants, in whose interest were insurance contracts.

Last, one might wonder how mistrust persisted among the descendants of the populations in areas with stronger exposure to feudalism for more than 150 years. One possibility refers to a rich literature of direct *vertical* transmission of cultural values inside families. If trust values are somewhat hard-wired, as some studies suggest (Cesarini et al., 2008), genetic transmission may rationalize this. On the other hand, there is a rich literature showing that cultural values are passed down in families through vertical socialization (Boyd and Richerson, 1988). If parents exhibit *imperfect empathy*, that is a bias to shape the offspring's cultural values close to their own preferences, low values of generalized trust may have passed through generations. Alternatively, if the feudal constraints themselves have persisted to some degree, for example through a slow-moving culture within localized institutions, *horizontal* socialization may explain the spatial inertia of trust. A last option is rooted in the interaction of cultural values and local institutions or vertical with horizontal forms of socialization. If mistrust and local institutions are complements, descendants from feudal areas will not only have the tendency to mistrust but also choose local institutions that discourage cooperation and thereby perpetuate low levels of trust dynamically.

1.3 Data Sources and Description

In order to test in reduced-form the hypothesis that contemporaneous generalized trust is associated with the historical severity of feudal restrictions, I draw on two main sources of data: (i) the German Socio Economic Panel (GSOEP) and (ii) the Prussian Economic History Database.

1.3.1 Measuring Trust

The German SOEP is a representative panel survey of the resident adult population in Germany that was initiated in 1984 and covers today roughly 22,000 individuals, from about 12,000 households (for a detailed description, see Schupp and Wagner (2002)). For this study, I focus on the 2003, 2008 and 2013 waves as these include the key attitude questions used in our analysis. In the 2003 wave, the SOEP survey included three questions about individuals' trust attitudes, similar to standard measures of trust used in other surveys, e.g. the General Social Survey. Subjects were asked to rate to what extent they agree or disagree with the following statements: (1) In general, one can trust people. (2) These days you cannot rely on anybody else. (3) When dealing with strangers, it is better to be careful before you trust them. The answer categories were labelled: strongly agree; agree somewhat; disagree somewhat; strongly disagree. An advantage of the GSOEP is the availability of other individual and household characteristics, that may plausibly be connected to the outcomes of interest. In the main specifications, the set of individual and household controls includes, gender, age, logarithmic household income and religion as well as a full set of categorical indicators for education, occupation, working and marital status. Last, I include as well a dummy variable that separates individuals that were raised in the former GDR and made culturally diverse experiences than native west Germans (Alesina and Fuchs-Schündeln, 2007). From total pool of survey respondents, I exclude (i) individuals with non-German roots, that cannot have the trace of ancestral exposure to feudalism. Then, for the main specifications, I project the precise geolocation of respondents on the map of the Prussian Empire and (ii) drop individuals who live outside the ancient Prussian boundaries.[7] For this purpose, the borders of Prussia as of 1849 is chosen, which corresponds to the date of the de-facto abolition of feudalism.[8]

1.3.2 Measuring Feudalism

Second, I use the Prussian Economic History Database to measure feudalism in the period before its demise. According to historians, as of the late eighteenth century, serfdom across Prussia was being practiced at varying levels of intensity, depending on the customary obligations of the peasants and the strength of their land tenure rights (see, e.g. Bowman (1980),

[7] The shapefiles with the historical boundaries of the German Union are retrieved from https://censusmosaic.demog.berkeley.edu/data/historical-gis-files.

[8] Historical variables that are measured at different points in time are converted to the borders as of 1849. The changes in the administrative borders of the Prussian Empire were necessitated by population growth or the expansion of the boundaries through military successes.

Harnisch (1984), Pierenkemper and Tilly (2004)). The data consist of digitized censuses from Prussia and provide high-quality county-level information published on behalf of the Royal Prussian Statistical Bureau in Berlin (see Becker et al. (2014) for more details about the data and sources). Most notably, the data allow me to measure feudalism at the extensive and the intensive margin:

Concentration of Land Ownership

Following Cinnirella and Hornung (2016), the first measure for feudalism is the ratio of large landholdings over the number of total landholdings per county. Large landholdings are defined as farms with more than 300 Prussian Morgen (circa 75ha) and are the highest rank category in the Prussian censuses in 1816 and 1849. Harnisch (1984) reports that manors exceeding 300 PM were usually not cultivated by the landlord but rather by coerced and paid laborers. An alternative measure of serfdom is the size of knight estates, which is reported at a later period (1856). Knighthood came along with large properties and noble privileges of labor coercion. Both variables are indirect measures of feudalism that rely on the notion that a higher concentration of land estates creates a higher demand for paid or forced workers and a higher share of the population exposed to the constraints set by feudal lords.

Serf Emancipation

I use administrative data on serf emancipation, published by the Prussian Statistical Office, first presented in Meitzen (1866) and used by Ashraf et al. (2018). Meitzen (1866) reports on all resolved emancipation cases as of 1848, that is former serfs, who, upon settlement, were liberated from the provision of labor services to their manorial lords. For each county, Meitzen (1866) lists the number of redeemed days of service as well as the compensation payment for redemption of the services, which are very well suited to proxy the average intensity of coercion. Servile duties as well as redemption costs contain multiple elements, e.g. "draft animal services", "hand labor services", etc.

Feudalism Index

Hence, the historical records take stock of the extensive *and* intensive variation in the severity of de facto feudal restrictions. I combine the four variables describing the feudal system via a principal component analysis to construct a single measure that reduces the dimensionality of the data and captures the systematic structure of the variation of the various elements of the feudalist system. I will use in the subsequent analysis the first principal component of the four pieces, that retains 78% of the combined variation across counties.[9]

[9] The scores for the eigenvectors are 0.52 for landownership concentration, 0.49 for the share of knight estates,

Figure 1.1 and Figure 1.2 plot the spatial distribution of the landownership concentration and, respectively, redemption costs.[10] At first, it is noteworthy that both measures largely overlap with the historical accounts on the geography of feudalism in Germany. They confirm a steep gradient between north-eastern territories with respect to the areas of Rhineland and Westphalia in the south-west of the Prussian Empire. Second, the graphs suggest, that the river Elbe constitutes a marked border with counties in the east showing stronger feudal restrictions. Third, the fraction of Germany that was also part of the historical territory of Prussia comprise both regions abundant of feudalist constraints and areas where these restrictions were virtually absent from the 16th century. Last, the intensity of feudal oppression varied also between neighbouring counties. This local heterogeneity endows me with the possibility to restrict the analysis to variation across counties but within-provinces, the next higher administrative unit in Prussia.

1.3.3 Determinants of Feudalism

Before starting with the estimation of the main hypothesis, I want to perform an analysis on the determinants of feudalism. The historical literature on the causes of feudalism itself is rife and, from an econometric perspective, this suggests that we cannot treat the incidence of feudalism as random. This first analysis intends (i) to see which observable historical variables are predictive of feudalism and (ii) to determine to which *degree* districts with strong feudal institutions were systematically different from districts that abstained from interfering with their peasants' lives. A severe unbalancedness in terms of observables that measure the historical geography, development, demographics and culture would give rise to concerns about unobservables that are related to our observables. However, in total Table 1.1 does not pinpoint towards a strongly skewed distribution of feudalism across the observable covariates.

The first set of covariates aims to incorporate the geographic and climatic endowments of counties that can have a direct effect on both trust values and feudalism but can also interfere indirectly shaping proximate institutional characteristics through various mechanisms. The geographical control variables include latitude, longitude, the share of loamy soils, the share of swamps and an indicator for counties east of the river Elbe. While latitude and longitude proxy for manifold unobserved geographical and climatic factors varying across space, the share of loamy soils captures the texture of the soil, a measure strongly related to agricultural productivity. The east-elbian counties were historically the area that the ancestors of the German

0.49 for servile duties and 0.49 for redemption costs.

[10] For a plot of the feudalism index consult Figure A.1.3.

lands colonized and where agriculture and coercive labor institutions evolved diverse from the west of the Prussian empire (Harnisch (1986), Melton (2000)). Moreover, east of the river Elbe, serf-emancipation had been exacerbated due to the involvement and political power of the vested interests of the nobility to perpetuate their large-scale agricultural enterprises (Harnisch (1986), Melton (2000)). The inclusion of the indicator thus corrects also for measurement errors in the feudalism index itself. Last, the share of swamps is introduced to pick up variation in the disease environment of counties. The susceptibility for pathogens in turn can alter preferences directly or through a tendency to foster civil conflict. Those geographic covariates paint exactly the same picture as observed in Figures 1.1 and 1.2. The positive and significant terms of latitude, longitude and eastelbe refer to the apparent gradient in the intensity of feudal institutions from the south-western territories of Prussia to the north-east.

In the next column I simultaneously introduce indicators for the presence of a city and a main road as well as the urbanization rate in 1816. These factors can account for the historical level of economic development as well as the integration with goods and factor markets and thereby, at least partially, test the prominent Domar-hypothesis regarding the emergence of serfdom. Domar (1970) argues that areas with high-land labor ratios adopted the strongest feudalist regimes, which can be interpreted as the perceived threat of labor scarcity due to better outside options and therefore bargaining power for peasants against the elites (Acemoglu and Wolitzky, 2011). Alternatively, these variables capture access to goods access, that is related to norms of cooperation per se (Henrich et al., 2010).[11] In line with a literature on the effects of serfdom in Russia, I find, that areas strong in feudalist restrictions are also underdeveloped in terms of economic development, measured as urbanization.[12] On the other hand, the existence of cities or roads does never explain the severity of feudalism.

Alternative theories about the determinants of feudalism advance differences in inheritance regimes (Kaak (2013)), adoption in formerly slavonic areas (Carsten (1941)), protestantism (Ekelund, Hébert and Tollison (2002)) If those factors, that capture compositional differences across county populations additionally affect cultural practices and values or the propensity to attract migrant groups from diverse ethnic backgrounds, the main regression will be biased. Column 3 introduces an indicator that specifies whether the predominant form of inheritance of peasant landholdings in a county was primogeniture or partible. On the one hand, partible succession practices resulted in smaller landholdings, which is mechanically correlated to the

[11] Note, that also the share of swamps will contribute to control for population density.

[12] This pattern is insensitive of using alternative measures such as urbanization in 1849 or population density.

measures of emancipation through exclusion from the de jure process until 1850. On the other hand, one may argue that partible inheritance resulted in a scattered possession landscape, that complicated village consolidation for landlords and thereby lowered the extent of feudalism. Moreover, I include the share of Protestants in 1816 and the share of the population with non-Germanic (mostly Slavic) forefathers in 1861. Yet, none of these variables is associated to the index of feudal ties in any specification.

Then, I add distance to the province capital. This variable tries to proxy for areas where the sovereign was incapable of exercising it's power and thereby manorial lords set themselves up as the de facto power. These areas thereby have developed more despotic political institutions, that have been linked to limited morality (Tabellini, 2008*a*).[13] Furthermore, I include the per-capita number of steam engines in mining. This measure is predictive of industrialization and accounts for the fact that capital-skill complementarity has incentivized landlords to accelerate serf-emancipation as a way to free agricultural labor for their industrial purposes and thus explain differences in emancipation costs (Ashraf et al., 2018). Again, neither distance to province capital nor early-industrialization can add explanatory power in the spatial occurrence and intensity of feudalism.

Last, I perform the analysis only from within-variation of historical provinces in column 5. The sample includes 12 Prussian provinces that are on the nowadays German territory. This specification demonstrates that the differences in feudalism even within provinces seem to follow the general south-west to north-east gradient and were underdeveloped. Apart from that there doesn't seem to be a strong relationship between this sizeable set of historical determinants and feudalism. Still, this cautions against a causal interpretation of the OLS-results, through unobserved factors related to historical geography and development. One should however take note, that the R^2 of the models estimated from within-province variation achieves a value of 0.70, restricting the feature space for unobserved confounders. The main regressions will take into account this large list of possible confounders. In addition, I will use individual-level covariates that rule out persistence in trust values arising from the historical underdevelopment through occupational, educational and income-related mechanisms.

[13] Clearly, this is a coarse way of implementing the hypothesis, as through time the seat of the Prussian province capital may not have been most appropriate ruler.

1.4 Estimation Strategy and Empirical Results

1.4.1 OLS-Estimates

I begin by estimating the relationship between and individuals level of current trust and the historical incidence of feudalism with the following OLS regression:

$$Trust_{i,r,d,p} = \alpha + \beta Feudalism_{d,p} + \gamma \mathbf{X_{i,r,d,p}} + \delta \mathbf{G_{d,p}} + \theta \mathbf{H_{d,p}} + \Omega_p + \epsilon_{i,r,d,p} \tag{1.1}$$

where i indexes the individual, r denotes her residence, d refers to the historical district and p the historical province. The variable $Trust_{i,r,d,p}$ denotes one of the three measures of trust, which varies across individuals. $Feudalism_{d,p}$ represents one of the measures for the severity of feudal restrictions in the historical district d, located in province p. The values assigned to individuals are obtained based on the spatial overlap of the current residence r, i.e. her geo-coordinates, and the historical district d, i.e. $r \in d$. The coefficient of interest β estimates the reduced-form relationship between the historical exposure to feudalism and contemporaneous trust.

To limit doubts on the unconfoundedness of the OLS-estimates, I employ an extensive set of controls and fixed effects: $X_{i,r,d,p}$ constitutes the vector of socio-demographic controls. The intention to account for this set of individual controls arises from the desire to the measure the pure effect of *cultural* transmission of feudalism to modern trust. As (Buggle and Nafziger, 2018) show, Russian provinces that were strongly relying on serfdom exhibit contemporaneously underdevelopment. Expecting the same relation for feudalism, I would capture the effect of the intergenerational transmission of income on trust. $G_{d,p}$ is a vector of geographical controls for the historical districts, that comprises latitude, longitude, soil quality, the historic disease environment (share of swamps), and indicator for the presence of a city, the distance to the province capital plus an indicator whether the historical county was east of the river Elbe. $H_{d,p}$ reflects a vector of historical socio-demographic confounders, i.e. the share of urban population in 1816 and 1849 (proxying historical development), the share of steam engines and factories per capita (proxying susceptibility for industrialization), the share of the non-German population per capita, the share of the protestant population, an indicator whether the county had a *chaussee*, i.e. a main road (proxying for historical trade patterns) and a dummy variable indicating the inheritance system. I also include fixed effects for administrative units, indexed by Ω_p, which can be a modern federal state or a historical province. Upon inclusion the β-coefficient is identified by within-province variation and precludes that the results are driven by provinces without feudalism in 1816/1849 or cultural differences between provinces. Standard errors are clustered at the level at which the treatment occurs, i.e. the historical district, as suggested in Abadie et al. (2017).[14]

Results

The results from estimating Equation (1) are portrayed in Table 1.2. In total, results pinpoint towards a robust and negative long-term effect of feudal repressions on two measures of contemporaneous trust. Starting with Column (1), I find that historical feudalism impairs individuals generalized notion of whom to trust, conditional on household, geographic and historical controls. I introduce fixed effects for Prussian provinces, that eliminate concerns arising from time-invariant unobserved cultural, geographical and institutional features at the province-level, in Column (2). These factors slightly *increase* rather than dwarf the magnitude of feudalism. Finally, Column (3) tests the effect of feudalism, within contemporary state-borders. Importantly, German states have considerable degrees of legislative freedom in their level of legal enforcement, which might interact with the trust (Bigoni et al., 2016). If trust pays off more under stricter enforcement, the average individual will 'generally trust people' more, given the same social norms and preferences. Accounting for those differences doesn't alter the significance of the shadow of feudalism. In Columns (4) - (6), I introduce an alternative and probably more precise measure of mistrust. A conceptual concern with the statement to 'generally trust people' is the frame of 'people' that is activated in respondents. This reference group might refer to family, neighbours or unrelated others. Instead, 'distrust in foreigners' clearly puts the emphasis on unknown third parties. The coefficients now change the sign, as we measure *distrust* but apart from that are roughly of the same magnitude and statistical significance. This stability prevails, irrespective of restricting the variation to within historical provinces or nowadays state borders.[15]

Robustness Checks

The OLS-results are robust to a battery of sensitivity checks, suppressed here and relegated to the appendix for the sake of brevity. (1) Initially, I check for sensitivity with respect to the dependent variable. First, I use the remaining survey question 'trust in nobody' as the outcome

[14] Results are robust to clustering at the level of individuals.

[15] I cannot discuss the economic significance here, as I am lacking the mean values of the trust measures in the sample. I recall that the effects were quite small, in the range of 10-20% of a standard deviation.

of interest. Second, I include survey-fixed effects into the main regressions. These survey-wave dummies filter out general time trends in trust and hence capture the more stable part of trust. Results hold as well, when we use the neglected third question, on trusting nobody. Third, I exchange trust attitudes with self-reported trusting *behavior* as outcome. Survey responses are always subject to apprehensions of either non-truthfulness and/or only measuring intent, that is unrelated to real-world behavior. Therefore, I use the SOEP questions on actual behavior to overcome those worries. The questions were advanced in Glaeser et al. (2000), implemented and experimentally validated, through dictator games with SOEP respondents by Naef and Schupp (2009) and read as follows: (i) How often do you lend personal possessions to your friends?; (ii) How often do you lend money to your friends?; (iii) How often do you leave your door unlocked?
(2) Then, I perform robustness exercises on part of the regressors. Specifically, I replace the compound treatment of the index of feudalism, by it's ingredients. Additionally, I define define feudal origins based on the municipality of birth instead of the current location. Note, that this alters the sample of the respondents used for the analysis, which lends further credence to the hypothesis.
(3) Last, I alter the sample of respondents to demonstrate insensivity of the results with respect to the selection of respondents, most notably exluding the religion fixed effects from the regression, as up to 35% of the observations haven't answered this question. The results for these additional tests are displayed in Table A.1.1 - A.1.5. Taken as a whole, the coefficient sizes and statistical significance are remarkably stable to changes in the sample (the sample size varies within bounds of -36% to +33%), definitions of feudalism or measures of trust. This corroborates the credibility of the previous findings.

Irrespective of previous attempts to control for confounding factors, the OLS-method is subject to concerns of omitted variable bias. As the set of historical covariates and province fixed effects in the main regression explains already 70 % of the variation in the main independent variable (R^2, Table 1.1), these doubts of selection into feudalism based on unobservables are limited but certainly not eradicated.
One candidate are *time-invariant* omitted variables, for example geographical features not included in the regression. In particular, proximity to natural resources gave territorial lords sources of financing their budget unrelated to the landlords. Landlords contributed to the federal budget directly by money lending and indirectly by collecting taxes (von Below, 1923) and leveraged this position by acquiring rights of their peasants (Carsten, 1947). If contempo-

raneous trust is related to historical proximity to natural resources per se or through factors out of the researchers control, the OLS results are biased.

On the other hand, *time-varying* confounders threaten the measurement of the true relation. In the context of the so-called Brenner debate, it was argued that the autonomy of village communities was a decisive element of the severity of feudalism (Aston, Aston and Philpin, 1987). Brenner (1976) argued that bottom-up village institutions of economic and political self-government fostered cooperation and ultimately resistance against landlords in Western Germany, while lacking in Eastern Germany. One might imagine that this type of organization was conducive to inter-personal trust historically, but in contemporaneous impersonal markets actually breeds mistrust above village borders. If this is the case, historical village autonomy represents a time-varying unobserved variable.

1.4.2 Epidemiological Approach

In order account for time-invariant unobserved characteristics, I draw on what the cultural economics literature calls the epidemiological approach. The foundation of this approach is that (i) individuals who live in the same (administrative) area face similar incentives and constraints created by the cultural norms, geographical conditions and the economic and institutional environment. The epidemiological approach then (ii) estimates outcomes across different migrant groups residing in the same area. The gist is that individuals with a diverse cultural background will have different beliefs and take different actions despite sharing a common institutional and economic environment. Adopted to my setup, the epidemiological approach relies on variation in survey responses among movers in Germany, that reside in the same contemporary county, while their exposure to feudalism is at least co-determined by their county of *birth*. By constraining the variation to come from movers, time-invariant unobservables cannot contribute to the variation in outcomes, as they confound the current environment in the same way for all migrants. Thereby, this approach allows to isolate the portable element of culture from the stationary- and potentially confounded part of the original economic and institutional framework.

I estimate, the following regression

$$Trust_{i,o,d,p} = \alpha_r + \beta Feudalism_{d,p} + \gamma \mathbf{X_{i,o,d,p}} + \delta \mathbf{G_{d,p}} + \theta \mathbf{H_{d,p}} + \epsilon_{i,o,d,p}. \tag{1.2}$$

This estimation equation is different from the OLS-regression above in two respects. For one, the feudal ancestry $Feudalism_{d,p}$ is now defined by the **municipality of origin** o instead of the residence of individuals. In particular the individual gets assigned the value of the historical districts which overlaps with its municipality of birth, i.e. $o \in d$. The analysis is thus confined to individuals whose municipality of origin is different from their current residence. Second, the reduced form equation now entails a residence fixed effect α_r, and hence identification comes from individuals with different backgrounds in the same external environment. This fixed effect absorbs all time-invariant unobserved (e.g. geographical, institutional or cultural characteristics) and the culturally embodied effect of feudalism is separately identified. The residence can be a contemporaneous county ($N = 294$) or the next larger administrative unit 'regional planning districts' (*Raumordnungsregion*, $N = 96$). As the identification is exclusively based on individuals who don't live in their municipality of birth, differential selection can be an issue. If movers are on average more trusting, as one would expect, this would bias our estimates to finding a null-effect.

Results

The results of the epidemiological regression are reported in Table 1.3 and show mixed evidence in favor of the working hypothesis. Column (1) and Column (2) would suggest, that the OLS-coefficients were indeed plagued by omitted variables that shaped both historical institutions and contemporaneous trust. On the other hand, Columns (3) through (5) tend to confirm the notion of persistence in distrust through imposing immobility. While Column (3) is marginally insignificant, this can be attributed to a lack of statistical power. Remember, that the estimation is performed with the full set of county fixed effects in combination with a considerable number of controls, thus cell sizes tend to be small. When I exclude the frequently unanswered survey question on the religious denomination from the regression, the coefficient in Column (5) remains exactly the same in terms of magnitude, but the standard error decreases such that the relationship turns statistically significant again. This notion is supported when collapsing the entity of administration into regional planning districts.

One way to interpret these findings is simply to take them literal. Feudalism manifests in today's mistrust towards foreigners while the historical persistence of trustworthiness in people in general can be rather explained by unobserved characteristics. It is not hard to imagine a lack of historical shadow when generalized trust denotes trust towards the family, the kinship, neighbours or the village. My notion of feudalism has not anything to say on the relationship towards the kinship. In fact, the feudal constraints may have strengthened ties between the nuclear core of people on the lot of the estates. It was specifically the deprivation of contacts with unrelated strangers, that feudalism fostered.

1.5 Instrumental Variable Approach

However, the epidemiological approach is not suited to correct for the OLS-bias arising from time-varying unobservables, that may have been the true underlying drivers of both, feudal institutions and trust values. I use the combined variation of the share of sand and population losses in Thirty Years War (1618 - 1648) in an IV-LATE approach, to isolate an exogenous shift in the adoption of feudal institutions.

1.5.1 Explaining the IV Methodology

The general idea behind the IV is to use a historical determinant of feudalism in combination with a geographical pre-condition and exploit the quasi-randomness of the two combined. In this section I will explain how this shock influenced the feudal institutions and why we can think of it as exogenously and exclusively working through feudalism.

Share of sandy soils

The geological composition of the soil, called soil texture, is determined by the proportions of sand, silt and clay. Sandy particles are relatively round and offer more space between each other as compared to the slim and flat particles of silt and clay. These spaces translate into a lower capability of sandy soils to retain water. Differences in soil texture thus relate to the probability of draughts and crop failures.[16] Ex ante, it is difficult to build up an expectation whether the *singular* impact of a high content of sand, or respectively low agricultural productivity, should affect feudalism positive or negative.

On the one hand, homelands with low marginal value experienced historically a relatively lower demand for land. This lack of demand is accompanied by less accentuated land fragmentation and less secure property rights (Binswanger and Rosenzweig, 1986). Lower land fragmentation decreases the number of landlords in a village and drives down the competition between seigneurs up to the point where a landlord is the solitary ruler of a village and has certainly more scope to raise obligations then, say a scenario, where there is a second entity in the same reach (Brenner, 1976). Inferior property rights ensured higher obligations directly as issuing a new contract was used to generate revenues by landlords and indirectly because

[16]See Carsten (1947) (p.159) on the relevance in the German context: 'At the peak of the colonization, villages had probably been founded on soil which in the long run did not produce sufficient yields, or they were ill situated, endangered by floods, dearth of water, or soil erosion'. Thus, exactly the corollaries of a high share of soil.

the frequency to raise obligations increased.[17] Both, strengthen the bargaining position of the landlord vis-a-vis peasants and eased the tying of tenants to the soil.
On the other hand, as Eddie (2013) elaborates, demesne farming constitutes of a bundling of risk on the side of the seigneurs. Feudal lords, whose mode of farming constituted of coerced labour bore the entire risk of crop failures. On the other hand, manorial contracts that relied entirely on fixed monetary rents, shielded the landlords income from arising draughts and missing harvests. Thus, a higher content of soil could, gives landlords incentives to switch to a system of monetary rents and eliminate income risk with respect to crop failures.

Population Losses in the Thirty Years War

It is recognized, that the period after the *Thirty Years War* (1618 - 1648) was characterized by a general re-feudalization of the German society (Schilling (1989), Whaley (2012)). It is also known, that land-labor ratios rose disruptively during the *Thirty Years War* as the Holy Roman Empire lost over a third of its inhabitants (Voigtländer and Voth, 2012*b*). The historical literature hypothesized for a long time that increases in land-labor ratios are associated strongly with labor coercion. Domar (1970) argues that in economies where wages were high because labor was scarce relative to land, landowners devised institutions such as serfdom and slavery to ensure they could get labor to work their land at a lower cost than would be the case in a non-coerced labor market. Depopulation thus led landlords to impose peasant clearance, mobility restrictions, loss of hereditary status of the land and ultimately higher obligations (Brenner (1976); Aston, Aston and Philpin (1987)).[18] Thus, regions with higher population losses in the Thirty Years War (1618-1648) exhibited jumps in land-labor ratios and tightened feudal restrictions.

Interaction Term

The core of this IV-methodology is the use of the combined variation of the share of sand and population losses. Thereby, the interaction-term tests whether the relationship between feudalism and population losses was different on sandy soils with respect to population losses on soil with more clay contents. I argue, that the jump in land-labor ratios arising from the Thirty-Years-War induced a relative *de-feudalization* in areas with a high share of sand. The key difference between this depopulation shock in the early 1600s and other periods was that it interacted with changes in the entrepreneurial choices of the nobility.

[17] For example from a hereditary contract, where obligations were fixed even through generations to an non-hereditary contract, that allowed an adjustment at least when handing the land to a successor Ogilvie (2014).

[18] Recently, Klein and Ogilvie (2017) proposed that this relation holds causally.

To rationalize this choice, two changes in the occupational environment were fundamental. First, 'the long-run depression from 1618 to 1650 [...] terminated the era of profitable agricultural expansion and checked the Junker's entrepreneurial career, which was not to be resumed on a grand scale before the middle of eighteenth century' (Rosenberg (1944), p.236f.). This meant, that the market equilibrium in which the landlords operated shifted to low prices for agricultural commodities and land, rising real wages and high production costs (Rosenberg (1944), p. 239). Second, the aftermath of the Thirty Year War was coined by the rise of *Absolutism*. As documented in the "Recess" of 1653, the Great Elector of the Holy Roman Empire, Frederick William, announced his standing army financed by an annual military tax. This created lucrative non-agricultural trajectories as professional army officers, bureaucratic administrators, diplomats and courtiers in the kingdoms (Rosenberg, 1943). In fact, the territorial princes granted the feudal lords a monopoly on those posts (Carsten, 1947).

Consider a revenue maximizing landlord, who faced the depletion of the labor supply and the widespread desertion of the land due to the Thirty Years War. Now, the argument is that that landlords in areas with relatively higher soil productivity decided to cope with the market environment and found it necessary to impose mobility restrictions on their peasants relatively more often. Alternatively, landlords on sandy soil were rather incentivized to shift their agricultural enterprises from the active managing of large demesnes to passive receivers of monetary grants and pursuing a trajectory entering lucrative state services. Thus, seigneurs possessing more sandy soil relatively more so lightened peasants' burdens and advanced their emancipation, in order to induce their tenants to stay.

Figure 1.3 plots the graphical representation of this interaction term.[19] Indeed, we find that the North-Eastern areas, where feudalism was thriving, heavy wartime losses met fertile land. In contrast, in the regions of the Rhineland, that were largely spared of the turmoil of the Thirty Years War, we do find the milder versions of feudalism (compare 1.2 and 1.3).

Exogeneity: Random Assignment

Exogeneity requires that conditional on the set of covariates the probability of being *assigned* to treatment, i.e. having higher population losses paired with higher contents of sandy soil, is as good as random. Clearly, wartime losses during the Thirty Years War could reflect underlying differences across areas and be endogenous to economic preferences or development. Historians noted that population losses were concentrated for example near main roads and along a corridor from the North-East to the Middle-West of Germany (Franz (1979); Whaley

[19] See Figure A.1.2 and Figure A.1.3 for plots of the individual components.

(2012)). However, the exogeneity assumption refers to the interaction term. On this regard, soil texture ensures the identification of exogenous variation. Soil texture is a natural feature, constant over time and out of the scope of human intervention. Thus, the interaction term can be regarded as distributed randomly.

Exclusion Restriction: Feudalism as the Channel

To fulfil the exclusion restriction the population coefficient of stronger population losses interacted with sandy soil can affect individual trust only through its effect on feudalism.

Soil texture is clearly a strong determinant of crop choice. Crop choice in turn has been related to time preferences (Galor and Özak, 2016) and may bias the results if feudal and non-feudal areas produce systematically different field crops that require different amounts of cooperation and hence are beneficial or detrimental in building-up trust relations between individuals or communities. Moreover, wars have been shown to deteriorate trust. This may occur directly through a reinforcing process between conflicts and ethnic cleavages (Rohner, Thoenig and Zilibotti, 2013*a*) or indirectly harming future inter-ethnic trade, that in turn erodes trust building (Rohner, Thoenig and Zilibotti, 2013*b*). One could alleviate these concerns through the inclusion of (available) historical information on per hectare yields for the most common grains of that time (wheat, rye, oat, barley, and potatoes) or arguing that ethnic markers hardly played a pivotal role in ancient Germany.

However, considering the components in isolation is not enough to possibly violate the exclusion restriction. Instead, only factors that are related to **both** variation of population losses *and* the soil texture of counties run against feudalism as the single causal channel of influence. I am not aware of any historical narrative or current research that demonstrate the effects of soil texture in combination with depopulation shocks on outcomes potentially related to distrust in strangers.

1.5.2 Instrumental Variable Estimates

For the instrumental variable design, I estimate the following first stage regression:

$$Feud_{d,p} = \alpha + \beta_0 sand_{d,p} + \beta_1 pop - losses_{d,p} + \beta_2 Z_{d,p} + \delta \mathbf{G_{d,p}} + \theta \mathbf{H_{d,p}} + \Omega_p + \epsilon_{d,p} \tag{1.3}$$

The extent of feudalism is regressed on the control variables from the baseline specifications plus $sand_{d,p}$ the share of sandy soils in county d, $pop-losses_{d,p}$ the population losses in percent that arose from the Thirty Years War in county d and $Z_{d,p}$ that reflects an interaction-term between the two.[20] This instrumental variable strategy requires thus additional information on soil texture and population losses in the Thirty Years War (1618-1648). The data for the share of sandy soils were collected by an 1866 census and reported in (Meitzen, 1866), which assessed the composition of the soil in a county by gathering information on three main soil categories: the area of clay soils, of loamy soils and sandy soils. The variable $sand_{d,p}$ is defined as the area of sandy soils as a percentage of the total land. To gauge the population losses from the Thirty Years War, I have digitized historical maps from various sources i.a. de La Blache (1894), Franz (1979) and Vasold (1993). From those maps, I created the spatial overlap with the Prussian counties and averaged across the historical maps to get a diverse and complete picture of the losses.

Table 1.4 displays the results for the first stage of the IV approach and builds up the intuition sequentially. Column (1) demonstrates a strongly negative relationship between the share of sand in a county and the adopted feudal institutions.[21] This occurs even if I include latitude and longitude, that capture the impact of unobserved geographic and climatic factors, for example temperature or precipitation. Thus the share of sand had a unique influence on feudalism net of the geography of historical areas. Column (2) introduces the population losses from the Thirty Years War and confirms that this particular shock in the land-labor ratio reinforced feudalization. Both coefficients retain sign and significance when included simultaneously in Column (3). Column (4) then accounts for the differential effect of population losses depending on soil texture. The interaction term enters negative and significantly. Moreover, the F-Statistic transcends the critical threshold suggested in (Stock and Yogo, 2002).[22] Last, I estimate the same relationship from within-county variation. Again, the share of soil and population losses are strong predictors of the adoption of feudal regimes. However, we can now not longer exclude the possibility to run inference problems arising from weak-instruments.

The interpretation of the obtained parameter is the effect of feudalism for those counties

[20]The $sand_{d,p}$ has a mean of 0.2 and a standard deviation of 0.2. $pop-losses_{d,p}$ has a mean of 30 and a standard deviation of 17. The results are qualitatively robust to an alternative definition based on the relative dominance of soil texture.

[21]This result is opposite of the one found in Cinnirella and Hornung (2016). This may generally be the case due to a different sample of counties - Cinnirella and Hornung (2016) use the entire Prussia, dependent variable or control variables. Yet, if I align my specification to theirs, i.e. using the same dependent, independent and control variables for the sample of nowadays Germany, this relation persists to be weakly negative. To explain the difference, note that Cinnirella and Hornung (2016) motivate their positive effect through historically diverse land demand and farm sizes. As elaborated by Carsten (1947), farm sizes seemed to have evolved uniformly in the German lands, which shuts down this channel.

[22]Note, that the *coefficient* of sandy soil has changed sign. Yet, the effect of sandy soil can only be interpreted meaningfully in combination with the interaction term: $\frac{\delta y}{\delta share-sand} = 1.08 - 0.13 \times pop-losses$. In words, this states that the effect of sandy soil of feudalism was positive for counties that experienced population losses below 9% of their population. Less than 10% of the counties in the sample experienced population losses of this magnitude.

whose treatment status can be changed by the instrument. In other words, identification comes from counties that adopted a new feudal regime in response to the combination of soil texture and a depopulation-shock, but who would not have developed strong feudal institutions otherwise. This excludes for example regions who were never part of the manorial system such as cities as well as regions whose feudal institutions were fully developed before the Thirty Years War. Hence, the estimand captures the local average treatment effect (LATE).

Results

[Incomplete]

1.6 Conclusion

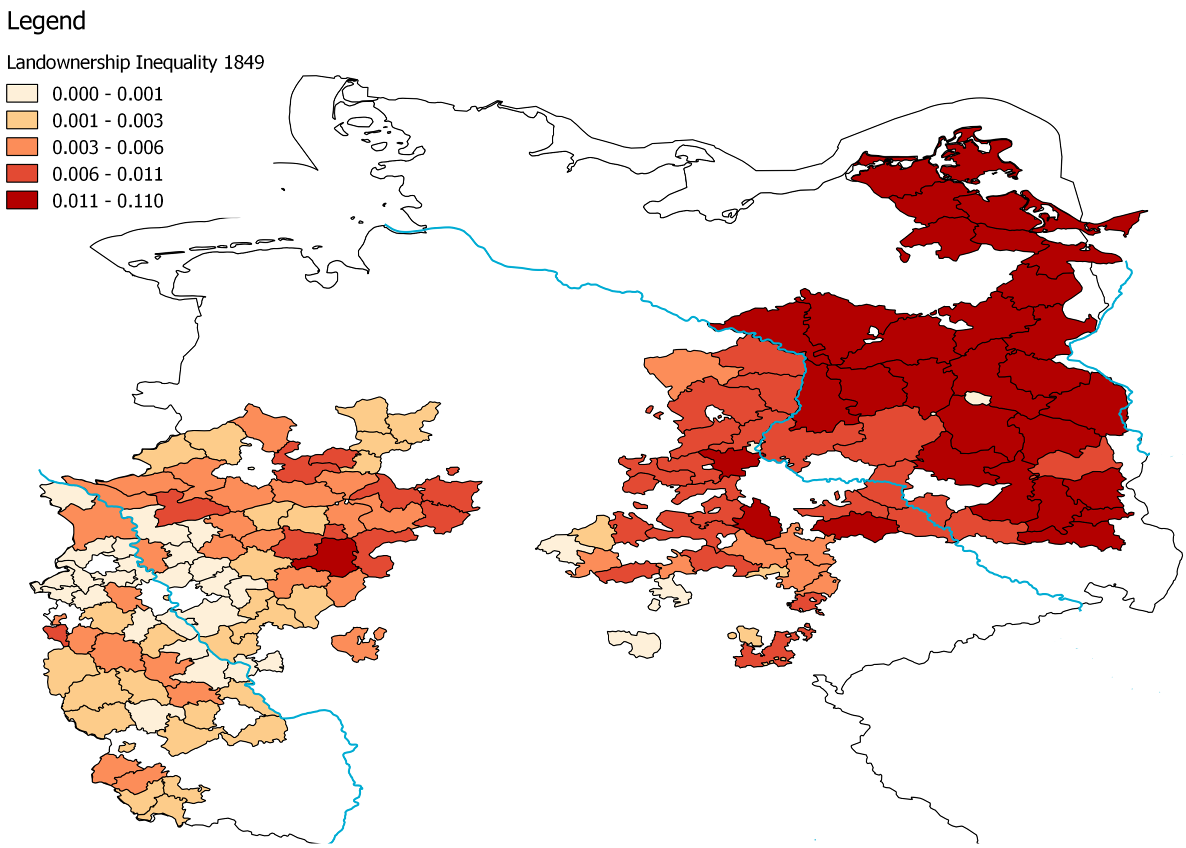
Perpetuating differences in mistrust have the potential to cause severe social and economic ramifications at individual as well as aggregate levels. This paper argues that for centuries the pivotal medieval institution of feudalism influenced the geographical pattern of generalized trust in Germany and still does so. Feudalism was accompanied by severe migration restrictions and coercive labor obligations that purportedly breeded amoral familism through lowering the access to goods markets and social closure.

I test this hypothesis within the borders of modern Germany by constructing a measure of ancestral exposure to feudalist constraints. I document empirical patterns that confirm my conjecture. The findings indicate a negative and significant relationship between contemporary generalized trust and historical feudalism. I document the robustness of this cross-sectional relationship accounting for a wide range of potentially confounding factors including household characteristics as well as geographical, historical and cultural characteristics of Prussian counties. I further show that the association remains qualitatively and quantitatively robust to different levels of province fixed effects, that wash out time-invariant unobserved heterogeneity at the level of historical provinces and contemporaneous federal states.

Complementing this analysis by applying the epidemiological approach shields the estimates from persistence in unobserved time-invariant confounders at the even narrower level of contemporary counties. I use internal migrants that carry the trace of feudalism through their municipality of birth to estimate their trust keeping their current institutional, cultural and geographic environment constant. This alternative identification strategy demonstrates that the long shadow of feudalism persists specifically in distrust against strangers rather than a general notion of mistrust.

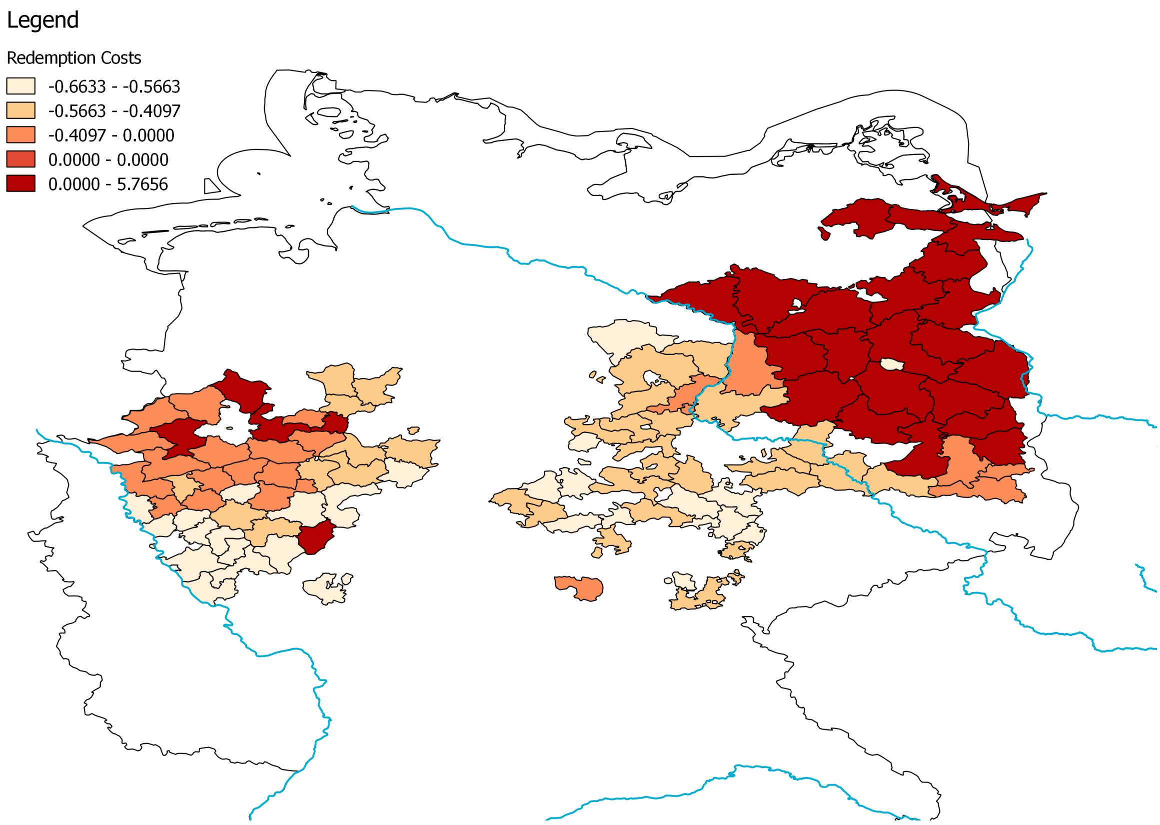
Conclusive evidence is sought via an instrumental variables methodology. The proposed instrument combines exogenous geographical variation with a historical determinant of feudalism to circumvent the outstanding endogeneity concerns. In specific, I predict the extent of feudalism in a Prussian county through the texture of its soil and the population losses in the Thirty Years War. While desertion of the land provided historically a stimulus towards demesne farming and mobility restrictions, the peculiar agricultural and occupational market environment after the Thirty Years War gave differential incentives to landlords depending on their soil texture. The first stage results confirm this narrative and are robust at conventional levels.

Figure 1.1: Landownership Concentration 1849



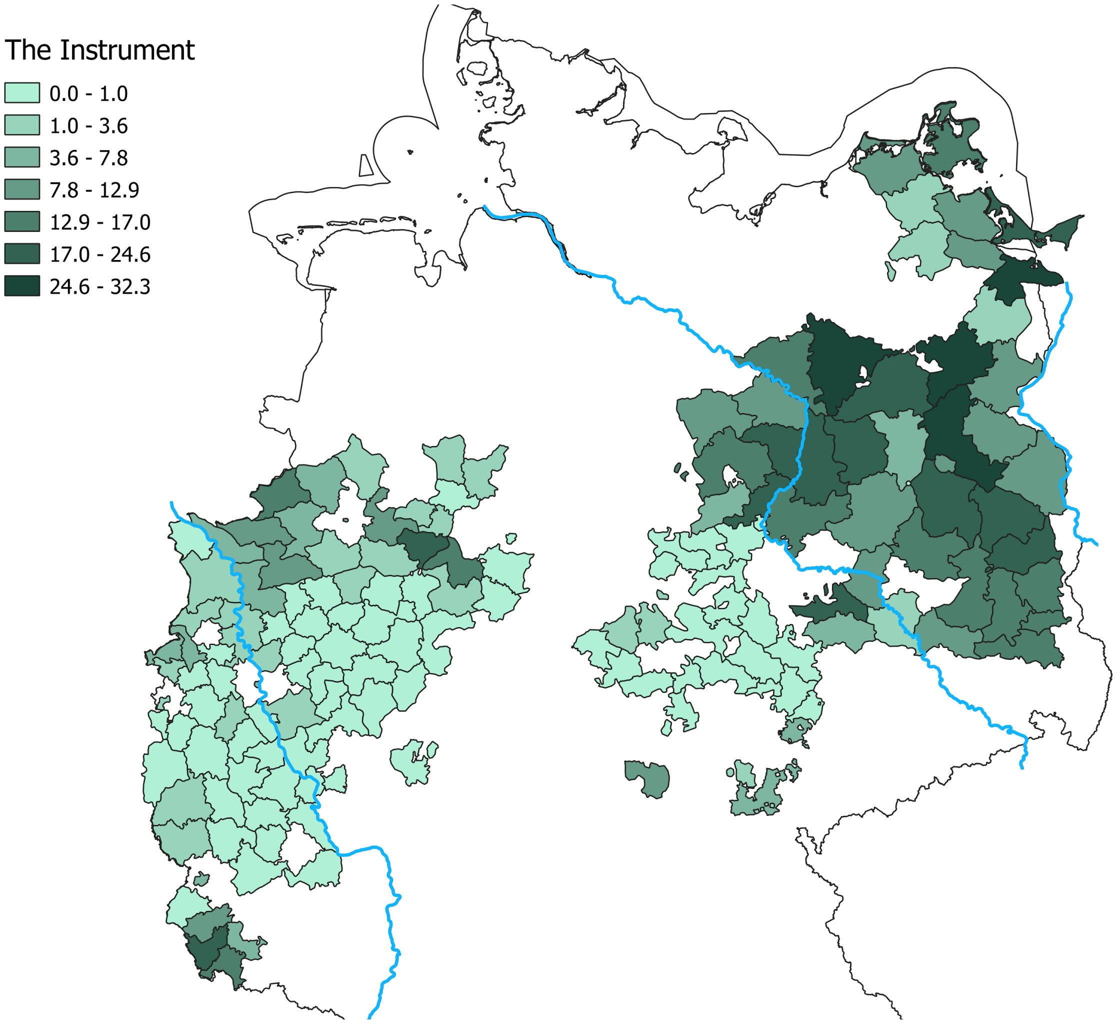
Notes. - Figure plots the spatial distribution of landownership concentration in Prussia as of 1849. The indicator is bounded between 0 and 1, with higher values representing a higher share of farms that own more than 300 Prussian Morgen.

Figure 1.2: Redemption Costs 1821 - 1849



Notes. - Figure plots the spatial distribution of redemption costs in Prussia between 1821 - 1849. The indicator is the first principle component of several redemption payments, with higher values representing a higher costs.

Figure 1.3: Instrument: Share of Sandy Soil x Population Losses



Notes. - Figure plots the spatial distribution of the interaction term between the share of sandy soil in a county, bounded between 0 and 1, combined with the population losses in the Thirty Years War (1618 - 1648), bounded between 0 and 100 percent.

Table 1.1: Determinants of Feudalism

	Feudalism Index				
	(1)	(2)	(3)	(4)	(5)
Latitude	1.080**	1.139**	1.122**	1.233**	1.533**
	(0.405)	(0.393)	(0.417)	(0.455)	(0.453)
Longitude	0.111**	0.173***	0.206**	0.194*	0.591*
	(0.0407)	(0.0495)	(0.0715)	(0.0874)	(0.301)
Soil Texture	0.144	0.220	0.148	0.0153	-0.943
	(0.997)	(0.933)	(1.013)	(0.965)	(0.833)
Eastelbe	1.464***	1.308**	1.305**	1.354**	-0.621
	(0.414)	(0.423)	(0.459)	(0.444)	(0.649)
Share Swamps	-2.513	-3.401	-3.430	-3.278	-4.542*
	(2.610)	(2.716)	(2.757)	(2.791)	(1.940)
City		0.256	0.288	0.293	0.158
		(0.226)	(0.240)	(0.245)	(0.277)
Urbanization 1816		-1.977**	-1.995**	-1.931**	-2.055**
		(0.626)	(0.647)	(0.609)	(0.767)
Chaussee		-0.231	-0.224	-0.254	-0.298
		(0.189)	(0.188)	(0.204)	(0.306)
Inheritance			-0.0130	-0.0180	-0.393
			(0.245)	(0.254)	(0.370)
Share No-Germans			-0.297	-0.607	-0.764
			(1.192)	(1.156)	(1.746)
Share Protestants			-0.236	-0.303	-0.223
			(0.282)	(0.320)	(0.491)
Distance Province Capital				3.924	6.060
				(3.797)	(5.385)
Steam Engines Mining per Capita				1.099	0.771
				(0.643)	(1.664)
Observations	116	116	116	116	116
Fixed Effects	No	No	No	No	Province
R^2	0.570	0.608	0.609	0.614	0.703

Note: The table reports OLS estimates. The unit of observation is a historical county. Standard errors are robust to heteroscedasticity. List of variables includes latitude, longitude, soil quality, % swamps, an indicator for counties east of river Elbe, having a chaussee, a city or the common inheritance system, per-capita counts of non-germans, protestants, steam engines in mining as well as measures for urbanization in 1816 and the distance to the province capital. Historical and geographical controls refer to 1849.

*p<0.1; **p<0.05; ***p<0.01

Table 1.2: OLS Estimates

	Dependent variable:					
	Generalized Trust			Distrust in Foreigners		
	(1)	(2)	(3)	(4)	(5)	(6)
Feudalism	-0.029***	-0.036***	-0.026***	0.027***	0.022**	0.022**
	(0.007)	(0.007)	(0.007)	(0.010)	(0.009)	(0.008)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	No	Province	State	No	Province	State
Observations	11,966	11,966	11,964	11,969	11,969	11,967
R^2	0.054	0.059	0.057	0.058	0.067	0.063
Clusters	113	113	113	114	114	114

Notes. - Note: The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. Household controls include gender, age, log household income, a dummy for former GDR and a full set of dummies for education, occupation, religion, employment and family status. Geographic controls include latitude, longitude, soil quality, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.
*p<0.1; **p<0.05; ***p<0.01

Table 1.3: Epidemiological Approach

	Dependent variable:				
	Generalized Trust		Distrust in Foreigners		
	(1)	(2)	(3)	(4)	(5)
Feudalism	-0.0062	-0.0069	0.018	0.020***	0.018**
	(0.011)	(0.007)	(0.011)	(0.006)	(0.008)
Household Controls	Yes	Yes	Yes	Yes	Y, w/o Religion
Geographic Controls	Yes	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes	Yes
Host Fixed Effects	County	Province	County	Province	County
Observations	7,474	9,511	7,478	9,516	9,516
Clusters	113	113	113	113	113

Notes. - The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. The controls mimic the ones in the main specification.

Table 1.4: First Stage

	Feudalism Index				
	(1)	(2)	(3)	(4)	(5)
Share Sandy Soil	-2.373*** (0.881)		-2.029** (0.779)	1.086 (1.000)	3.063*** (1.098)
Pop Losses Thirty Years War		0.0410*** (0.0142)	0.0345*** (0.0123)	0.0615*** (0.0185)	0.0685*** (0.0135)
Interaction				-0.127** (0.0567)	-0.174*** (0.0329)
Observations	116	116	116	116	116
Geographical Controls	Yes	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes	Yes
Fixed Effects	No	No	No	No	Province
R^2	0.656	0.652	0.692	0.737	0.792
F-Stat	10,56	10,16	10,17	10,52	7,21

Note: The table reports OLS estimates. The unit of observation is a historical county. Standard errors are clustered at the level of the historical county. Geographic controls include latitude, longitude, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.

*p<0.1; **p<0.05; ***p<0.01

Chapter 2

Revolution in 140 Characters? - U.S. Politicians on Social Media[1]

Abstract

In this paper, I study the behavior of U.S. politicians on social media. I first build a novel dataset that includes the Twitter handles of 18,000+ politicians and 61+ million tweets from 2008 - 2021. The data span from the local level to the president and include both, legislators and candidates. The analysis of this novel dataset delivers several new insights. First, I find substantial partisan differences in Twitter adoption, Twitter activity and audience engagement across all levels of government. Second, I use established tools to measure *ideological* polarization on text to provide evidence that online-polarization follows similar trends to offline-polarization, at comparable magnitude and reaches unprecedented heights in 2018 and 2021. Third, I apply sentiment-analysis to tweets about the political in- vs out-group to show a marked increase in *affective* polarization across the sample period. Fourth, I show that sudden jumps in audience engagement reinforce political attacks on opponents.

[1] I am much obliged to Paolo Masella and Alessandro Saia for comments and advice.

2.1 Introduction

In 2020, nearly 3 out of 4 Americans used social media for news consumption, as opposed to little more than 10 percent in 2008.[2] At the same time, the use of social media among politicians as a tool to communicate with voters is becoming more and more widespread. Figure A.2.1 shows that over the same time horizon, the share of political candidates that have a Twitter account grew from 2% to about 56%. In 2021, well over 90% of the members of congress use Twitter and the median legislator reaches out to voters on average 3 times per day.

Likely this triumph of social media will have multifaceted consequences on the political arena. While the low entry barriers of social media could allow extremist candidates, previously sidelined from the political and media establishment, to garner widespread attention (Levitsky and Ziblatt (2018), Zhuravskaya, Petrova and Enikolopov (2020)), the same holds true for groups that are disadvantaged in politics, for example newcomers or minorities (Petrova, Sen and Yildirim, 2021). Direct interaction with politicians enables voters to express (dis-) content with policy actions or - proposals and could thereby increase responsiveness among politicians. On the other hand, divisive messages receive more audience engagement on social media (Rathje, Van Bavel and van der Linden, 2021), which could tempt politicians to polarize more and intensify political conflict. A key challenge to answer these questions is the lack of comprehensive data on politicians' social media presence and activity.

In this paper, I present a novel dataset that includes Twitter handles of 17,000+ U.S. politicians, from the president to the mayor, and 61+ million tweets from 2008 - 2021. I first document a set of stylized facts about the evolution of Twitter adoption, Twitter activity and audience engagement, disaggregated by layer of politics and political party. Second, I rely on established machine-learning tools to characterize the level and evolution of *ideological* polarization on social media. The social-media time-series exhibits similar trends at slightly higher magnitudes compared to congressional speeches, as an offline benchmark, and climbs to unprecedented heights in 2021. Third, I develop and apply a new tool for the quantification of *affective* polarization in political speech and observe a steady and unbroken increase in political conflict along the *emotional* dimension. Fourth, I test whether politicians' communication reacts to sudden jumps in audience engagement and whether different types of user responses (positive and negative) amplify or dampens future communication.

Social media data provide some crucial advantages over more traditional sources, such as

[2]The 2008 value is taken from Pew Research Center 2008 Biennial Media Consumption Survey. The 2020 value is taken from the Pew Research Center American Trends Panel Wave 73, August 2020.

roll-call votes or congressional speeches, to study political speech or measure polarization. The first advantage is that social media data come at much higher temporal granularity and eases also the collection of large-scale data on politicians. Due to the volume of text generated, I can identify shifts in polarization at a higher time-frequencies as the common two-year horizon of a congress (as in Gentzkow, Shapiro and Taddy (2019)). Due to the wide spectrum the wide spectrum of politicians observed (from mayor to president; candidates and incumbents) I can construct a more complete picture of the extent of polarization, including the state - or local level and also to characterize the geography of elite polarization.

The second advantage concerns identification. Both, roll-call votes and congressional speeches, are subject to substantial party influence and thus not pure manifestations of a politicians' ideology.[3] Party leadership exerts control not only on the legislative agenda, so which *issues* are discussed/voted, but also determines which politicians' speak about which topic. Tightening party-discipline has been demonstrated to account for large proportions in the rise of polarization in roll-call votes.[4] Social media data alleviate, at least partly, these concerns as party leadership has arguably less *direct* control over the topics that arise on social media and who expresses her opinion towards these topics.[5]

I rely on data from Vote Smart that covers the universe of politicians at the federal - and state level. Since Twitter gained wide publicity in 2008, I observe 75,000+ politicians. Combining information on the social media presences from Vote Smart, Ballotpedia and own collection efforts, I am able to match 17,2000+ politicians to Twitter handles (around 22%). In sum, these politicians have sent 61+ million tweets and combine 737+ million followers. In 2021, only the members of Congress sent 63,000+ tweets in a typical month and, on average, their tweets are retweeted more than 200 times and received 1100+ likes. Politicians active on social media are more likely to be from groups under-represented in politics (women, non-white), have less legislative experience and more likely identify as Democrats.

As a first step, I use this novel dataset to provide stylized facts about politicians' behavior on social media based on *(i)* the adoption of Twitter, *(ii)* tweeting behavior as well as *(iii)* the feedback received from the audience: *(i) Twitter adoption* varies widely over space and time. I

[3]This argument is made both political economy literature (e.g., Moskowitz, Rogowski and Snyder (2018) or Canen, Kendall and Trebbi (2020)) and the political science literature (e.g., Proksch and Slapin (2012)).

[4]Canen, Kendall and Trebbi (2021) find that party pressure accounts for around 65% of polarization in recent decades; Clinton, Katznelson and Lapinski (2016) present patterns of "polarization" simulated from changes in the composition of the agenda holding constant ideological distances; Lee (2008) attribute more than a third of polarization in the Senate to changes in the agenda.

[5]Depending on which extent of party control one assumes for congressional speech relative to social media, the observed differences in magnitude might thus vary substantially.

find that the share of politicians that have a Twitter account ranges from around 25% in Maine, Wyoming and North Dakota to more than 70% in New York, Arizona and California. A first wave of widespread Twitter adoption, with 1000+ accounts opened took place in the beginning of 2009. Ever since Twitter adoption peaks in the start of federal election years. I find significant differences in Twitter adoption across Democrats and Republicans from 2015, with Democrats opening substantially more accounts.

I observe a similar partisan-gap with similar timing also in *(ii) tweeting behavior*. While, the median member of congress for both parties steadily increases her tweets per months from 10 in 2010 to 50 in 2015, Democratic politicians remain on this steady path and reach 120 tweets in 2021, whereas the curve for Republican politicians stagnates at 50 tweets until 2019 but trends upwards to reach 75 tweets per month in 2021.

Likewise, *(iii) audience engagement* evolves disproportionately over time. Starting in 2015 members of the Democratic party receive more likes and retweets than Republicans. The average tweet for the median legislator of the Democratic (Republican) party gets around 70 (40) likes and around 18 (16) retweets. In contrast, tweets from the Republican get sufficiently more *negative* audience engagement, starting from 2017. This is also reflected in the time-series of Twitter followers: The median Democratic account increased its followers from 16,000 Twitter users in the beginning of 2017 to 38,000 in the end of 2021. Instead, followership rose for the median Republican account only from 14,000 to 21,000.

By means of rich fixed-effects regressions, I show that Twitter adoption and activity are associated with a substantial and significant increase in the probability of winning an election. Using within-candidate changes, I estimate that the likelihood of being an electoral winner in a general election improves by about 4 percentage points (or about 8% of the average probability of winning) after opening a Twitter account.

The second part of the paper, quantifies trends in affective polarization among political elites. As in Iyengar et al. (2019), affective polarization is defined as the extent to which politicians express more positive sentiment toward their own party than toward the other party. To measure affective polarization, I initially identify tweets about politicians/political parties using *(i)* tags of twitter handles and *(ii)* mentions of names/ideology. I then determine the affect carried in those tweets, using VADER, a state-of-the-art sentiment-dictionary, attuned to social media. Among tweets from the members of congress, I find 970,000+ tweets that speak about one of the parties, and calculate a monthly measure of affective polarization.

Five findings emerge. First, affective polarization among members of congress rose steadily between 2010 and 2021. This wedge is driven by both, more negative out-party sentiment and more positive in-party sentiment. The increase in affective polarization is more pronounced but not limited to election years. While affective polarization between politicians grew by 64% in the last ten years, its level in 2020 was well below commonly estimated values for voters. Second, the rise in affective polarization is significantly stronger when politicians speak *about* each other, that is mention others names/ideology. but approximately flat when speaking *with* each other, that is when they tag other twitter handles. Third, this time-trend in partisan affection is present for both parties, albeit with differences in the timing. Fourth, I observe a strong increase in affective polarization for congressional candidates until 2016, with a decline thereafter. The level of affective polarization among state politicians is about 35% lower than for members of congress, but also trending continuously upwards. Fifth, across Congress members affective polarization is higher for members with ideologically more extreme roll-call voting records, for Democrats, for women, for the opposition party and for less-educated legislators.

In the third part, I focus on ideological polarization. I adopt the penalized regression method for high dimensional data developed by Gentzkow, Shapiro and Taddy (2019). This method explicitly addresses finite-sample bias, that is the tendency of naive estimators to interpret words that are said mostly by one group by chance as polarization.[6] I apply the penalized regression to estimate polarization on the tweets from members of congress. I construct two time-series: One from actual data and one from hypothetical data, that randomize party membership, which allows me to quantify the amount of finite-sample bias remaining. The permutation test suggests that finite-sample bias is low.

Ideological polarization rises from 2009 to 2013, weakly decreases for two years but 2016 the partisan divide among members of congress trends again upwards and reaches unprecedented heights in 2021. The observed trends are largely consistent with trends in congressional speeches, though at comparable level. Spikes in polarization do not occur during presidential election year 2016/2020 but thereafter. Congressional candidates exhibit downward trending levels of polarization until 2016, but are strongly polarized thereafter.

In the fourth and last part of the paper, I address the question whether audience engagement influences politicians' communication. One of the distinctive features of social media are horizontal flows of information, where politicians' receive direct and fine-grained feedback on

[6] This bias arises because the observed amount of speech is small relative to the number of words a politician could choose.

their communication. For each single message sent, politicians see how (much) social media users interact with the content and can thus learn which way of communication, for example which framing of an issue, garners attention.

I test whether audience engagement on social media reinforces or is potentially able to discipline a specific type of communication: attacks on political opponents. To do so, I focus on attacks that attracted spikes in audience engagement, calculated as a jump in reactions that exceeds the *politician-specific* time-varying reference point by 50% or 100%. Then, I compute the number, likelihood and fraction of attacks on political opponents in the 3,5,7 and 9 days before and after the sudden increase in audience engagement. Using within-politician variation over time, I estimate how responsive politicians are to two forms of audience engagement: A spike in the number of likes, as a proxy for *positive reception* of the tweet; and a spike in the 'ratio', as proxy for *negative reception* on social media.[7]

Unexpectedly, both types of audience behavior are associated with the same reaction among politicians: an increase in the number, likelihood and fraction of the number of attacks on political opponents. The effects are marginally stronger for *positive* compared to *negative* feedback, show no sign of decay over the course of 9 days, and are present irrespective of whether the jump in reactions is 50% or 100%. Instead, when I randomly assign a spike in reactions to an attack from a different politician on the same calender-day, I observe no rise of future attacks. While the response to positive feedback is intuitive, negative feedback could reinforce attacks, if social media algorithms disproportionately reward "unpopular" reactions, such as replies and quotes, by amplifying those posts and therefore bring visibility for the politicians.

This study contributes to a set of papers that document how social media changed politicians' behavior. Notably, Bessone et al. (2020) show that constituency access to social media leads politicians' to increase their online activity at the expense of offline efforts. Petrova, Sen and Yildirim (2021) demonstrate that presence on social media can be a viable tool for political candidates to attract campaign contributions. This result is more pronounced for political newcomers and arises through dissemination of information about themselves. I contribute to this literature by providing stylized facts for a large set of politicians' behavior on social media, including Twitter adoption, activity, audience engagement. I also present first estimates, how politicians communication behavior interacts with newly available granular information on

[7] The 'ratio' calculates the extent to which quotes and replies outnumber retweets. Intuitively, when users comment on a tweet rather than sharing it with their followers, it indicates that a tweet was unpopular. See https://www.washingtonpost.com/outlook/2021/10/27/twitter-amplifies-conservative-politicians/.

audience reception.

Additionally, this paper builds on studies examining ideological divisions among political elites. The notion of intensified partisan conflict in recent decades has found robust empirical support on measures such as congressional voting records (McCarty, 2019), candidate survey responses (Moskowitz, Rogowski and Snyder, 2018), congressional speeches (Gentzkow, Shapiro and Taddy, 2019) and campaign donation measures (Bonica, 2014). Closest to this paper is Gentzkow, Shapiro and Taddy (2019), who study the evolution of partisan difference through text from floor speeches in the US-Congress. While this paper builds upon their methodological innovation, it differs by studying polarization on social media using a comprehensive set of legislators and social media posts. This adds to our understanding of political polarization by comparing online to offline speech and examining more short-term fluctuations.

I also relate to a literature that has established stark evidence of affective partisan divisions among citizens (e.g., Iyengar and Westwood (2015), Iyengar et al. (2019) and Boxell, Gentzkow and Shapiro (2020)). It is common sense in this literature that partisan affection has become sufficiently stronger between 2008 and 2020, most strongly though in the US. Cross-country evidence points towards a strong association of mass - and elite polarization. While, causality can flow in both directions, several studies in political science argue that strategic choices of politicians play at least some role (e.g.,Lupu (2015), Banda and Cluverius (2018)). This work advances our understanding of the *emotional* side of the partisan divide by quantifying its rise among political elites and studies the determinants of affective polarization in congress.

The remainder of the paper is organized as follows. Section 2.2 explains and informs about the data about politicians and their social media handles, while section 2.3 assesses the representation of politicians on social media. Section 2.4 documents partisan differences in politicians behavior and audience engagement over time. Section 2.5 presents the measurement and results for affective polarization and Section 2.6 does the equivalent for ideological polarization. Section 2.7 tests whether politicians' communication is receptive of audience engagement. Section 2.8 concludes.

2.2 Building a Database for Politicians' Twitter Use

Data on politicians is obtained from Vote Smart, a non-profit, non-partisan research organization. The Vote Smart archive dates back until 1996 and contains electoral information on the universe of politicians, that is candidates and office-holders, for federal and state offices for a total of 106,736 unique politicians.[8] More specifically, the unit of observation in the Vote Smart data is at the politician-election level and provides the name of the politician as well as electoral information on her party, the date of the election, the office up for election and the election status (e.g., won). Less comprehensively, Vote Smart collects biographical and contact information, voting records, committee assignments, as well as ratings from 1,500+ interest groups.[9] Additionally, Vote Smart has data on 33,000+ local politicians from the legislative and executive branch, for example mayors or city/county council members. While ample information on their names, offices and location (e.g., the precise election district) are available, Vote Smart does not record when the local politicians assumed office or if they are still active.

To connect politicians to Twitter handles, I focus on the set of politicians, who were in office or participated in an election from 2008. Although Twitter was launched in March 2006, it didn't attract much popularity until an advertising campaign at the South by Southwest festival in March 2007. This leaves me with 75,420 politicians at the federal and state-level. To identify politicians' Twitter account(s) I rely primarily on information (i.e., linkages to Twitter accounts) from Vote Smart and Ballotpedia. For all missing observations, I use the Twitter API to search for accounts that match the politicians' name. For each politician, I retrieve the 30 best results, ordered by Twitter, and search in the self-description on Twitter for a) politics-related keywords (e.g., "senator", "representative") and b) politician-specific keywords such as the office, district or state of the respective politician. For queries that satisfy both of these criteria, I keep the account with the highest number of followers.

The resulting database encompasses Twitter accounts for 18,693 politicians, of which more than 70% are collected by Vote Smart or Ballotpedia. Figure A.2.1 plots the percentage of politicians with Twitter account over the total number of candidates in an election year. From 2008 onwards, the fraction of politicians using Twitter climbs around 10 percentage points every two years and reaches 57% in the 2020 elections.

Table 2.1 lists the matching rates by level of politics between the Vote Smart and the Twitter data. For each politician I keep the latest election she participated and in case of multiple occurrences the higher level of politics. While I achieve higher matching rates for higher levels of politics, this reflects to some degree the selection choice. I am able to link more than 92% of members of congress to at least one Twitter handle but still 45% of the officeholders at the state-level. Moreover, matching rates for officeholders are substantially higher than for

[8] as of 03/05/2022

[9] As an example, we know the birth date of 35,000+, the birth place of 52,000+ and previous legislative experience of 22,000+ candidates.

candidates. While, this is plausibly true of politics in general, it is at least partly a feature of this database. It is conceivable that I find a lower share of candidates using Twitter due to unsuccessful candidates changing their self-descriptions after an election or deleting their accounts.

The Twitter API grants access to the universe of statuses (tweets + retweets) sent by the politicians for a total of 61+ million statuses. For those statuses, I know the shared content (text, image, url) as well as the audience engagement with the status. In particular, Twitter reports for each status the number of retweets, likes and comments. However, Twitter does not provide historical information on the number of followers for the Twitter accounts. Through historical files collected from the George Washington University Library, I reconstructed the time-series of 600 members of the 115th and 116th congress, that is since January 2017.

In Section 2.3 and Section 2.4 I use the universe of politicians in the analysis, unless stated otherwise. For consistency throughout the textual analysis part, the remainder of the paper considers only *active* politicians from the Democratic or Republican party, either in office or in the year they run for a *general* election. I slice politicians in four samples: *(i)* Members of Congress, *(ii)* Congressional Candidates, *(iii)* State Officeholders and *(iv)* Candidates for State-Office. Overall, these samples encompass 12,300+ politicians, responsible for 23.26 million Twitter statuses during their active period. For the analysis I use only original tweets, thus excluding 7.8 million retweets.

2.3 Representation of Politicians on Social Media

This section investigates the representativeness of politicians on Twitter based on demographic characteristics and geography.

2.3.1 Demographic Representation

Table 2.2 lists summary statistics for the full sample of politicians and subsequently splits politicians according to whether or not they use social media along six characteristics: gender, race, age, legislative experience, education and whether a politician identifies as Democrat or Republican. I infer politicians' gender from their first name using U.S. Social Security data on the proportion of boys and girls for each name. I infer their race using census statistics on the distribution of ethnicities by surname.[10] The remaining characteristics were collected from

[10] Both operations are conducted using the R-package predictrace. While the gender distribution is largely in line with more incomplete information from Vote Smart, the race predictor seems to over-estimate the share of white

Vote Smart or Ballotpedia.

By and large Table 2.2 shows that groups that are generally under-represented in politics (e.g., women or POC) use social media to a higher extent. I find that among the politicians who use Twitter, nearly 34% are female, as compared to 26% in the full sample. Moreover, Twitter politicians are, on average, younger, higher educated and have less legislative experience than politicians who don't use Twitter. Among politicians, who use Twitter, I find that 55% identify as Democrats and 45% identify as Republicans.[11] While these numbers are qualitatively in line with estimates on Twitter users by Pew Research Center (2019), the under-representation of Republicans on Twitter is clearly less pronounced among politicians: Pew Research reports that 60% of Twitter users identify as Democrat and only 35% as Republican. Panel B compares politicians who use Facebook instead. Despite, the different user composition on the Facebook platform, the conclusions remain largely unchanged. Notably, also on Facebook Democratic politicians are in the majority, which might indicate a higher propensity for social media use in general.

2.3.2 Geographic Representation

Figure 2.1 maps the geographic distribution of the Twitter politicians. Panel (a) depicts for each state the percentage of Twitter politicians compared to the total number of candidates, averaged over elections from 2016 - 2020.[12] Across states, Twitter adoption varies from about 30% to more than 70%. The share of Twitter users among politicians is highest in Virginia, Arizona and California and lowest in Maine, Wyoming and North Dakota.

Panel (b) portrays the geographic dispersion of Twitter politicians. I extract information on the home-city from Vote Smart or geo-locate the self-declared city from a politicians' Twitter profile. The map of this geo-information coincides broadly with patterns of population density in the United States. In particular, rural areas in which the least number of people are active of social media (see Pew Research Center (2015)) are among the least represented among politicians.

---

politicians.

[11] Including politicians from third parties in this figure would result in 51% Democrats, 42% Republicans and 6% from third parties.

[12] This longer time horizon is chosen to ensure that all states had state-wide elections in the time-period.

2.4 Twitter Use

In this section, I provide new stylized facts on the evolution of Twitter adoption, Twitter activity and audience engagement among politicians. I document substantial differences in Twitter adoption, Twitter use and audience engagement across the Democratic and Republican party, arising around 2014/2015. Further, I find that Twitter adoption and activity are associated with electoral success.

2.4.1 Twitter Adoption and Tweeting Behavior of Politicians

As a starting point, Figure 2.2 plots the monthly average of accounts opened by party. The first wave of Twitter adoption dates back to late 2008 and yearly Twitter adoption remains high until 2012. After that, the number of newly added Twitter accounts nearly halves by 2016, but re-accelerates until the 2020 election. At a monthly frequency, Figure 2.2 shows a strong cyclicality with spikes in the number of openings in the beginning of federal election years. Up until 2015, I find no differences in adoption across the two parties. From 2015 but especially after 2016 the new wave of Twitter openings is driven by members of the Democratic party, while adoption rates for the Republican party stagnate.

Turning to Twitter activity, Figure 2.3 plots the average number of tweets by month for the median member of congress across party. The median Democratic legislator increased her tweets from 10 per month in 2009 steadily to more than 100 (original) tweets per months in 2021, as opposed to 75 tweets for the median Republican legislator. This party divide is non-existent in the initial period, in fact Republican legislators tweeted slightly more until 2012. From mid-2014 however, Twitter activity flattens out among Republican legislators, while Democrats increase their tweeting volume steadily. Since 2017 I observe a re-acceleration from the Republican party, such that the gap slowly closes.

Figure A.2.2 displays the tweeting behavior for the other samples. In recent years, congressional candidates tweet at around 40% lower volume than officeholders. For congressional candidates, party differences in tweeting volume emerge from 2016, with Democratic tweeting about 50% more than their Republican counterparts in the 2020 election. Panel (b) summarizes the tweets of officeholders at the state level. State politicians tweet around 17 times per month in 2021, which is about 20% of the volume of members of congress. At the state level, the Democratic dominance in Twitter use commences already from 2012 and exhibit a tendency to *widen* in recent years, due to a decrease in tweeting volume from members of the Republican party that is evident since 2017. In the 2020 election, tweets from Democratic members of the state legislature outnumbered GOP-tweets by a margin of 2.5 (1.5 million versus 0.6 million).[13] Last, Panel (c) shows the time-series of the number of tweets for candidates at the state level. In line with the previous results, I observe a lower level of tweeting from state candidates relative to officeholders, albeit less pronounced, and widening party differences since the 2014 election.

Next, I include retweets and consider the share of original over total tweets. Overall, I find that the percentage of original tweets amounts to 78% for members of congress, 76% for congressional candidates, 61% for state officeholders and 62% for candidates for state office. In Figure A.2.3 depict the evolution of this measure over time by party. Generally, the share of original tweets decreases over time, but stabilizes mostly from 2018. For the members of congress, I observe that Republicans rather than Democrats relied more on original tweets until 2016 and less thereafter. For all other samples, Republican politicians write more original tweets than Democrats. However, from 2018 onwards, Democrats increase their share of original tweets and as of 2020 break even with Republicans.

2.4.2 Audience Engagement

One of the features that distinguishes social media from traditional media is that politicians get immediate and differentiated reactions to the content they share. Twitter users have the option to retweet, so to share the politicians' status among their followers, to like the status and to comment the status, through quoting or direct replies.

Figure 2.4 summarizes these reactions by calculating the average number of reactions per month for the median legislator by party. The average number for all types of reactions among both parties is close to zero until 2016. From that year onwards, all types of reactions grow steadily for both parties. In 2021, the median Democrat (Republican) received around 60 (40) likes, 16 (13) retweets, 1 (2) quotes and 7 (19) replies. January 2021 represents abnormal levels among all types of audience engagement, due to the Capital Riots. Moreover, from 2016 a dichotomy among the reactions is apparent: The majority of likes and retweets accrue to Democrats, the majority of quotes and replies accrues to Republicans.

Turning to the other samples of politicians in Figure A.2.4, A.2.5, A.2.6, three patterns emerge: First, audience engagement is lower for candidates as compared to officeholders and substantially lower for state-politicians. For example, the median number of likes received in

[13] This is not primarily a compositional effect, as the number of Democratic legislators with Twitter accounts in 2020 is just 7% higher than their Republican counterpart.

2021 is around 85% lower for state-representatives relative to members of congress. Second, in all samples the Democratic party receives markedly more audience engagement from 2016. As of 2020/2021 Democratic politicians receive 30% - 50% more likes than Republicans. Third, except for the sample of candidates for state office, Republicans draw more attention in form of replies.

Based on this dichotomy, I calculate a measure for *unpopularity*, called the "ratio".[14] A tweets is 'ratioed' when quotes and replies exceed retweets. When users refrain from sharing content with their followers and rather comment on a tweet, it indicates that a tweet was unpopular, likely because many people disagreed with their take. Figure 2.5 shows that since 2011, Republican tweets are without exception more unpopular than Democratic tweets, consistent with a left-leaning twitter audience. Note however, that the average Republican tweets gets shared more often than commented until the election of Donald Trump in 2016 and only afterwards the "ratio" jumps dramatically. The ratio also trends upwards for Democrats but remains far away from the neutral value of 0. The consequences of unpopular audience engagement with Republicans' tweets are unclear, but several observers have noted that Twitter's algorithms could interpret this as increased engagement and amplify these tweets.[15]

Audience engagements for politicians' *accounts* is portrayed in Figure 2.6 as the average number of followers for the median party legislator. While the median members of congress of both parties started from a roughly equal level of followers (16,000 versus 14,000) in 2017, the followership evolved strongly divergent thereafter. The Democratic legislators now reach out to 38,000 as compared to 21,000 for the median Republican legislator. There is neither a slowdown for Democrats, nor an acceleration for Republicans observable.

2.4.3 Twitter Use and Electoral Success

Given the substantial differences in the adoption and use of Twitter between Democrats and Republicans, a natural question to ask is whether these behaviors influence electoral success. I estimate the following specification:

$$Y_{c,s,o,e} = \beta X_{ce} + \Omega_e + \Psi_c + \Phi_s + \alpha_o + \epsilon_{c,s,o,e} \tag{2.1}$$

where $Y_{c,s,r,e}$ equals 1 if a politician c from state s running for office o in an election-year e

[14]The ratio for each tweet is computed as follows: $\frac{quotes+replies-retweets}{quotes+replies+retweets}$.

[15]Twitter reported that its timeline algorithm is more likely to amplify right-wing politicians than left-wing politicians.https://blog.twitter.com/en_us/topics/company/2021/rml-politicalcontent

has been winning the election, and 0 otherwise. The unit of observation is thus at the politician-election level. The main explanatory variables is X_{ce}, which is (i) a dummy whether a politician has opened a Twitter account before the election or (ii) the number of tweets sent in the election year (before the actual election). Ω_e, Ψ_c, Φ_s and α_r are full sets of election-year, candidate, state and office fixed effects. The main variation exploited is thus *within* the same candidate across time. The standard errors are clustered at the politician-level.

Two methodological concerns prevent a causal interpretation of the regression coefficients. On the one hand, the explanatory-variable is likely to exhibit measurement error. If unsuccessful candidates are more likely to delete their Twitter accounts, the β-coefficients are upwards-biased. On the other hand, time-varying omitted variables could jointly determine electoral success and the adoption or use of Twitter.

Table 2.3 provides the regression results for Twitter adoption. Throughout the point estimates are positive and highly statistically significant. Column (1) represents the baseline-regression and shows that the adoption of Twitter is associated with a 6 percentage point higher probability of winning the election. Taken at face value the adoption of Twitter would increase the chance of electoral success by about 15 percent. Altering the clustering of standard-errors to the politician-year level, adding time-varying controls for being an incumbent and or electoral district fixed effects don't markedly change the point estimate (see Columns 2 - 4). Next, considering only general elections lowers the electoral advantage from Twitter adoption to 4.3 p.p. Column (6) uses only observations from 2016 onwards. In line with Figure 2.2, that showed that party differences in Twitter adoption occurred mostly from 2016 onwards, Twitter adoption is more *strongly* related to electoral success in recent times. Note also, that in this more recent period the concerns for measurement error are weaker due to more extensive data-collection from Vote Smart and Ballotpedia. Column (7) selects only candidates from the two major parties and thus precludes the interpretation that the coefficients are driven by electoral advantages versus independent candidates. Last, in Column (8) I use only the sample of politicians for whom I have a Twitter account, for which measurement-error should the downward-bias the coefficient. Indeed, I observe that the point estimate drops to 1.5 percentage points, but remains significant at the 5-percent level.

Table 2.4 estimates the same series of regressions, using as the explanatory variable the log-number of tweets, replacing missing observations with 0. Across all columns, Twitter activity is a strong predictor of electoral success, with p-values below the 1-percent level. The magnitude of the effect varies across samples, with a 1% increase in the number of tweets being associated with a 3.9 p.p. higher chance of electoral success in the sample considering period after 2016, to a 2 p.p. higher probability of assuming office in the sample considering only general elections.

2.5 Affective Polarization

In this section, I develop a measure to quantify and track affective polarization among politicians on social media. I show that emotionality and affective polarization have increased substantially over time, especially among members of congress; that this rise is attributable to both political parties and that affective polarization is more pronounced among members of the opposition-party and ideological extremists.

2.5.1 Measuring Affective Polarization in Political Speech

Affective polarization is the gap between individuals' positive feelings toward their own party and negative feelings toward the opposing party (Iyengar and Westwood, 2015). This concept is thus largely distinct from differences in policy positions between Democrats and Republicans, as it captures divisions based on affect and identity (Iyengar et al., 2019). In a host of studies, this concept is identified through survey self-reports where respondents are asked to place Democrats and Republicans on a "feeling thermometer" that ranges from cold (0) to warm (100). Using information on the respondents party affiliation, affective polarization is then measured as the difference between in-party feeling minus out-party feeling (Boxell, Gentzkow and Shapiro, 2020).

I adapt this concept to messages from political elites in two steps. In the first step, I identify tweets that refer to the political in-group or the political out-group. I mimic the various ways how politicians can speak about each other by considering tweets that *(i)* tag other politicians (e.g., @mikepence), *(ii)* mention variants of name and titles of a politician (e.g., Nancy Pelosi/Speaker Pelosi/Rep Pelosi) or *(iii)* contain keywords related to conservative/liberal ideology (e.g., democrats, GOP, conservative, left-wing). In the second step, I determine the sentiment of each tweet using VADER, a lexicon and rule-based sentiment analysis tool.[16] VADER provides for each tweet a positivity score, a negativity score and a neutrality score that form a compound sentiment index that takes into account the number of neutral words. As recommended by the developers of VADER, I use the compound index and normalize it between 0 and 1, neutral sentiment being thus 0.5.

[16] https://github.com/cjhutto/vaderSentiment.

VADER offers several advantages over other sentiment-dictionaries that have been employed in the economics literature (most prominently LIWC): VADER is trained on social media and has been empirically validated against human judges. It thus incorporates emoticons, sentiment related acronyms (LOL) and slang words. Moreover, VADER incorporates negation and syntactical rules that switch polarity (e.g., but). Last, VADER has nearly twice the vocabulary of other dictionaries (LIWC, Finn) and takes into account both polarity and the *intensity* of sentiment (great > good > okay).

For each month, I then compute in-party sentiment as the average sentiment among the tweets from Democrats (Republicans) about Democrats (Republicans), out-party sentiment as the average sentiment among the tweets from Democrats (Republicans) about Republicans (Democrats) and affective polarization as the difference between in-party and out-party sentiment. I exclude from the analysis retweets, tweets in which the author refers to him/herself and tweets that mention members from both parties.

2.5.2 Emotionality

Before turning to differential sentiment towards members of the out-party, I investigate overall differences in emotionality in social media speech. Emotional language is strongly on the rise in congressional speeches (Gennaro and Ash, 2022) and commonly regarded as tool to push policy positions (Jerit, Kuklinski and Quirk, 2009). On social media, moral-emotional language fosters the diffusion of political messages, especially concerning polarizing issues (Brady et al., 2017).

Figure 2.7 depicts the time-series for an emotionality index for the members of congress, Figure A.2.7 for the other samples. I use the fact that VADER calculates the proportion of neutral words for every tweet and construct a simple index of emotionality that is 1−neutrality score. Emotionality increases from 0.16 in 2009 to its peak around 0.22 in 2016 and trends weakly downwards thereafter. From the beginning of the sampling period in 2009 until the end of 2016, the smoothed trend in emotionality of the Democratic party lies above that of the Republican party. After that, the two trends converge. Consistently, emotionality reaches its high-point also in the sample of congressional candidates in 2016. Democratic candidates for Congress lead on emotionality before 2016, lag behind in 2016, and remain on par in the aftermath. At the state-level emotionality climbs until 2018 and flattens out thereafter. Republican state-office holders start to close the emotionality gap from 2013 and surpass Democrats by 2017. This switch in emotionality occurs already in 2014 for state-candidates. On average, candidates choose slightly higher levels of emotionality than officeholders.

Figure A.2.8 replicates the calculations using only retweets. Two patterns are striking. For one, the level of emotionality in retweets is considerably lower than in original tweets. For another, the trend in emotionality is entirely flat for members of congress, while much less pronounced in other samples.

2.5.3 Affective Polarization Increases Steadily

Using the information about whom tweets speak, I find a total of 979,231 unique tweets mentioning at least one of the (members of) political parties (around 20% of all original tweets). Out of those tweets, 44% refer to the political in-group and 56% speak about the out-group. The average sentiment towards the political in-group is around 0.67, the average sentiment towards the political out-group is 0.49 and thus marginally negative. The average gap in the sample is 0.18 from members of the Democratic party and 0.17 from members of the Republican party.

Figure 2.8 depicts the evolution of in-party sentiment, out-party sentiment and affective polarization over time. It shows that *(i)* in-party sentiment is always more positive than out-party sentiment; *(ii)* in-party sentiment and out-party sentiment grew apart over time, with a flat segment between 2014 and 2017 (except the 2016 election), such that affective polarization reaches its' peak in late 2021; *(iii)* more positive in-party sentiment drives affective polarization before 2014, and conversely more negative out-party sentiment from 2018; *(iv)* intuitively, the affective polarization measures follows a cyclical pattern with high-points around federal elections, driven presumably by stronger in- versus out-group rhetoric (consistent with Lau et al. (2017)) but is not limited to election years. Over the last ten years, the level of affective polarization rises from around 0.13 in 2011 to around 0.24 in 2021 (or an increase of 66%).

This development is consistent with the trends in affective polarization among *voters* observed in Iyengar et al. (2019) and Druckman and Levy (2021).[17] Both report that voters' out-party negativity and affective polarization grows strongly over from 2008 - 2020. More nuanced, they observe a relatively flat trend between 2012 and 2016 and a growing partisan divide thereafter. Similarly, in Figure 2.8 the out-party negativity and affective polarization grow in increasing pace after 2016. Despite, this overlap in trends and timing, there are two notable differences. First, in-party sentiment is becoming more positive among politicians, but remains stable for voters. Second, the level of partisan affect is severely higher among voters than among politicians. The reported level of affective polarization in the 2020 American

[17] Similar to the time-series for the US, Boxell, Gentzkow and Shapiro (2020) find a strong correlation in cross-country estimates between mass- and elite polarization.

National Election Studies is roughly double the level estimated in my time-series.

Several reasons might come to mind to explain this discrepancy. On the one hand, partisan divisions in surveys have been criticized to be subject to expressive survey responding and thus may overestimate the affective polarization (Prior et al. (2015); Bullock and Lenz (2019)). On the other hand, my time-series may underestimate the true extent of affective polarization. Note first, that VADER is trained on social media but not specialized to political language. For example, the word "corrupt" conveys no sentiment in VADER, but in connection with a politician is very negative. Second, private survey response are different from speaking *publicly observable* about and with other politicians on social media.[18]

Robustness. This trend and growth rate in the way parties express emotion towards each other is undiluted for a host of alterations in the way affective polarization is computed. Across the sensitivity tests employed, the minimum increase in affective polarization in the last 10 years never falls below 41%. In Figure A.2.10 I show the evolution of affective polarization is unchanged (i) when computing affective polarization within the politicians and averaging thereafter (Panel a.), (ii) excluding tweets about the presidents (Panel b.) and (iii) using an alternative sentiment dictionary, called AFINN (Panel c.). Exercise (i) and (ii) ensure that the results are not driven by a small set of politicians that either tweet extensively or are extensively tweeted about. Exercise (iii) considers the AFINN sentiment lexicon, that is also trained to classify microposts but has only $\frac{1}{3}$ of the vocabulary. Consequently the observed *level* of affective polarization is lower, yet the trend is similar to Figure 2.8.

Finally, Figure A.2.11 shows the time-series of affective polarization for the other samples of politician. Several patterns are noteworthy. First, affective polarization among congressional candidates rises strongly until the 2016 election (by 44%), driven by out-party negativity, but declines thereafter. The average distance between in- and out-party among congressional candidates is 0.15, as compared to 0.18 among members of congress. Second, state-officeholders are considerably less polarized than their federal counterparts, driven by both, less favorable in-party sentiment and more favorable out-party sentiment. Nevertheless, a steady upwards trend is discernible, such that affective polarization in 2021 is 0.15 which is about 65 percent of the level for members of congress. It is also worth noting that average out-party sentiment

[18]To underscore this I construct three time-series based on the type of mention in Figure A.2.9. Panel (a.) plots the estimates when politicians "tag" each other. Despite strong surges in affective polarization before elections, no upward trend is visible until 2020. Note also, that the out-party affect is positive for the majority of the time. Instead, the gros of the rise in affective polarization from Figure 2.8 is driven by mentioning another politicians' name - without tagging her - and references to the other parties ideology (see Panel b. and Panel c.). The interpretation of these finding is straightforward: Politicians that speak with each other are more positive towards their counterpart, as when they speak about each other.

among state-officeholders is around 0.54 and therefore, in contrast to members of congress, clearly positive. Third, also candidates for state-office exhibit a weak increase in affective polarization, peaking in 2018.

2.5.4 Determinants of Affective Polarization

In this section I assess how affective polarization varies across members of congress, taking into consideration partisanship in roll-call votes, politician characteristics and party opposition status. To test for statistical significance, I regress those factors on affective polarization aggregated at the politician-year level. Affective polarization is calculated on the sample of tweets that refers to other members of congress or the other party, hence excluding tweets about presidents or governors. Following Gennaro and Ash (2022), I include chamber-year fixed effects and cluster standard errors at the politician level, hence allowing for serial correlation in the error term by politician over time.

Table 2.5 reports the regression results. To gauge the relationship between ideological extremism and affective polarization, I use the squared DW-NOMINATE, a standard measure based on roll call votes, such that more extreme voting takes larger values (as in Gennaro and Ash (2022)). Column (1) shows that more ideological voting is associated with more affectively polarized speech, which holds when including politician demographics (Column 9). Note, that this relationship is ex-ante not entirely clear, as ideological extremists also appeal to fringe voters by spurring political conflict within their own party (Kirkland and Slapin, 2018).

Next, I test for party asymmetry in affective polarization. Several studies have shown that the surge in ideological polarization from the 1980s to 2010 can be attributed to a large degree to an alienation of the Republican from moderate positions (McCarty (2019), Moskowitz, Rogowski and Snyder (2018)). As initial visual evidence Figure A.2.12 disaggregates the time-series of affective polarization by party. Overall, both parties exhibit clear upward trends in affective polarization, yet with different inflection points. For the Democratic party, I observe modest increases in affective polarization from the beginning of the sample period until 2016. Since then, affective polarization and especially out-party negativity are more strongly on the rise and peak in January 2021. In contrast, the Republican Party exhibits a pronounced ascent in affective polarization until the end of 2016, followed a substantial two-year drop for two-years and fast acceleration since 2019. Overall, Column (2) finds members of the Republican party to communicate less divisive in their tone about other politicians.

Column (3) shows that female politicians communicate more polarized than their male counterparts. This finding is in line with evidence on gender differences in affective polarization among voters (Ondercin and Lizotte, 2021) and consistent with female politicians higher levels of emotionality in general (Gennaro and Ash, 2022). Instead, higher levels of education mediate divisive communication across in- and out-party members (Column 4). Within-member shifts in legislative experience are not related to shifts in affective partisanship.

Last, I investigate the role of a parties opposition status. Previous research has shown that minority-party politicians communicate more emotional in general (Gennaro and Ash, 2022). In addition, it is plausible that minority politicians are more prone to blame/attack members of ruling party than vice versa (Green, 2015), while majority-party politicians speak less negativity about their colleagues as they depend more on their collaboration. Consistent with the expectation, Table 2.5 shows a strongly significant and positive association of opposition status with higher affective polarization, even when netting-out the candidate-specific mean values (Column 8).

These relationships remain statistically significant when pooling all factors together in a single regression (Column 9). In terms of magnitude, I find that political extremism and majority status are the strongest determinants of affective polarization.

2.6 Ideological Polarization

Building on the work of Gentzkow, Shapiro and Taddy (2019), I quantify ideological differences in politicians' speech on social media in this section. Partisanship is defined as ease with which a neutral observer could infer a congressperson's party from a single expression. To measure partisanship, I rely and lay-out the model of political speech developed Gentzkow, Shapiro and Taddy (2019). In subsequent steps, I delineate the pre-processing of the data and present the estimates. Different from Gentzkow, Shapiro and Taddy (2019) my (preliminary) estimates do *not* show a strong drop in partisanship in the mid 2010s and slope upwards after 2016.

2.6.1 Measuring Partisanship in Political Speech

I assume a multinomial model of text-generation:[19] A tweet by politician i at time t, for example a year, is represented as a vector of word counts, $\mathbf{c}_{it}$, over J available words. A politician i chooses $\mathbf{c}_{it}$ of size $m_{it} = \sum_j \mathbf{c}_{it}$ according to the choice probabilities $\mathbf{q}_{it}^{P_i} = q_{it1}, ..., q_{itJ}$.

[19]This modelling approach has been successfully applied in related contexts, for example as a measure of media bias (Groseclose and Milyo, 2005)

This model can be written as

$$\mathbf{c}_{it} \sim MN\left(m_{it}, \mathbf{q}_{it}(\mathbf{x_{it}})\right) \tag{2.2}$$

and is fully characterized by the total amount of speech by a speaker i at time t and the choice probabilities $q_t^{P_i}(x_{it})$, defined over the available words J for speaker i from a party P ($P_i \in \{Democrat, Republican\}$) at time t.

The choice probabilities are estimated through the following set of multinomial logistic regressions

$$\begin{aligned} q_t^{P_i}(\mathbf{x_{it}}) &= \frac{e^{u_{ijt}}}{\sum_j e^{u_{ijt}}}, \\ u_{ijt} &= \alpha_{jt} + x_{it}'\gamma_{jt} + \phi_{jt}\mathbf{1}_{i\in R_t} \end{aligned} \tag{2.3}$$

and depend on three parameters: the parameter α_{jt}, that represents the baseline usage of a phrase j in session t; the vector γ_{jt}, that represents the effect of the covariates x_{it} on the probability to choose phrase j in session t and last the parameter ϕ_{jt} that represents the partisan effect on the choice to use phrase j at time t.

The inclusion of covariates follows the usual intuition in economics, namely to control for factors that are correlated with the outcome (speech) and the variable of interest (party differences), but are not indicative of **ideological** cleavages. For example, politicians from Arizona might refer more often to the Arizona Senate (#AZsen) and at the same time have a relatively high probability to be Republican. However, increases in the usage of #AZsen over time reveal nothing about diverging political ideology. Therefore, the baseline specification includes as set of covariates x_{it} indicators for the chamber, gender, state, Census region and majority control. Additionally, I add for year times Census region fixed-effects to account flexibly for different paths of regional speech.

To quantify partisanship, we define polarization at $\mathbf{x}$ as the posterior probability a neutral observer assigns to a politicians' true party after hearing word j:

$$\pi_t(\mathbf{x}) = \frac{1}{2}\mathbf{q}_t^R(\mathbf{x}) \times \rho_t(\mathbf{x}) + \mathbf{q}_t^D(\mathbf{x}) \times (1 - \rho_t(\mathbf{x})) \tag{2.4}$$

where

$$\rho_{jt}(\mathbf{x}) = \frac{q_{jt}^R(\mathbf{x})}{q_{jt}^R(\mathbf{x}) + q_{jt}^D(\mathbf{x})}$$

Partisanship is thus the sum of two terms: The first term on the right-hand side of equation

2.4 is the product of the observers prior (1/2), the propensity to use word j by the Republican party (given x) and the posterior probability ϕ_t. The second term on the right-hand side is then the posterior belief of an observer of a neutral prior to a politician being Democrat if the politician chooses phrase j in session t and has characteristics $\mathbf{x}$. Average polarization, which is the value of polarization reported in the results, averages equation 2.4 across characteristics $\mathbf{x_{it}}$. Intuitively, partisanship will be high when, given characteristics $\mathbf{x}$, some phrases are spoken relatively of one party relative to the other.

However, because the set of phrases a politician chooses from is large in comparison to the actual number of phrases spoken, many observed differences between parties are due to chance rather than partisanship. Not accounting for this finite sample bias can lead to severely upward biased estimates of the true extent of party divisions. I tackle this problem through the inclusion of a Lasso penalty on the coefficients of interest. Specifically, as in Taddy (2015), the model parameters $\{\alpha_t, \gamma_t, \phi_t^T{}_{t=1}\}$ are estimated using distributed multinomial regressions with a lasso-penalty. The estimator is given by the following minimization problem:[20]

$$\hat{\alpha}_{jt}, \hat{\gamma}_{jt}, \hat{\phi}_{jt} =_{\alpha_{jt},\gamma_{jt},\phi_{jt}} \sum_t \sum_i [m_{it} \exp(\alpha_{jt} + x'_{it}\gamma_{jt} + \phi_{jt}\mathbf{1}_{i\in R_t}) - c_{ijt}(\alpha_{jt} + x'_{it}\gamma_{jt} + \phi_{jt}\mathbf{1}_{i\in R_t}) + \psi(|\alpha_{jt}| + ||\gamma_{jt}||_1 + \lambda_j|\phi_{jt}|] \quad (2.5)$$

The Lasso penalty, $\lambda_j|\phi_{jt}|$, helps to control the amount of small sample bias by shrinking the polarization coefficients towards zero and yields a sparse solution, as some coefficients will be set exactly to zero.[21] The last step computes partisanship $\bar{\pi}_t$ inserting these estimates in equation 2.4.

Validation

While the estimator severely reduces the amount of bias arising from words said by legislators from different party purely by chance, finite sample bias cannot be controlled fully. I determine the extent of finite sample bias through a permutation test. I draw 100 samples, each containing 20% of the speakers, from the actual data and compute partisanship for each sample. I then

[20]To estimate equation 2.6 I use the *dmr* package in R.

[21]As in Gentzkow, Shapiro and Taddy (2019) the lasso penalty λ_j is chosen computationally. In particular, the optimal λ_j is determined by a imposing a large value of λ_j such that all party loadings are forced to zero, and then by a stepwise decrease in λ_j up to the value that minimizes a Bayesian Information Criterion (see Taddy (2015) for details). Moreover, to assure convergence I follow Gentzkow, Shapiro and Taddy (2019) and include a constant penalty on the other parameters of the model. The minimal penalty is $\phi = 10^{-2}$ and higher than the value in Gentzkow, Shapiro and Taddy (2019), but note that we operate on samples that are 1/10 their size.

re-estimate the partisanship on the same 100 samples but re-assign political parties randomly. More specifically, for each congress session the probability to get assigned the Republican party label is given by the share of Republican politicians in that session. In this "random" series the true value of partisanship $\pi_t = \frac{1}{2}$ and the deviation from $\frac{1}{2}$ quantifies the finite sample bias.

2.6.2 Pre-Processing

The steps to prepare data are chosen try to mimic Gentzkow, Shapiro and Taddy (2019) as close as possible in order to ensure that differences in the estimated time-series arise purely through the means of communication, that is a parliamentary speeches versus social media.

Given the much shorter time-horizon, the number of unique members of congress, with active social media accounts, is 929 (12% of Gentzkow, Shapiro and Taddy (2019)) and the number of politician-sessions is 3,500 (10% of Gentzkow, Shapiro and Taddy (2019)). In the construction of the vocabulary I remove punctuation, capitalization, digits and stopwords, including place names and hashtags (cities, counties, states); dates (months, weekdays), full politician names and procedural words. Words are stemmed, using the Snowball stemmer. Finally, I include in the vocabulary only features that are mentioned at least 10 times in at least one session, mentioned in at least 10 unique speaker-sessions, mentioned by at least 5 unique speakers and mentioned at least 50 times across all sessions. The resulting vocabulary contains around 34,000 unique features, 88% bigrams and 12% hashtags (or 7% of the vocabulary in Gentzkow, Shapiro and Taddy (2019)).[22] The twitter data are then aggregated into a document-term matrix, where an observation summarizes all tweets by a given speaker in a year. The data on the set of covariates comes from Vote Smart.

2.6.3 Results

Figure 2.9 plots the series of average partisanship over the 100 samples. The gray shaded area characterizes the estimates for polarization from the data in which politicians party assignment is randomized. The upper and lower bounds correspond to the 5th and 95th quantile from the placebo estimates. Accordingly, the pink time-series represents the estimates and confidence-interval on the real data. The hypothetical series shows no pronounced trends, is close to zero for each year and gets more precisely estimated over the years due to higher verbosity. Thus, finite sample is properly controlled for in the estimation and trends in the actual data stem from

[22] This discrepancy in *total* vocabulary will be largely abated in the estimation as language use in each single draw is relatively more consistent in my shorter time-span compare to the 140-years in Gentzkow, Shapiro and Taddy (2019). My average sample uses more than 90% of the total vocabulary.

real divergence in how politicians speak rather than noisy estimation. Except in the starting year 2009, observed polarization lies always outside the placebo distribution, providing strong statistical evidence of polarization on social media.

In Panel (a), I re-estimate the time-series in Gentzkow, Shapiro and Taddy (2019) with using tweets-by-year instead of bi-annual congressional speeches. Average partisanship increases up to the 112th congress (until 2013) and moderately declines up to 2016. Ideological party conflict rises again to reach its previous level in 2018, trends weakly downward for two years but shoots through the roof in 2021. Disaggregating the time-series by year, provides some nuanced insight: ideological polarization does not shoot off during years of presidential elections (see 2016 and 2020) but in the years after. This may emphasize the role of replacement as a driver of polarization in those years.[23]

Comparing my measure for social-media polarization to the measure in Gentzkow, Shapiro and Taddy (2019) on congressional speeches, I find broadly coinciding time trends, that is an increase in polarization until 2013 and a somewhat less pronounced decline towards 2016. One potential reason for this difference is a substitution effect from off-line modes of communication towards social media. Bessone et al. (2020) find evidence for such a switch in the number of mentions of the municipality. The point-estimates in the social media time series moderately above the congressional time series, but not statistically different. Yet, the 2021-endpoint of the estimation is nearly double the size of the peak in polarization in the Gentzkow-time series.

On the other hand, one could argue that the estimates from social media time-series are *markedly* higher than congressional speeches, as they are less selected sample of political speech. Party leaders are widely known to influence the congressional agenda and whom to delegate floor time on which policy issue (see e.g., Canen, Kendall and Trebbi (2020) or Proksch and Slapin (2012)). Like this party leadership can maximize party-unity, that is consistent party-message and suppress (expressed) dissent.[24] As party leadership has no procedural control who speaks on social media, we may observe a more truthful distribution of legislator preferences in contrast to floor speeches. In that case, my estimates from social media would represent substantially higher levels of preference divergence among politicians.

Panel (b) of Figure 2.9 presents the time-series for congressional candidates. I observe a

[23]Figure A.2.13 provides the estimates for the bi-annual congressional speeches and shows the estimates for a minimal penalty of $\phi = 10^{-3}$, that is closer to the originally proposed. The trends in both time-series are unchanged, but I observe an increase in the signal to noise ratio in the latter.

[24]Dissent may arise even if politicians share the *same* preferences with the party leadership, simply due to differences in the types of districts represented (Polborn and Snyder Jr, 2017).

declining level of ideological polarization from 2010 to 2016 up to the point, where the point estimates from the random time-series and actual polarization are indistinguishable. In contrast, I see strong increases in partisanship for the 2018 and 2020 elections. The level of polarization of congressional candidates is distinctly lower in the 2012-2016 elections and almost the same in the 2018 and 2020 elections. Comparing candidates' ideological polarization to affective polarization in Figure A.2.10, I note exactly opposite time trends. Affective polarization rises until 2016 but decreases thereafter. This pattern is however not found for members of congress.

2.7 Does Twitter Backlash Discipline Politicians?

In comparison to the traditional media environment, social media have the distinctive feature that politicians receive granular and real-time feedback on their communication. Each single message is liked, retweeted or commented by social media users and gives politicians a chance to learn whether certain types of messages *a)* attract audience engagement at all and thus bring visibility and *b)* how users interact with those messages, for example more positive or more negative.

In what follows I investigate whether and how politicians' communication is receptive to sudden leaps in audience engagement. I consider the case of a politician, who attacks an opponent, measured as a tweet that refers to (a politicians of) the other party and has a sentiment score below 0.5, which is the neutral value. To test whether Twitter users engagement reinforces or dampens future attacks, I look at attacks that attracted abnormal levels of user engagement. I define as abnormal user engagement reactions that exceed the politician-specific average in the two weeks before, by more than 50% (100%). I consider two types of reactions: Likes as a way to express positive feedback and the 'ratio', so the extent to which quotes and replies outnumber retweets (as explained in Section 2.4).

Within this setup, I estimate the following specification:

$$Y_{c,s,o,(d+t)} = \beta \, \text{post-engagement}_{cd} + \gamma_y + \Omega_d + \Psi_c + \Phi_s + \alpha_o + \epsilon_{c,s,o,d} \tag{2.6}$$

where I aggregate tweets at the candidate-day-level and c indexes a politicians, s the states, o an office, y the years and $d+t$ a time-interval. The outcome of interest takes the form of *i)* number of attacks, *ii)* an indicator for whether a politician attacks, *iii)* number of tweets and *iv)* the number of attacks over the number of tweets in a period $d+t$. I consider tweets in the interval of $t \in (3, 5, 7, 9)$ days before and after engagement at a given attack jumps, excluding the day of the spike from the analysis. The main explanatory variables is $post - engagement_{cd}$, an indicator variable that takes a value of 1 in the t days after a jump in reactions and 0 in the t days before. Ω_d, γ_y, Ψ_c, Φ_s and α_r are full sets of calender-day, year, candidate, state and office fixed effects. The standard errors are clustered at the politician-level.

Overall 50,000+ (36,000+) attacks on other politicians attracted spikes in likes (ratio). Considering subsequently only cases in which I observe a single attack with abnormal audience engagement in the period, thus excluding cases with multiple attacks with jumps in audience engagement, leaves me with 26,000+ (21,000+) attacks. This coding choice ensures the presence of a uncontaminated pre-period, but the strong drop in the number of attacks with necessitates further sensitivity tests regarding this choice.

Starting with the effect of *positive* feedback on communication, Table 2.6 shows the results of regression 2.6, where Column (1) - Column (4) are the four dependent variables and Panel (a.) through (d.) stand for the different time-windows of 3, 5, 7 and 9 days. Note first, that all 16 coefficients are positive and highly-significant. As dependent variables are standardized throughout, I find that in the three days after a sudden jump in likes, the same candidate increases her number of attacks by 0.12 standard deviations compared to before the spike in engagement. Over time, this effect decays only weakly. While Column (2) shows that the strong audience engagement increases the likelihood to attack an opponent, we observe in Column (3) that also Twitter activity, measured as number of tweets sent, is on the rise. Last, I find in Column (4) that also the ratio of attacks over total tweets climbs persistently over 9 days.

Table 2.7 replicates the same series of regression, however looking at periods before and after a jump in effect of *negative* feedback. Surprisingly, again all coefficients are *positive* and significant, indicating that abnormal backlash from Twitter users on a given attack, induces *more* attacks in the future. This holds true across all columns, such that politicians attack more and send more tweets, while the fraction of attacks over tweets increases as well. Also in the case of backlash the effect is strongly persistent in magnitude and significance.

Before interpreting this unexpected results, I test the sensitivity regarding the choice of what constitutes a spike in the ratio and conduct a permutation test. Table A.2.1 and Table A.2.2 use spikes in the number of likes and respectively the ratio of more than 100%. This cuts down the number of attacks considered by around 50% but results are consistent throughout: Jumps in likes and the ratio for an attack are followed by an increase in future attacks. Table A.2.3 and Table A.2.4 present the results of a permutation test, that re-estimate regression 2.6

selecting politicians that were launching attacks on the same day but did not receive a spike in engagement. Throughout, I find that coefficients are often precisely estimated zeros. Out of 32-coefficients, I find 4 being significantly different from 0. As those 4 coefficients refer to different specifications in different time-intervals, I conclude that they are rather false-positive than actual increases in the attacks.

One reason why Twitter backlash *reinforces* rather than dampens politicians' attacks on opponents is that backlash may result in visibility on social media. Twitters' algorithm has been shown to amplify more right-leaning politicians (and content) compared to left-leaning politicians (see Huszár et al. (2022) and footnote 13). Observers noted, that conservative politicians have inflated 'ratios' and argue that social media algorithms might interpret this as heightened engagement and amplifies them (see footnote 6). In line, the Facebook papers reveal that from 2016 its algorithm started to overweight posts with an increased number of comments and replies compared to likes in its decision what to amplify.[25] Figure 2.5 demonstrates that Republicans have persistently higher ratios than Democrats, with trends for both groups sloping upwards from 2016. Thus, if backlashed attacks resulted in higher amplification for politicians, an attention-maximizing strategy would be to increase the number of attacks.[26]

2.8 Concluding Remarks

A growing body of research investigates the welfare effects of the rise of social media, with ambiguous conclusions (Allcott et al., 2020). While not assessing the effect of social media, this paper establishes a set of novel results about the communication of a comprehensive set of politicians in the United States using the universe of their tweets from 2008 to 2021.

First, politicians on Twitter are more like from under-represented and disadvantaged groups in politics such women, non-whites and newcomers. Second, I observe pronounced partisan-differences in social media usage and audience engagement between Democrats and Republicans, across all levels of government - with a steady increase in activity from Democrats but flattening activity among Republicans. I show that social media adoption and activity increase the likelihood of subsequent electoral success. Third, a large set of findings documents a widening inter-party hostility. Politicians from both parties use social media increasingly to speak negatively about the political opponent. This behaviour is bipartisan, not driven by a

[25] https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/, accessed 10/04/2022

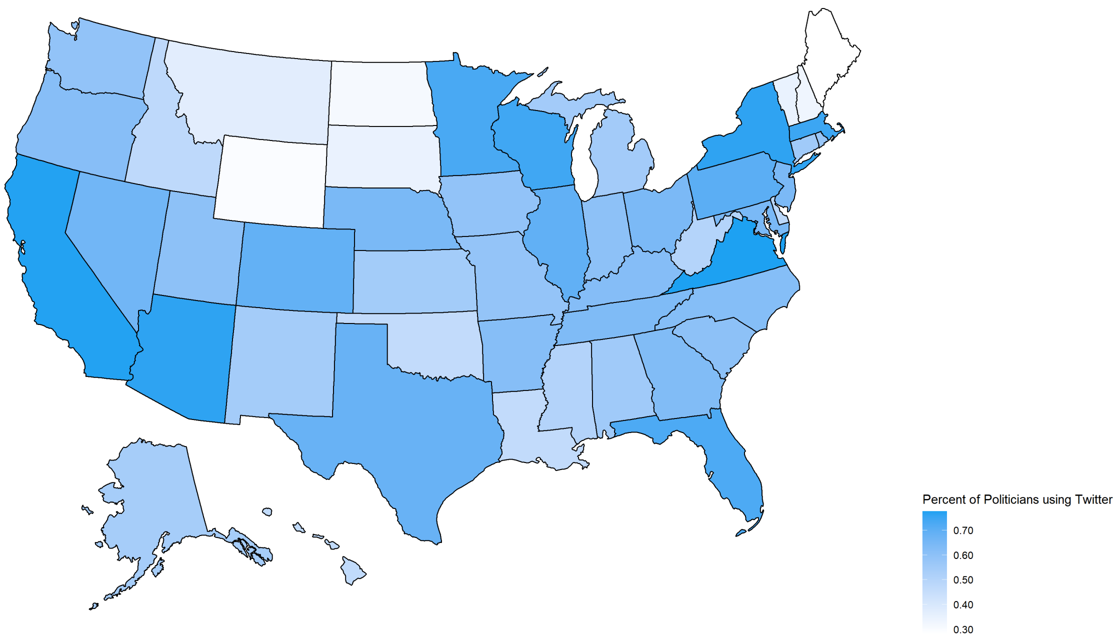
[26] Tweets that attack opponents have a mean (normalized) ratio of 0.377 as compared to a mean ratio of 0.369 for all other tweets. The t-value of the difference is 14.

small set of politicians, and less prevalent at lower levels of government. Opposition status and ideological extremism are the strongest covariates for affective polarization. Fourth, ideological polarization on social media is at least as high as polarization in congress speeches and reaches unprecedented heights in 2021. Fifth, upon observing jumps in audience engagement for attacks on political opponents, politicians strongly scale up further attacks.

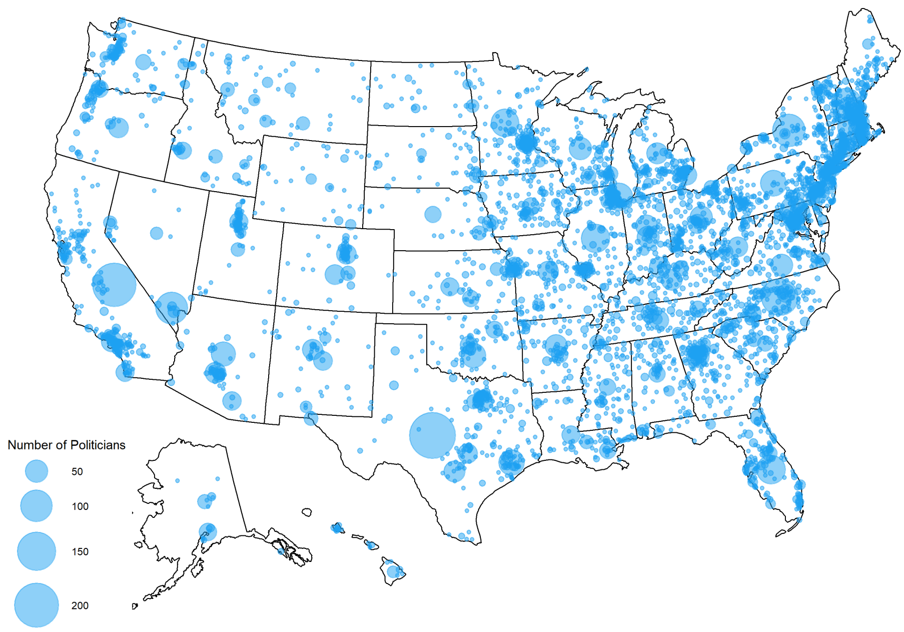
Overall, this paper presents substantive evidence that political cleavages and inter-party hostility among U.S. politicians are deeper than ever. This growing political antagonism has pervasive *political* costs, due to reduced efficacy of government (Hetherington and Rudolph, 2015) arising through political gridlock (Binder, 2004), obstructionism, delays in fiscal stabilizations (Mian, Sufi and Trebbi, 2014) but also leads to *economic* inefficiencies, for example through policy uncertainty (Baker et al. (2014), Baker et al. (2020)) and fosters *societal* cleavages by increasing the homophily of social groups (Iyengar, Sood and Lelkes (2012), Iyengar et al. (2019)).

Moreover, political communication is not living in a vacuum but shapes the political discourse as well as beliefs and behavior of citizens. Original tweets of members of congress only have attracted 3,6+ billion reactions on social media. Political framing affects the public opinion across a host of issues (see e.g., Druckman, Peterson and Slothuus (2013)). Existing evidence shows that social media posts of politicians (persistently) influence the perception of immigrants (Giavazzi et al., 2020) and predict hate crimes (Müller and Schwarz, 2020). The growing level of elite polarization is thus likely to diffuse and amplify the division of citizens along party membership.

Figure 2.1: Geographic Representation of Politicians on Twitter



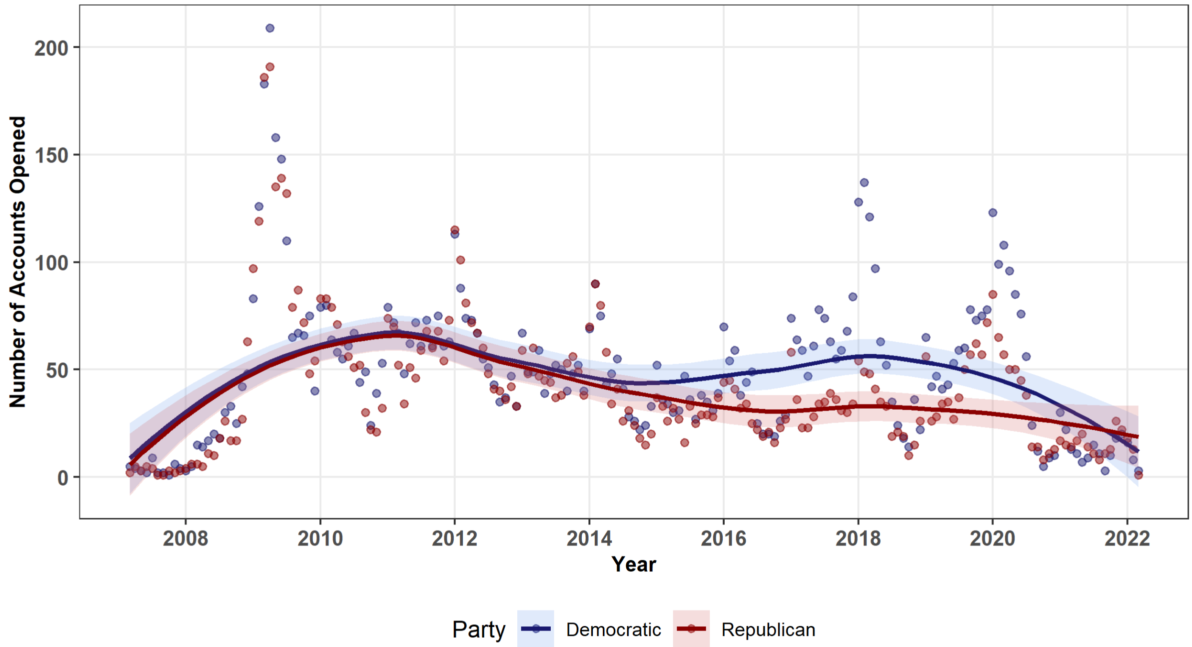
(a) Share of Politicians Using Twitter, by State, 2016-2020



(b) Geographic Distribution of Homecity of Politicians Using Twitter

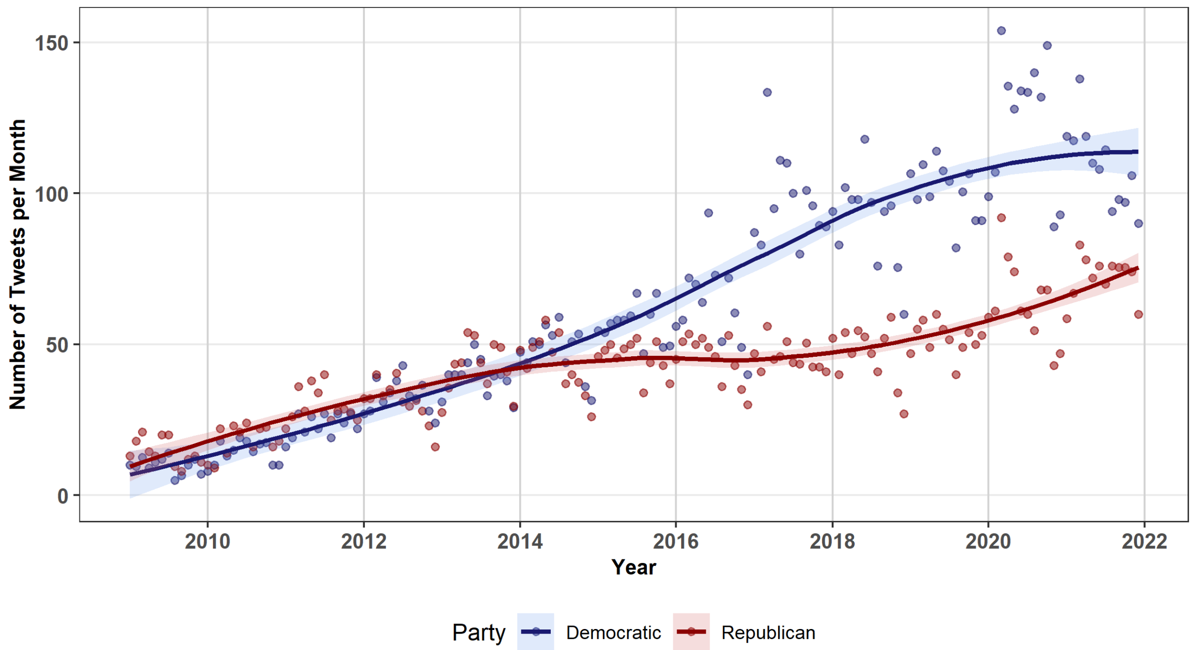
Notes. - Panel (a) maps the share of politicians, who have a Twitter account, in relation to the total number of candidates in each state averaged over the years 2016 - 2020. Panel (b) maps the geographical distributions of the home-city or self-declared city in Twitter profile of all 12,800 politicians with available data.

Figure 2.2: Twitter Adoption by Party



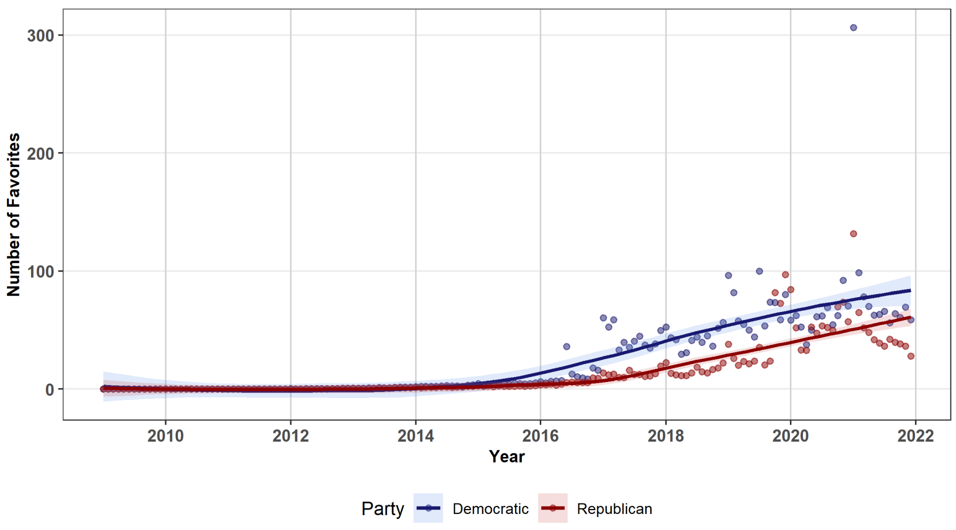
Notes. - The figure plots the monthly number of new Twitter accounts opened by party. Sample: Full Sample

Figure 2.3: Average Number of Tweets per Month for Median Legislator by Party

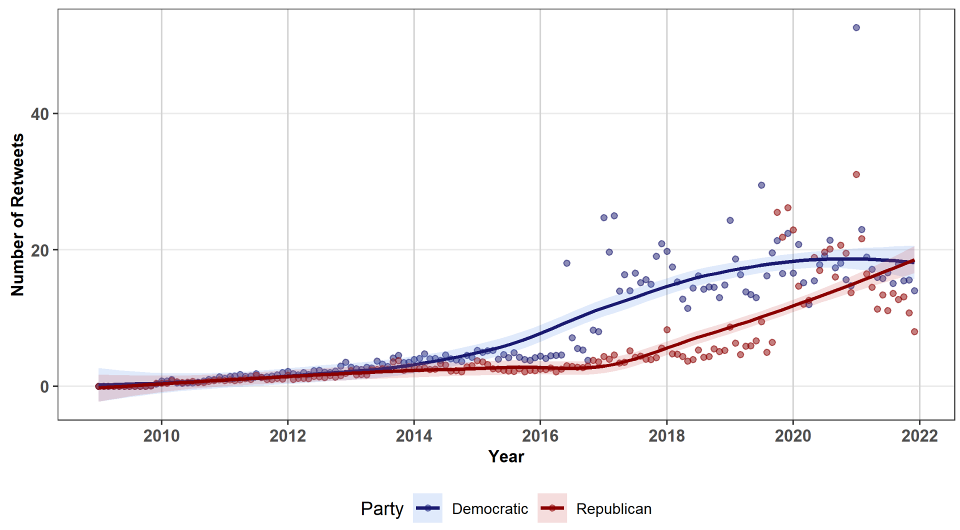


Notes. - The figure plots the average number of tweets per month for the median legislator, by party. For each month, I compute the average number of tweets by politician and display the value for the median legislator in each party. The numbers are based on tweets and retweets. Sample: Members of Congress.

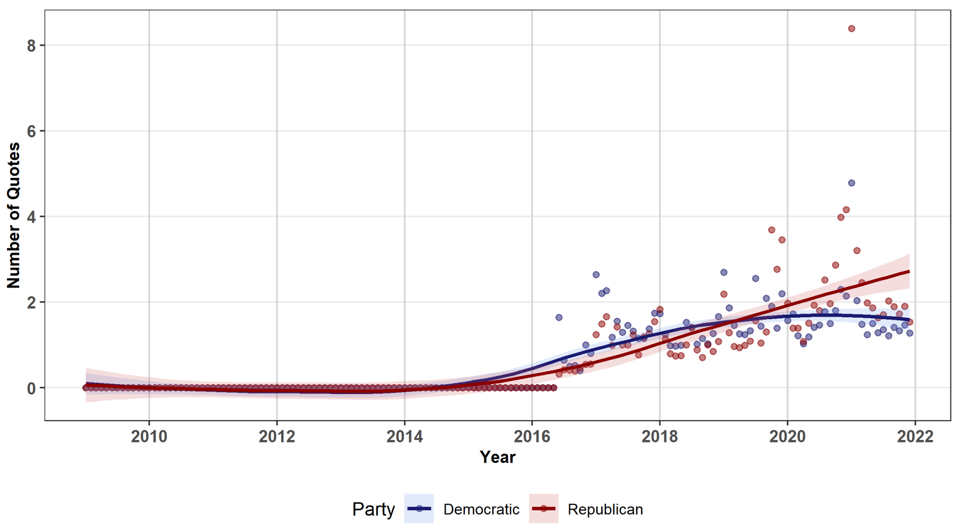
Figure 2.4: Audience Engagement for Average Tweet of Median Legislator by Party



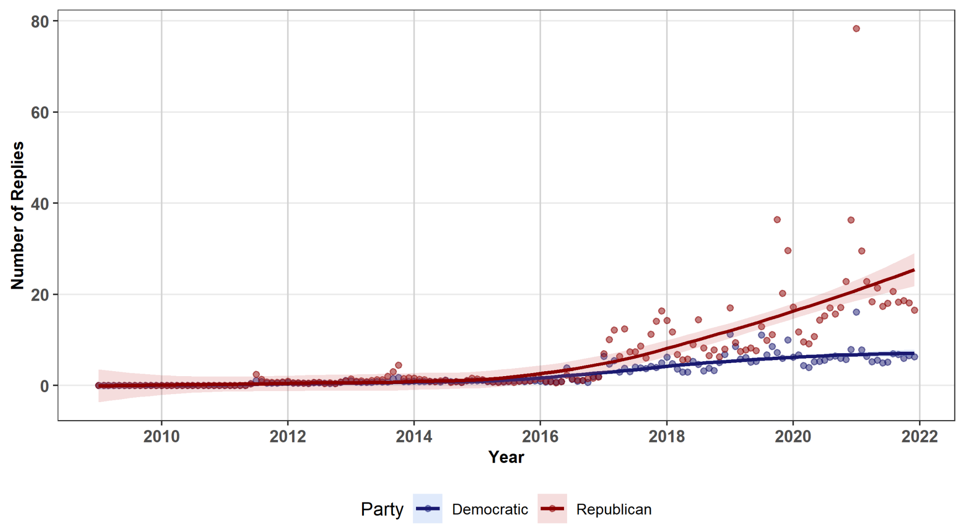
(a) Favorites



(b) Retweets



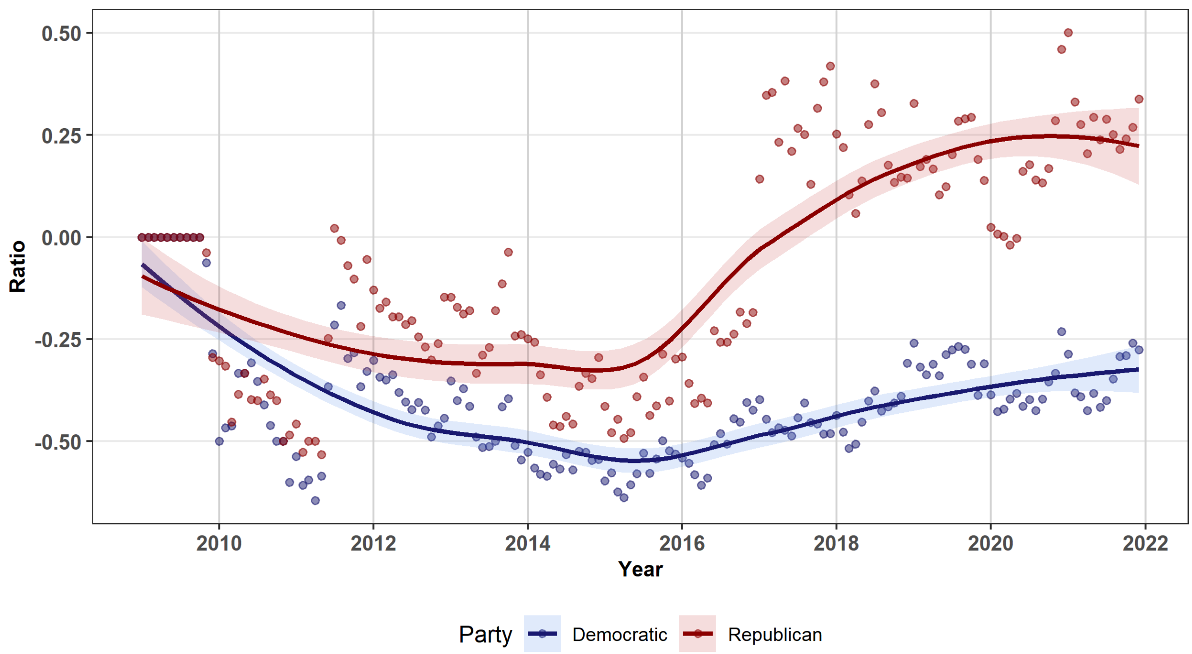
(c) Quotes



(d) Replies

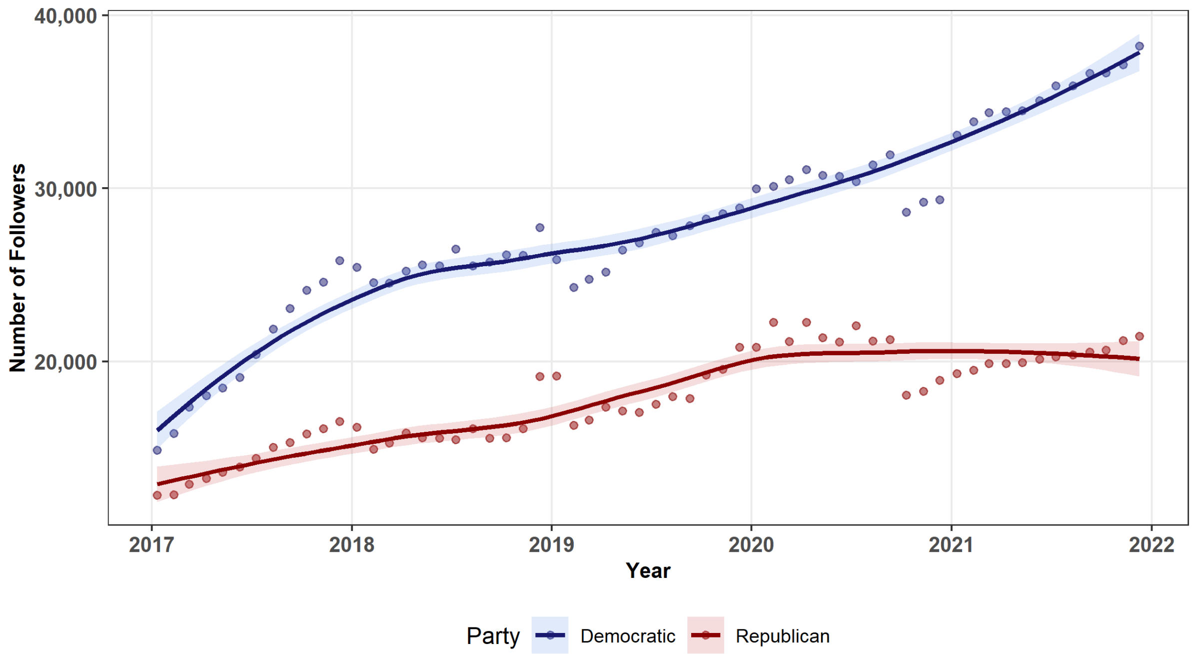
Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: Members of Congress.

Figure 2.5: Average "Ratio" per Tweet for Median Legislator by Party



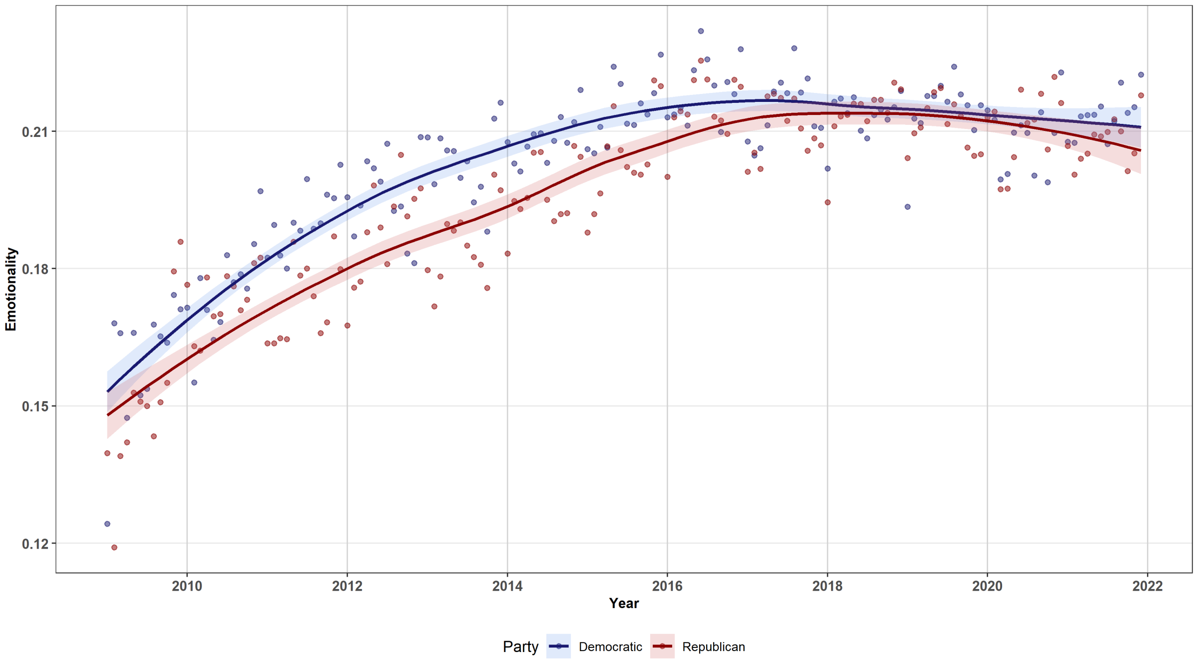
Notes. - The figure plots for each month the average "ratio" for the median legislator by party. For each tweet, I compute (replies + quotes - retweets)/(replies + quotes + retweets). I then compute the average "ratio" by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: Members of Congress.

Figure 2.6: Average Number of Followers per Month for Median Legislator by Party



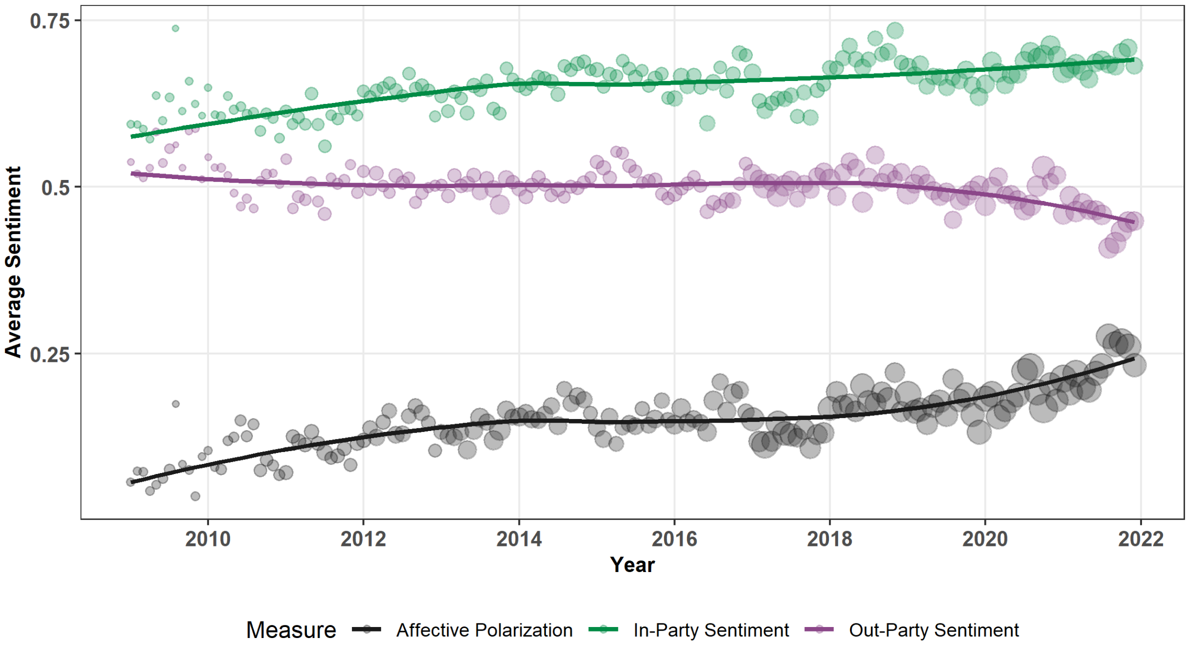
Notes. - The figure plots the average number of followers per month for the median legislator, by party. For each month, I compute the average number of followers by politician and display the value for the median legislator in each party. Sample: Members of Congress.

Figure 2.7: Average Emotionality per Month for Median Legislator by Party



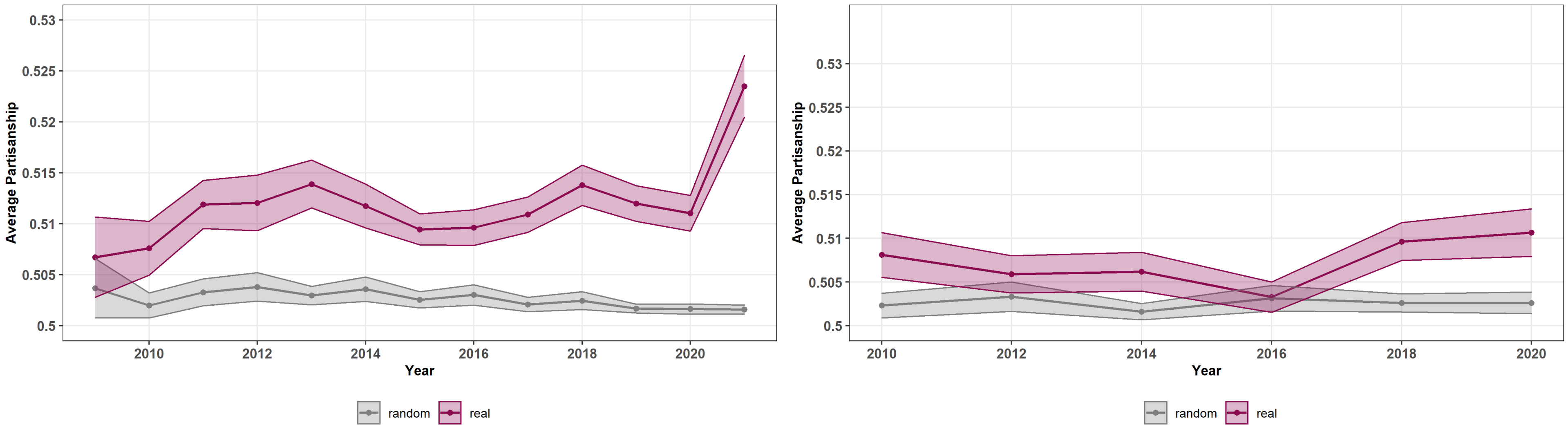
Notes. - The figure plots the average emotionality per month for the median legislator, by party. I compute emotionality as 1 - share of neutral words at the tweet level and then calculate the average monthly emotionality by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: Members of Congress.

Figure 2.8: Affective Polarization, In-Party Sentiment and Out-Party Sentiment, 2009 - 2021



Notes. - The figure plots the average in-party sentiment, out-party sentiment and affective polarization by month. For each month, I compute the average sentiment across all tweets that tag another politician or mention their name/ideology. In-party sentiment is computed from tweets by Democrats (Republicans) about Democrats (Republicans). Out-party sentiment is computed from tweets by Democrats (Republicans) about Republicans (Democrats). Affective polarization is the difference between in-party sentiment and out-party sentiment. The numbers are based on original tweets. Sample: Members of Congress.

Figure 2.9: Partisanship on Social Media, 2009 - 2021



(a) Sample: Members of Congress

(b) Sample: Congressional Candidates

Notes. - The figure plots the average partisanship from the penalized estimator $\hat{\pi}_t$. Panel A plots the yearly estimates for partisanship for members the 111th to 117th Congress. Panel B plots and estimates partisanship for congressional candidates. In each plot the "real" series is from actual data, while the "random" series assigns politicians' identity at random with the probability that a politician is Republican equal to the average share of Republican politicians in a session of Congress or the share of Republican vs Democratic candidates. The shaded areas are pointwise 95% - confidence intervals based on 100 subsamples of the data. For each subsample, I randomly draw 20 percent of all members of congress active on social media and compute the penalized estimate. The numbers are based on original tweets.

Table 2.1: Matching Rates by Level of Politics

	Votesmart	Twitter	Matching-Rate
	A. Presidential Elections		
Officeholders	5	5	100%
Candidates	1,386	191	14%
	B. Congressional Elections		
Officeholders	1,018	941	92%
Candidates	10,228	2,983	29%
	C. Gubernational Elections		
Officeholders	217	181	83%
Candidates	1,556	357	23%
	D. State Elections		
Officeholders	17,556	7,869	45%
Candidates	43,445	4,737	11%
	E. Local Politicians		
Officeholders	31,119	1,429	5%
Total	106,539	18,693	17,5%

Notes. - The table presents for each level of politics and type of politician the number of unique politicians observed in the Vote Smart data (Column 1), and the number of unique politicians for whom I have identified a Twitter account (Column 2). The matching rate is the ratio between the two.

Table 2.2: Representation of Politicians On Social Media Platforms

	A. Twitter						
	Vote Smart		No Twitter		Has Twitter		
Variable	N	Mean	N	Mean	N	Mean	Difference
Female	66200	0.264	49943	0.239	16257	0.339	0.10***
Age	16862	56.64	11218	56.79	5644	56.39	-0.04*
White	66149	0.900	50025	0.908	16124	0.879	-0.03***
Leg Experience	18659	10.8	9770	11.08	8889	10.45	-0.63***
Education (y)	34631	15.84	21952	15.74	12679	16.01	0.26***
Democratic	66845	0.500	50511	0.486	16334	0.544	0.06***
	B. Facebook						
	Vote Smart		No Facebook		Facebook		
Variable	N	Mean	N	Mean	N	Mean	Difference
Female	66200	0.264	46608	0.231	19592	0.342	0.11***
Age	16862	56.64	11869	56.61	4993	56.72	0.11
White	66149	0.900	46659	0.906	19490	0.887	-0.018***
Leg Experience	18659	10.8	10065	11.58	8594	9.842	-1.74***
Education (y)	34631	15.84	21271	15.81	13360	15.89	0.8***
Democratic	66845	0.500	47119	0.489	19726	0.527	0.038***

Notes. - Female presents an indicator equal to 1 if the politicians is a female. Age presents the age of a politician in 2020. White is an indicator equal to 1 if a name-algorithm has classified the first name of a politician as associated with a white race in Census Statistics. Legislative Experience is calculated from the Vote Smart Data. Information on the highest educational degree for a politician is obtained from Vote Smart and Ballotpedia and subsequently transformed into years of education. Democratic is a dummy equal to 1 if a politician identifies with the Democratic party and 0 if she identifies with the Republican party. Independents are excluded. Statistical significance markers: * p<0.1; ** p<0.05; *** p<0.01

Table 2.3: Effect of Twitter Adoption on Electoral Success (OLS-Coefficients)

	Dependent variable: Indicator(Election Winner)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Has-Twitter	0.060***	0.060***	0.057***	0.052***	0.043***	0.070***	0.065***	0.015**
	(0.006)	(0.007)	(0.006)	(0.005)	(0.007)	(0.014)	(0.006)	(0.007)
Mean Dep Var	0.382	0.382	0.382	0.382	0.518	0.375	0.424	0.651
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	No	No	No	Yes	No	No	No	No
Sample	Full	Full	Full	Full	Gen. Elect	>2015	Dem|Rep	Twitter Pol
Observations	126,995	126,995	125,822	125,822	92,879	54,423	113,255	42,387
R^2	0.875	0.875	0.878	0.902	0.904	0.931	0.873	0.818

Notes. - The table reports the OLS estimates from specification 2.1. The dependent variable is an indicator that is equal to 1 if a politician c from state s running for office o in an election-year e has been winning the election, and 0 otherwise. The unit of observation is at the politician-election level. The main explanatory variables is a dummy whether a politician has opened a Twitter account before the election. All regressions include full sets of election-year, candidate, state and office fixed effects. Column (5) includes only general election results; Column (6) considers only elections from 2016 onwards; Column (7) selects only politicians associated with the Democratic or Republican party; Column(8) uses only politicians linked to a Twitter account. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 2.4: Effect of Twitter Activity on Electoral Success (OLS-Coefficients)

	Dependent variable: Indicator(Election Winner)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ln (1 + N-Tweets)	0.034***	0.034***	0.034***	0.032***	0.020***	0.039***	0.036***	0.033***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Mean Dep Var	0.382	0.382	0.382	0.382	0.518	0.375	0.424	0.651
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Race FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	No	No	No	Yes	No	No	No	No
Sample	Full	Full	Full	Full	Gen. Elect	>2015	Dem|Rep	Twitter Pol
Observations	126,995	126,995	125,822	125,822	92,879	54,423	113,255	42,444
R^2	0.878	0.878	0.882	0.905	0.905	0.935	0.877	0.828

Notes. - The table reports the OLS estimates from specification 2.1. The dependent variable is an indicator that is equal to 1 if a politician c from state s running for office o in an election-year e has been winning the election, and 0 otherwise. The unit of observation is at the politician-election level. The main explanatory variables is a the natural logarithm of the number of tweets (+1) before the election. All regressions include full sets of election-year, candidate, state and office fixed effects. Column (5) includes only general election results; Column (6) considers only elections from 2016 onwards; Column (7) selects only politicians associated with the Democratic or Republican party; Column(8) uses only politicians linked to a Twitter account. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 2.5: Determinants of Affective Polarization

	Dependent variable: Affective Polarization								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(DWnom1)	0.102^{***}								0.127^{***}
	(0.019)								(0.020)
Republican		-0.032^{*}							-0.037^{*}
		(0.019)							(0.022)
Female			0.044^{**}						0.040^{**}
			(0.018)						(0.018)
Education				-0.034^{*}					-0.032
				(0.020)					(0.020)
Leg Exp					0.052^{**}	0.009			0.049^{**}
					(0.023)	(0.726)			(0.022)
Majority							-0.075^{***}	-0.054^{***}	-0.070^{***}
							(0.015)	(0.016)	(0.015)
Chamber-Year Fe	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Candidate Fe	No	No	No	No	No	Yes	No	Yes	Yes
Observations	5,052	5,071	5,071	4,974	5,059	5,059	5,071	5,071	4,955
R^2	0.053	0.044	0.045	0.043	0.045	0.320	0.048	0.322	0.066

Notes. - Each column shows a OLS regression of the standardized affective polarization score in a given politician-year on the respective covariate. The sample is composed of all original tweets by active Democratic or Republican members of Congress between 2009 and 2021. All specifications include chamber-year fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 2.6: Effect of Jump in Likes on Future Attacks on Opponents

	Dependent variable:			
	N-Attacks	1(Attack)	N-Tweets	N-Attacks/N-Tweets
	(1)	(2)	(3)	(4)
	Panel A. Period T +- 3			
Post-Like	0.120***	0.128***	0.026***	0.099***
	(0.006)	(0.006)	(0.006)	(0.005)
Observations	123,394	123,394	123,394	123,394
	Panel B. Period T +- 5			
Post-Like	0.095***	0.119***	0.034***	0.093***
	(0.006)	(0.005)	(0.005)	(0.005)
Observations	157,512	157,512	157,512	157,512
	Panel C. Period T +- 7			
Post-Like	0.079***	0.117***	0.033***	0.093***
	(0.007)	(0.005)	(0.005)	(0.004)
Observations	184,653	184,653	184,653	184,653
	Panel D. Period T +- 9			
Post-Like	0.078***	0.117***	0.035***	0.092***
	(0.006)	(0.004)	(0.005)	(0.004)
Observations	203,201	203,201	203,201	203,201
Pol FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	Yes	Yes
Calender-Day FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full
R^2	0.166	0.166	0.327	0.122

Notes. - The table reports the OLS estimates from specification 2.6. The unit of observation is at the politician-day level. Column (1) uses a dependent variable the number of attacks (per politician-day), Column (2) uses an indicator whether an attack occured, Column (3) uses the number of tweets and Column (4) uses the number of attacks divided by the number of tweets. The explanatory variables is an indicator that takes the value of 1 in the t days after a politician attacked an opponent and this attack was 'liked' by 50% more than the two-week rolling average of the politician, and 0 in the t days before this attack. Panel (a.) - Panel (d.) present time-windows for $t \in (3, 5, 7, 9)$ days before and after the attack. All regressions include full sets of candidate, state, office, calender-year and -day fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 2.7: Effect of Jump in Backlash on Future Attacks on Opponents

	Dependent variable:			
	N-Attacks	1(Attack)	N-Tweets	N-Attacks/N-Tweets
	(1)	(2)	(3)	(4)
	Panel A. Period T +- 3			
Post-Backlash	0.077***	0.100***	0.038***	0.069***
	(0.008)	(0.006)	(0.007)	(0.006)
Observations	98,294	98,294	98,294	98,294
	Panel B. Period T +- 5			
Post-Backlash	0.078***	0.096***	0.040***	0.069***
	(0.007)	(0.005)	(0.006)	(0.005)
Observations	130,947	130,947	130,947	130,947
	Panel C. Period T +- 7			
Post-Backlash	0.072***	0.095***	0.039***	0.066***
	(0.006)	(0.005)	(0.005)	(0.005)
Observations	159,362	159,362	159,362	159,362
	Panel D. Period T +- 9			
Post-Backlash	0.073***	0.096***	0.042***	0.068***
	(0.006)	(0.005)	(0.006)	(0.004)
Observations	179,954	179,954	179,954	179,954
Pol FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	Yes	Yes
Calender-Day FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full
R^2	0.181	0.208	0.318	0.140

Notes. - The table reports the OLS estimates from specification 2.6. The unit of observation is at the politician-day level. Column (1) uses a dependent variable the number of attacks (per politician-day), Column (2) uses an indicator whether an attack occured, Column (3) uses the number of tweets and Column (4) uses the number of attacks divided by the number of tweets. The explanatory variables is an indicator that takes the value of 1 in the t days after a politician attacked an opponent and this attack was 'ratioed' by 50% more than the two-week rolling average of the politician, and 0 in the t days before this attack. Panel (a.) - Panel (d.) present time-windows for $t \in (3,5,7,9)$ days before and after the attack. All regressions include full sets of candidate, state, office, calender-year and -day fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Chapter 3

Bad News and Political Deception - Evidence from Mass Shootings[1]

Abstract

This paper investigates whether politicians attempt to manipulate public perceptions when exposed to “bad news”. I treat mass shootings as bad news for politicians opposing gun control and combine shootings in the U.S. with data on political communication on Twitter, 2010 - 2020. Exploiting the random timing of mass shootings and applying quantitative text-analysis and deep-learning techniques to the twitter-data, I demonstrate that pro-politicians try to distract the public from the mass shooting and provide misleading information about the causes of mass shootings. The results are consistent with a conceptual framework in which politicians who face bad news communicate strategically to (i) prevent belief updating or (ii) distort the belief formation process of voters towards supporting policies other than gun-control.

[1] I am deeply grateful to Paolo Masella, James Snyder and Enrico Cantoni for their guidance and support throughout this project. This paper profited also greatly from advice of Ruben Durante, Daniela Iorio, Massimo Morelli, Vincenzo Scrutinio, Tommaso Sonno. Moreover, I wish to thank seminar participants at the University of Bologna for their comments and suggestions.

"Facts need testimony to be remembered and trustworthy witnesses to be established in order to find a secure dwelling place in the domain of human affairs."

Hanna Arendt, Lying in Politics, 1972

3.1 Introduction

What causes politicians to mislead the general public? Political communication diffuses widely through the media (e.g., Martin and Yurukoglu (2017)) and the rise of social media gives politicians direct access to an ever-increasing audience. While this connection facilitates unmediated conversation between voters and politicians that can improve responsiveness and representation, it also can be misused by politicians to manipulate public perceptions. For a long time, this misleading political rhetoric is attributed a crucial role in the emergence of misperceptions (e.g., Kuklinski et al. (2000)), a concern that is aggravated absent the vigilant gaze of media gatekeepers. Despite this peril, evidence on whether and, more importantly, the drivers of politicians' attempts to manipulate the public discourse is scarce.

In this paper, I investigate whether politicians disseminate distortive messages when exposed to bad news.[2] Specifically, I study the diffusion of misleading communication from pro-gun politicians in the aftermath of mass shootings. Mass shootings highlight the issue of gun policy, on which politicians that oppose gun-control (pro-gun) have a relative disadvantage.[3] Using data on political communication on Twitter in conjunction with mass shootings in the U.S. between 2010 and 2020, I rely on the plausibly exogenous timing of mass shootings to test whether pro-gun politicians spread distortive messages thereafter.

I document empirically that, after mass shootings, pro-gun politicians use two types of distortive communication: distraction and deception. *Distraction* constitutes a politician that uses her agenda setting power to deflect attention away from the original news, for instance by speaking about other newsworthy events to crowd-out the information provided by the bad-news event. A distracted citizen would not update his beliefs and remain with the prior.[4] *Deception* is defined as in Grossman and Helpman (2019) and refers to a misrepresenta-

[2] I provide a conceptual framework in Appendix A.3.3. Examples of bad news include reports about political malfeasance or poor performance in office. A rich empirical literature has documented the political costs of enhanced voter information (see for example Besley and Burgess (2002), Ferraz and Finan (2008), Snyder Jr and Strömberg (2010), Garz and Sörensen (2017) or Daniele, Galletta and Geys (2020)).

[3] Supporting evidence for this argument comes from Yousaf (2021), who finds that mass shootings causally decrease the vote share of Republicans (1.7 p.p.), which is most often congruent with being pro-gun.

[4] Take the following example for distraction: On September 1^{st} 2019, Donald Trump tweeted in the context of Hurrican Dorian that "In addition to Florida - South Carolina, North Carolina, Georgia, and Alabama, will most likely be hit (much) harder than anticipated." The tweet attracted wide reception in the news media and entailed the Hurrican Dorian-Alabama controversy. It also appeared the morning after the Midland-Odessa shooting, the

tion of the state of the world, such as the provision of misleading explanations for the news event. Rather than preventing the belief updating, this strategy aims at changing voters factual information to trigger support for alternative policies.[5]

This paper makes progress towards a better understanding of the supply of distortive political communication along two dimensions: First, I develop novel text-based tools for the systematic measurement of distraction and deception. The tools can be easily applied in other contexts in social science research or by media organizations to spot check-worthy claims among the myriad messages sent by politicians. Second, I apply and combine these text-tools with variation in the occurrence of mass shootings to provide the first direct evidence on the drivers of politicians to distort the public discourse.

To quantify *distraction*, I identify directly and indirectly politicians' attempts at shifting the focus of the public discourse from one topic (e.g., a mass shooting) to an unrelated topic (e.g., a hurricane). My *indirect* measure separates tweets that speak about mass shootings from non-shooting related tweets. To classify the tweets, I apply a simple keyword-based methodology, as in Demszky et al. (2019), as well as unsupervised topicmodelling. The topicmodel additionally distinguishes the non-shooting related tweets, into tweets with higher distractive value (e.g., political topics) from tweets that would unlikely deflect citizens (e.g., birthday wishes). Thus, I determine how much politicians cover other politically meaningful topics as opposed to mass shootings. I also compile a *direct* measure of distraction defining the set of distractive topics as the national news in the United States (excluding mass shooting news) and compute the textual similarity between the news abstracts and tweets.

I measure *deception* by detecting statements that misrepresent the state of the world. Specifically, I classify as deceptive, statements about the causes of mass shootings that are highly representative/predictive of a Fake News (FN) source relative to a Real News (RN) source. I collect news articles about mass shootings from the ten major daily newspapers in the U.S. (RN), ensuring a broad ideological spectrum, as well as from Breitbart and InfoWars (FN). I then develop and employ two neural networks, the state-of-the-art tool in Compute Science for text analysis, for the prediction of misrepresentative statements. In a first step, I train a neural network model to recognize statements about the causes of mass shootings. In the

third major mass killing in August 2019 only. Donald Trump received an "A" rating from the NRA in the 2016 and 2020 elections. Source: https://time.com/5671606/trump-hurricane-dorian-alabama/.

[5]Take the following example for deception: During the 2013 New South Wales bushfires, Australia's Prime Minister Tony Abbott said that "[...] these fires are certainly not a function of climate change. They are a function of life in Australia." Prior to these fires, Abbott had questioned the science of climate change and rejected Carbon Pollution Reduction Schemes. Source: https://www.theguardian.com/world/planet-oz/2013/oct/23/climate-change-tony-abbott-australia-bushfire-science.

second step, I use these statements as inputs for the main neural network model that learns to distinguish between statements representative of a FN-source relative to a RN-source. The algorithm uses for its prediction of misrepresentation only statements that are either exclusive to FN-outlets or heavily *over-represented* in the FN-sources.

My analysis relies on a dataset that comprises nearly half a million tweets from U.S. politicians in the days right before and after mass shootings. These novel data combine information on the timing and the location of highly visible mass shootings in the U.S. between 2010 and 2020 with data on political communication from Twitter. For this purpose, I connect the universe of politicians (elected officials and candidates) at the state and federal level between 2010 and 2020 with their respective Twitter accounts. I determine a politicians stance on gun control using ratings from gun interest groups (e.g., the NRA).

I exploit the random timing of these mass shootings and use event studies to identify the effect of negative publicity on distractive - and deceptive communication from exposed politicians. In particular, I compare the outcome, distraction or deception, of a pro-gun politician right before versus right after the mass shooting.

The main identification assumption is the absence of pre-trends in the outcomes. This includes, for example, no pre-emptive distortive communication from pro-gun politicians before the bad news shock, that is anticipatory tweets that mislead on the causes of mass shootings and that affect the perception of citizens or the response of media to the events. The event studies do not show any pre-trend in distractive or deceptive communication before mass shootings, supporting this assumption. Additionally, I estimate the same regression on the sample of anti-gun politicians and test for differences. In the case of distraction, using only tweets from pro-gun politicians could raise concerns regarding the interpretation of the results as speaking about a different topic might simply reflect a lack of media interest in the mass shooting. The comparison of the event studies reveals strong differences in communication between pro-gun and anti-gun politicians after the occurrence of the shooting and precludes this concern.

My analysis proceeds in four steps. Initially, I study politicians' tweeting behavior in response to the shootings. Then, I investigate the dissemination of my main outcomes, distraction and deception. I conclude by testing the strength politicians' reaction using variation in the type and timing of the mass shooting.

I start by exploring the Twitter activity of politicians in the 5 days before and after a mass shooting. Pro-gun politicians slightly decrease the number of tweets sent after mass shootings.

I observe a similar reduction in the number of tweets sent by anti-gun politicians. This suggests that, instead of staying quiet, pro-gun politicians cope with the negative consequences of mass shootings through altering the *content* of their messages.

The second part of my results focuses on distraction. I find that pro-gun politicians speak about the mass shootings, but significantly less compared to anti-gun politicians. In other words, pro-gun politicians post-event tweets disproportionately focus on non-shooting related topics. Tweets from pro-gun politician are 5.8 p.p. more likely to speak about other political topics after the shooting (or about 7% compared to baseline), an effect that arises promptly and persists for at least 5 days after the shooting.[6] Second, turning to the direct measure of distraction, the results indicate that in the three days following the mass shooting, the tweets of the pro-gun politicians get closer to the national news. This pattern is specific to pro-gun politicians after a shooting; it is not detectable for pro-gun politicians before a shooting or for anti-gun politicians either before or after a shooting. The estimated coefficients imply that the news similarity of pro-gun tweets after the shootings increases by 32% relative to the baseline.

The third part of the results shifts the focus to deception. Using the predictions from the deep-learning algorithm on the Twitter data, I find that pro-gun politicians spread falsehoods immediately upon the arrival of the news of a mass shooting, but never before the event. The effect tends to be short-lived and persists for two days after the mass shooting. The diffusion of misleading narratives from pro-gun politicians is significantly stronger than among anti-gun politicians. The effect reflects an increase in the spread of misleading statements about mass shootings of 54%. In contrast, a large part of the provision of accurate information is attributable to anti-gun politicians. [7]

Finally, I conduct a series of heterogeneity tests that accentuate the highly strategic nature of distortive messages. To begin with, I find that my results are stronger for shootings with a high number of victims as well as shootings that occur close to elections. This suggests that the extent of distortive messaging is responsive to the size of the potential costs of bad news. Moreover, I show that incumbents draw more often to deception than candidates. Given that incumbents are at the center of political responsibility for the shootings, it is harder for them to duck away and not speak about the shootings. Consequently, incumbents draw more

[6] Note, that the baseline category "distraction" is extremely high as before the event nearly 100% of the tweets are not about gun policy.

[7] Overall, the share of misleading tweets compared to the total number of statements about the causes of mass shootings is more than 4 times higher for pro-gun politicians (40.8%) relative to anti-gun politicians (9.7%). The number of "misleading" tweets from anti-gun politicians is not zero as the algorithm captures as well causes exclusively as well as causes that are over-represented in FN-sources. A small share of the ladder explanations are discussed in RN-sources and also in tweets from anti-gun politicians.

often on deception. Last, I observe stronger effects for distraction when mass-shootings and natural disasters co-occur. In opposition, I find weaker effects for deception in this situation. This result presents prima-facie evidence how politicians decide on their rhetorical strategy: politicians respond to sudden shifts in the returns of the respective strategy.

Collectively, my results demonstrate that politicians react to policy shocks in which they have a relative disadvantage by *(i)* steering the public discussion away from the occasion and *(ii)* employing manipulative rhetoric to mislead the general public.

These findings have important implications for political accountability and question the welfare effects of social media. On the one hand, political attempts to distort the public discourse can directly alter voters' belief formation. On the other hand, politicians' rhetoric on social media can change *how* traditional media discuss upcoming news and change the trajectory of *which* news are reported. Recent work by Cagé, Hervé and Mazoyer (2020) shows that newsroom production decisions of mainstream media are influenced by the popularity of a story on social media, particularly from users with high centrality. Thus, the strategic use of distortive messages of politicians' can curb medias' ability to inform citizens, sets limits to political accountability and raises doubts on the overall welfare provided by social media.

Related Literature. First and foremost, this paper contributes to previous work on the diffusion of political misinformation. Descriptive evidence shows that factually false information spreads "farther, faster, deeper, and more broadly than the truth" (Vosoughi, Roy and Aral, 2018). It is estimated that before the 2016 U.S. election one in four Americans visited a fake news website and about three fake news stories have been read per American voter (Guess, Nyhan and Reifler (2018); Allcott and Gentzkow (2017)). More worryingly, experimental evidence demonstrates that exposure to misleading statements has long-lasting effects on voting intentions, even after a fact-checking intervention (Barrera et al., 2020). Last, Grossman and Helpman (2019) study theoretically the role use of misinformation in electoral competition. My contribution to this literature is threefold. First, I develop a novel *content*-based measure for misleading information. Previous attempts to track fake news focused on the spread of *single* fact-checked articles from fake news websites (Allcott and Gentzkow, 2017) or references to fact-checkers in the comments (Vosoughi, Roy and Aral, 2018), an information that is not necessarily accessible for researchers (e.g., on Twitter). Therefore, these attempts have a limited capacity to detect misleading information on a broader scale and require expert judgement on whether an entire article is factually false. Instead, my measure works on the mild judgement whether a text makes a factual statement and exploits differences in the representation of those facts across RN-sources and FN-sources. Although my measure is tailored to detect misleading content about the causes of mass shootings, it is scalable and can be broadly applied to other contexts. Second, I propose a rationale for when politicians provide factually false accounts of reality. Last, I apply my text-based measures for the first empirical evidence for the diffusion of misleading information from political elites in general and specifically as a strategic response to bad news.

Second, this research informs a large literature in political economy on the importance of information provision to discipline elected officials (Besley and Burgess (2002), Ferraz and Finan (2008), Snyder Jr and Strömberg (2010)). The crucial pre-condition for political accountability is citizens' attention. A body of work has documented that politicians exploit periods when media and citizens are distracted from politics to vote with interest groups (Kaplan, Spenkuch and Yuan, 2018) or to release controversial information (Djourelova and Durante, 2019) to avoid critical coverage on their actions. I contribute to this literature by *(i)* shifting the focus to the complementary situation, that is when politicians are already subject to negative press and by *(ii)* measuring and assessing the reaction in terms of a different strategic tool: distractive communication. My results demonstrate how politicians seek to actively diminish media/citizen attention on a critical issue through deflective messages.[8]

Third, I contribute to a nascent literature which investigates the "polarization of reality" (Alesina, Miano and Stantcheva, 2020). This literature shows that Republicans and Democrats view the same reality through different lenses (heterogeneous priors) and as a result may hold different views about policies (e.g. Alesina, Miano and Stantcheva (2018)). In line with this interpretation, a literature on the effects of mass shootings on political outcomes (Luca, Malhotra and Poliquin (2020), Yousaf (2021)) has demonstrated that despite observing the *same* signal, a mass shooting, Republican and Democratic voters update their policy views in opposite directions in the aftermath. My results offer an alternative explanation for the observed divergence of (gun) policy preferences: exposure to a *different* signals about the causes and thereby implied best-policies of mass shootings from partisan actors.

The remainder of the paper is organized as follows. Section 3.2 informs about the data and the data-selection process, while Section 3.3 lays out the empirical specifications. Section

[8]In a similar vein, Lewandowsky, Jetter and Ecker (2020) show that Donald Trump reacts to critical reporting (media coverage of the Mueller investigation) by tweeting about unrelated issues. This paper differs in three important aspects: First, I present *systematic* evidence by investigating the Twitter handles of a large set of politicians. Second, I construct broadly applicable measures of distraction instead of relying on a narrow, ex-ante defined set of keywords to identify potentially distractive tweets. Third, mass shootings are arguably a more exogenous source of variation for a bad news shock than New York Times coverage of the Mueller investigation, that could be launched strategically.

3.4 presents the measurement and results for distraction and Section 3.5 the equivalent for deception. Section 3.6 displays some tests for heterogeneity and Section 3.7 concludes.

3.2 Data

The analysis combines five sources of data. I gather information on highly visible mass shootings from Wikipedia for the period I study: 2010 - 2020. To investigate political communication around the shootings, I collect comprehensive information on all politicians running for state-level and federal offices. I connect those politicians to their Twitter accounts and gather the universe of their tweets. I determine the gun-stance of the politicians through gun-interest group ratings. To measure if politicians are distracting from mass shootings, I use the content of the tweets and additionally data on national news for the major U.S. broadcast TV networks. To measure if politicians are deceiving, I collect news-articles covering mass shootings from the major U.S. daily newspapers as well as from Breitbart and Infowars.

3.2.1 Mass Shootings

Data on mass shootings come from Wikipedias' "List of mass shootings in the United States".[9] The list includes the most notable incidents of firearm-related violence involving several victims. All of the shootings on Wikipedias' list have an their own Wikipedia article and thus received "significant coverage in reliable sources that are independent of the subject" (e.g., newspaper, radio, or TV).[10] The Wikipedia article provides information regarding the location of the shooting, the date, the number of deaths and injured, and at times additional information on the context, e.g. the shooters and their motives. In total, Wikipedia lists 125 mass shootings for 2010 - 2020.

3.2.2 Political Communication

Data on politicians come from Vote Smart, a non-profit, non-partisan research organization. The Vote Smart data cover the (near) universe of politicians, who ran for or held state or federal offices, 2010 - 2020.[11]

To connect these politicians to their Twitter account(s) I rely on information (i.e., linkages to Twitter accounts) from Vote Smart and Ballotpedia. For all missing observations, I use the

[9]Accessed: January, 13 2021.

[10]As stated in Wikipedias' notability guideline: https://en.wikipedia.org/wiki/Wikipedia:Notability. Accessed: April, 7 2021

[11]Vote Smart counts 62,239 politicians in that period.

Twitter API to search for accounts that match the politicians' name. For each politician, I retrieve the 30 best results, ordered by Twitter, and search in the self-description on Twitter for a) politics-related keywords (e.g., "senator", "representative") and b) politician-specific keywords such as the office, district or state of the respective politician. For queries that satisfy both of these criteria, I keep the account with the highest number of followers.
The resulting database encompasses Twitter accounts for 16,126 politicians, of which more than 70% are collected by Vote Smart or Ballotpedia, and 51 million Twitter statuses.[12]

3.2.3 Gun Stance

I use two sources to determine each politicians' gun stance. The first source is Vote Smarts' interest group rating: these evaluations are provided from all special interest groups to Vote Smart for 2010 - 2020 and solely distributed through the platform. I use all ratings from all 33 interest groups that appear under the issue of gun-control. The interest groups rate politicians on a scale from 0 to 100. I express all ratings in terms of how strong a politician is *against* gun control that is a higher rating corresponds to a stronger opposition of gun control measures.[13]
The second source are NRA grades and endorsements collected from Everytown for Gun Safety, available for 2010 - 2016.[14] The NRA grades range from A+ to F-. I classify a politician as *pro-gun* if her rating of *opposing* gun control is above 50, the NRA grade is above B+ or the candidate received an NRA endorsement. I classify a politician as *anti-gun* if her rating of *opposing* gun control is below 50, the NRA grade is B+ or lower or the candidate did not receive an NRA endorsement.[15]

3.2.4 News Articles about Mass Shootings

I collect newspaper articles about mass shootings for the years 2010 - 2019 from the following two sources.

Real News. I use the top 10 newspapers in the United States by average weekday circulation in 2019 and CNN as my sources for real news.[16] I then search in the ProQuest Newslibrary for

[12]The resulting dataset is not representative for all US-politicians on Twitter. It is conceivable that for example unsuccessful candidates change their self-description after an election or retired elected officials delete their accounts.

[13]For instance, suppose a politicians receives two ratings: a rating of 100 from an interest group that is against gun control (e.g. the NRA), and a rating of 20 from an interest group that is in favor of gun control. The total ranking of this politician is then 90 ($\frac{100+(100-20)}{2}$).

[14]The data can be found here: https://www.everytown.org/nra-grades-archive/.

[15]Politicians with ratings lower than B+ never receive an NRA endorsement.

[16]The news sources are "Washington Post", "New York Times", "Los Angeles Times", "Wall Street Journal", "USA Today", "Chicago Tribune", "Boston Globe", "Newsday", "New York Post", "Star Tribune", "CNN Wire Service". The

news articles using the location-specific keywords,e.g. "Thousand Oaks Shooting" and articles about mass shootings in general.

Fake News. I chose the Breitbart News Network and InfoWars as sources for fake news. Both outlets are involved in numerous cases of publishing intentionally misleading news stories. InfoWars is even subject to a lawsuit for false claims about mass shootings. Both outlets have been used in the computer science literature as sources for text-based fake news classification (Zellers et al., 2019). The comparative advantage of Breitbart and InfoWars with respect to other potential fake news outlets is that both were active throughout the entire sample period.[17] Any differential reporting about mass shootings can therefore be attributed to the editorial choices of the news outlets and are not driven by data availability. I have scraped all available articles for both websites.

To ensure that the respective news articles are indeed covering mass shootings, I restrict my sample to news articles whose title or more than 50% of the paragraphs contain the words from the dictionary developed in Section (3.4.2). This amounts to 8,675 articles from Real News sources and 5,562 articles from Fake News sources.

3.2.5 National News

Data on news coverage of other noteworthy events comes from the Vanderbilt Television News Archive (VTNA).[18] These VTNA data contain daily information on the content and duration of every news story aired in the national evening news (henceforth, national news) from ABC, CBS, NBC, CNN and Fox News. Following Kaplan, Spenkuch and Yuan (2018), I restrict the sample to the newscasts of the "big three", ABC, CBS, NBC. The average number of news stories per day and program is 9. The average duration of a news story is around 2 minutes.

3.2.6 Sample Construction

The construction of the final sample proceeds in four steps: First, I restrict the sample of mass shootings to *public* shootings with one perpetrator in one location (e.g., Luca, Malhotra and Poliquin (2020)). This excludes, for instance, incidents of familicides, gang violence, or workplace shootings, that are more likely to bring about debates about social policy and crime rather than gun policy (Krouse and Richardson, 2015). I refrain from excluding shootings with

list has been retrieved from: https://web.archive.org/web/20190722203322/https://www.cision.com/us/2019/01/top-ten-us-daily-newspapers/. Accessed: April, 20 2020.

[17]The list in Zellers et al. (2019) includes also wnd.com, bigleaguepolitics.com, naturalnews.com. News articles on Natural News, for example, date back only until 2016. For a more extensive list of sources, see Dicker (2016).

[18]https://tvnews.vanderbilt.edu/

less than four deaths. All shootings in my sample were covered at least by local news outlets.[19] Last, when mass shootings occur in the event-window of another mass shooting, I keep the first shooting. This leaves me with 70 mass shooting events.

Second, to determine which politicians are affected by a mass shooting I use the following rule: *(i)* for all mass shootings, I use politicians from the state in which the mass shooting occurred and *(ii)* for high-visibility mass shootings, I use all politicians at the federal level. This decision is grounded on the political responsibilities for the regulation of firearms in the United States: while highly salient shootings call for a regulatory intervention from the federal government, wide parts of gun policy are under the responsibility of state governments.

Third, I include only politicians that have received at least one gun-rating, are politically active (excluding unsuccessful candidates after the elections) and had an active Twitter account at the day of the mass shooting. In total, these criteria apply to $5,231$ politicians, of which 47% are pro-gun and the residual 53% are anti-gun. For this sample of politicians I scrape all tweets in the five days before and after the mass shooting.

Fourth, I pre-process the tweets. I apply standard rules in the text-as-data literature and remove stop-words, stem words and subset the vocabulary according to a very mild tf-idf criterion (Gentzkow, Kelly and Taddy, 2019).[20] Given the diversity of the tweets, the tf-idf threshold outselects the most rare words, that occur less than 4 times for a total vocabulary of $29,649$ words. Last, I filter out tweets that contain less than 4 words to ensure that the text-algorithms capture meaningful attempts to distract or deceive. The final sample encompasses $494,359$ tweets from $4,065$ politicians.

3.3 Empirical Strategy

To test for the relationship between distortive messaging of politicians and bad news, I exploit the random timing of mass shootings. Specifically, I estimate a high-frequency, non-parametric event-study specification of the following form:

$$\mathrm{Y}_{j,i,d,m} = \sum_{k=-5}^{k=5} \beta_k \times D^k + \Omega_m + \Psi_i + \epsilon_{j,i,d,m} \tag{3.1}$$

where $Y_{j,i,d,m}$ equals 1 if a tweet j from a politician i d days before ($k < 0$) or after ($k \geq 0$) a mass shooting m is either distractive or deceptive, and 0 otherwise. D^k is a set of relative

[19] I obtain consistent results restricting the main analysis to shootings with at least 4 deaths, as in (Yousaf, 2021).

[20] Using the tf-idf-heuristic is a widely applied tool to separate informative terms of a given document in a corpus.

event-time indicators, that take the value of 1 if a day d is k days before or after the mass shooting. These dummies are identified up to a constant and are therefore normalized to $k = -1$, the day before the mass shooting. The coefficients of interest are the β_k's, measuring the change in distortive communication of a politician within a mass shooting event k days relative to the pre-shooting day. Ω_m and Ψ_i are full sets of event and politician fixed effects. The standard errors are clustered at the politician-level.

The key identifying assumption of my design is the absence of pre-trends in baseline outcomes. This assumption is credible when the timing of mass shootings is indeed random. Specifically, pro-gun politicians should not be able to predict the occurrence of mass shootings to pre-emptively disseminate distortive messages that could affect how citizens interpret the shooting or how media reporting about mass shooting unfolds. More importantly, the pattern of estimated effects offers indirect tests of my identification assumption: if mass shootings induce pro-gun politicians to disseminate distortive communication, then the β_k's should be approximately flat in the days prior to the shooting.

While it seems reasonable that politicians cannot anticipate the exact timing of mass shooting, one source of concern is the 'copycat' effect, i.e. media coverage of mass shootings induces subsequent shootings. Jetter and Walker (2018) estimates a positive and statistically significant effect of media coverage on the probability of a new mass shooting lasting up to 10 days. To rule out concerns regarding copycat shootings, I restrict my sample to mass shootings that are at least 11 days after a previous shooting.

Additionally, I compare pro-gun politicians to anti-gun politicians. I test for aggregate statistical differences in communication patterns these two groups in the event-window, by estimating the following specification.

$$\mathrm{Y}_{j,i,d,m} = \alpha\,\text{Pro-Gun} + \delta\, I_{j,k\geq 0} + \gamma\, I_{k\geq 0} \times \text{Pro-Gun} + \Omega_m + \Psi_i + \epsilon_{j,i,d,m} \tag{3.2}$$

The indicator $I_{k\geq 0}$ signals tweets from the day of the shooting and thereafter, the indicator *Pro-Gun* identifies politicians opposing gun-control. The coefficient of interest γ measures the change in outcomes of pro-gun politicians in the aftermath of the mass shooting, relative to the pre-treatment event window and compared to the change in outcomes for anti-gun politicians.

Despite the fact that the group of anti-gun politicians receives *good* news, the comparison of the coefficients is informative. On the one hand, contrary to the intuition that politicians who receive *good* news emphasize these news, I find that anti-gun politicians are not more active on Twitter after the shooting. In contrast, Figure A.3.1 shows that Twitter activity of *both* groups drops identically after the event and converges to pre-shooting levels after three days. More formally, the event-study estimates in Figure A.3.2 don't show statistical differences in Twitter activity between pro-gun politicians and the corresponding effect for anti-gun politicians.[21] Table A.3.2 largely confirms the insignificant differential reaction of pro-gun relative to anti-gun politicians in their Twitter activity after the shooting. Therefore, the differences for distraction and deception from regression 3.2 are purely identified by shifts in the *content* of the messages as opposed to the numerosity of tweeting. On the other hand, in order to gauge if pro-gun politicians behave abnormal, one needs a benchmark to compare them to. If pro-gun politicians focus in the advent of mass shootings primarily on alternative topics, one cannot rule out the interpretation that mass shootings didn't attract sufficient attention in the public sphere without a reference group.

3.4 Distraction

I propose to measure distraction in two distinct approaches. In the indirect approach I identify tweets on mass shootings and treat all non-shooting related tweets as distraction. I impose increasingly strict requirements for the non-shooting related tweets to be classified as distraction. In the direct approach I define the national news in the U.S. as the set of other events to distract with and measure the textual similarity to the tweets. All in all, I find that in the aftermath of mass shootings pro-gun politicians' tweets are more likely to emphasize topics unrelated to the shooting, especially about politics, relative to anti-gun politicians and get closer to the national news.

3.4.1 Pro Gun Politicians' Tweets Focus on Topics Unrelated to Shooting

To begin with, I identify if a tweet speaks about a mass shooting with a keyword-approach and classify the remaining non-shooting related tweets as distraction.

A tweet is defined as speaking about a mass shooting if *(i)* it mentions at least one of the events' location-based keywords, such as "Thousand Oaks", and *(ii)* mentions at least one lemma from the list *"shoot", "gun", "kill", "attack", "massacre", "victim"*. This narrow definition was originally proposed by Demszky et al. (2019) and is best interpreted as direct *reporting* of the event. I call it henceforth the "event-and" definition. I expand this dictionary method to

[21] The event-study plot in panel (a) displays cyclicality in the Twitter activity of both groups. The conclusion is unchanged, when I enrich specification 3.1 with a weekday fixed effect in panel (b). I include a specification with weekday fixed effects in the robustness tests for my main results.

capture more broadly the debates provoked by mass shooting and relax the criteria to require a tweet only to mention one location keyword *or* one lemma from the list.[22] The broader measure is less prone of type-2-errors, but will more likely pick up also tweets before the event that mention the keywords. I call it henceforth the "event-or" definition.
The narrow "event-and" definition classifies 8,909 tweets (or 1.8%) as commenting on the shooting. The broader "event-or" definition instead identifies 38,454 tweets (or 7.8%) as speaking about the shooting.

I estimate specification 3.1 with the indicator whether a tweet comments the mass shooting as outcome on the relative event-time, separately for the pro-gun and the anti-gun sample. Figure 3.1 condenses the regression coefficients in a single graph. Lending credence to the identification assumption, mentions of mass shootings jump on the day of the shooting but do not display pre-trends using either classification of the tweets. Additionally, the increased political attention devoted to the mass shooting events on social media persists for at least 6 days. Last, the narrow event-and measure fades faster than the broader event-or version. I interpret this pattern as the occurrence of a tragic event, that is directly mentioned the days after the incident and then fluidly transitions into a public debate about the causes and consequences of the event.

Figure 3.1 highlights that after a mass shooting, pro-gun politicians are more likely than anti-gun politicians to discuss non-shooting-related topics. While panel (a) suggests that pro-gun politicians are less likely to report the occurrence of mass shootings, there are more striking differences in the broader measure in panel (b). The differences across the two measures indicate that pro-gun politicians report the shootings but are much less likely to discuss the policy implications that arise from the shootings. Given that I observe no differences in the number of tweets sent by pro-gun vis-a-vis anti-gun politicians, the lower attention on mass shootings implies by construction an increased focus on other topics and is thus indirect evidence for distraction.

Additionally, I estimate specification 3.1 on the full sample and interact the relative event dummies with an indicator for being pro-gun. The corresponding plots in Appendix A.3.3 overall underscore the existence of a differential effect. Table 3.1 summarizes the treatment dynamics in the aggregated form of regression 3.2. While it is evident that pro-gun politicians comment on mass shootings (columns 1 and 3), the probability of commenting on a mass

[22] In the short time-frame of five days after the shooting voters would likely connect a tweet that solely mentions the word "shooter" to the recent event.

shooting for a pro-gun politician relative to an anti-gun politician is 0.6 p.p. (4.3 p.p.) lower (columns 2 and 4).

3.4.2 Pro Gun Politicians' Tweets Focus on Topics Related to Politics

To strengthen the interpretation that emphasis on tweets unrelated to the shooting is indeed evidence for distraction, I employ an unsupervised machine learning approach called topic-modelling. The topicmodel has the crucial advantage that the extracted topics are not only informative whether a tweet speaks about a mass shooting, but also which issues are discussed in tweets unrelated to the shooting. I separate tweets that address topics with higher distractive value (e.g., political issues) from tweets with lower distractive value, such as anniversaries or celebrations (e.g., mothers' day). Thereby, the differences in non-shooting related tweets are entirely attributable to other politically relevant topics.

Topicmodelling involves two steps. In the first step, a topicmodel expresses any given text corpus as a mixture of topics, where the topics are learned from variation in word co-occurrence patterns. In the second step, the topicmodel expresses each text as a probability distribution over topics. I apply the Biterm Topic Model and extract 100 topics from the tweets.[23] The topics, reported in Appendix Table A.3.3, are straightforward to interpret. Some refer clearly to mass shootings/gun policy (Topic 9 or 21). Others represent topics that I refer to as political, such as the supreme court nomination (Topic 38), North Korea (Topic 16) or immigration (Topic 62 and 69). In accordance with Barberá et al. (2019) I find a sizeable amount of topics that deal neither with mass shootings nor with political issues. Instead, these topics cover anniversaries, sports or upcoming events, such as town hall meetings. I identify 30 of these topics (as opposed to 47 in the Barberá et al. (2019) paper).

I extract for each tweet the topics with the two highest probabilities and select out tweets as follows: *(i)* If neither of those covers one of the 70 meaningful topics, I disregard the tweet. *(ii)* Moreover, I out select tweets where the topicmodel is insufficiently sure, if the topic relates to a mass shooting - or distraction topic. I measure the ratio between the top-two probabilities and select out tweets that are below the 25th percentile of that ratio.[24]

Employing these selection rules leaves me with 400,205 tweets as compared to 494,359 from the full sample. The fraction of tweets from pro-gun politician is virtually unchanged (37.4%

[23] I use the BTM rather than the more conventional topic-models (i.e. Latent Dirichlet Allocation), as the BTM was specifically developed for topic-modelling within short texts and significantly outperforms the conventional models on those (Yan et al., 2013).

[24] More than 93% of the outselected tweets accrue to non-informative topics.

as compared to 36.6%). The topicmodel assigns 45,921 (or 11.4%) tweets to commenting the event and the remaining 354,284 tweets to other political topics, based on the topic with the highest probability.

The estimates from the event-study regressions with the outcome defined by the topic-model are displayed in Figure 3.2. The treatment dynamics are remarkably similar to the corresponding results in Figure 3.1, both in terms of significance and magnitudes. Despite the use of a completely different classification approach and a sizeable cut in the number of tweets (19% of the full sample), the differences of speaking about mass shootings between pro-gun politicians and anti-gun politicians persist. The elimination of potentially noisy tweets reinforces the indirect evidence for distractive communication from pro-gun politicians as, by construction, the topic-choice is now limited between speaking about the mass shootings or other, politically relevant topics.

The magnitude of the estimated effects is sizeable. The coefficients from Table 3.1 indicate that pro-gun politicians have a 5.8% lower probability to engage in the mass shooting topic in the aftermath of the shooting compared to anti-gun politicians (column 6). Relative to the baseline probability, the three different measures find pro-gun politicians to comment the shootings 33% to 54% less. To infer the extent of distraction, we can reverse the sign of the coefficients from the even columns in Table 3.1, i.e. a pro-gun politician is 5.8% more likely to talk about a different topic than mass shootings after the event, compared to an anti-gun politician. The implied magnitude of this coefficient relative to baseline is rather low (about 7%) which largely owes to the fact that non-shooting topics have a share of nearly 100% before the shooting.

So far, the (indirect) relationship between mass shootings and distractive messaging is robust across different measures from different text-analysis methods as well as an elimination of nearly a fifth of the data. Note as well, that the three measures consider a wide range of tweets (1.8% - 11.5%) as commenting on mass shootings. Table 3.2 further assesses the sensitivity of the distraction effects to the choice of clustering, fixed effects and construction of the sample of mass shootings. The results are insensitive to (i) using heteroskedasticity-adjusted standard errors or (ii) alternative clustering at the state-level, (iii) using weekday fixed effects, (iv) using a mass shooting times politician fixed effect and (v) restraining the set of mass shootings to exceed four deaths as suggested by Yousaf (2021).

Taking stock of the results presented so far, I demonstrate that pro-gun politicians are far less likely to comment mass shooting events with respect to anti-gun politicians and instead emphasize other political topics with the potential to sway away public attention from mass shootings.

3.4.3 Pro Gun Politicians' Tweets Align to National News

To corroborate the finding that pro-gun politicians aim to deflect from mass shootings, I also construct a direct measure for distraction. I use the national news as a proxy for a potential set of other, unrelated events to distract with. Measuring the linguistic overlap between tweets and news abstracts, I show that in the aftermath of the shooting, tweets from pro-gun politicians are getting more similar to national news in absolute terms and relative to anti-gun politicians.

There are three reasons why national news are a good proxy for the set of topics to distract with. First, topics from national news are already salient and don't require a costly search for an attention-grabbing topic. Second, by choosing from the set of issues covered in the news politicians can avoid the accusation of distracting, that voters might associate with a lack of empathy for the victims for example. Third, the focus on national news doesn't require the costly coordination of a distraction strategy among pro-gun politicians.

The measurement of textual similarity between national news and the tweets of politicians comprises three steps. First, I exclude news stories that cover the mass shootings. Specifically, I sort out news whose *(i)* title contains the words *mass shoot* or *school shoot* or *(ii)* whose corpus combines the location-specific with the gun words from the dictionary from Section 3.4.1. Then, for each day, I extract the major 5 news stories for each of the news networks, ABC, CBS, NBC. Following Eisensee and Strömberg (2007) this is measured by air-time.
Second, I follow the text-as-data literature (e.g., Gentzkow, Kelly and Taddy (2019)) and capture the content of news stories by using two-word phrases (bigrams). I restrict attention to two-word phrases that appear at least 4 but no more than 200 times in the news abstracts and keep the highest ranking 4000 bigrams based on the tf-idf statistic.[25] Thereby, I eliminate for example procedural phrases that are common in news abstracts, such as "report introduced", but uninformative of the news topic. The remaining phrases are clearly informative of news topics, such as "hurrican dorian" or "government shutdown". Given that the imposed cutoffs are arbitrary, I show that the results are robust to tightening or loosening the cut-offs for the number of bigrams as well as considering the major 3 or 7 news stories.

[25] To be sure to capture only the gist of the news story, I ex-ante excluded also two-word phrases that contain the title of the shows, i.e. "NBSP", "ABC", "NBS", "CBS" or procedural content, i.e. "report", "introduce", "statement", "shown", "detail". These phrases are not excluded by the tf-idf statistic as they are informative of some documents in the corpus.

Third, for any given day, I measure the cosine similarity between the news abstracts and the subset of tweets, that does not speak about the mass shooting, as defined from Section 3.4.1.

Figure 3.3 depicts the dynamics of regressing news similarity on the days around mass shooting events separately by gun-stance. It is evident that in the days directly before the shooting, the alignment of tweets with the national news is not different from zero for both pro-gun and anti-gun politicians. Instead, in the 3 days following the shooting, the tweets from pro-gun politicians become significantly more similar to national news relative to the baseline. In contrast, the tweets from anti-gun politicians don't show an increased overlap to national news after the shooting, excluding concerns regarding a general trend of higher news reporting of politicians after mass shootings. Note also, that relative to anti-gun politicians, the news similarity of the pro-gun in the aftermath of the shooting is always higher and close to significant for most days.

Table 3.3 reports the results from the aggregated pre-post regression 3.2. The event-study in column (1) points towards an increased closeness of news and pro-gun tweets after mass shootings, however insignificant at conventional levels (p-value 0.15). Restricting attention to the 3 days following and preceding the mass shooting (column 2) shows that the insignificance is primarily driven by the strong drop in news similarity at days 4 and 5 after the shooting, that occurs for both types of politicians. The differential focus on national news with respect to anti-gun politicians is confirmed in column (3) for the full sample and column (4) for the 3-day sample. Note that the point estimates for both columns are identical. The effect is quite sizeable: In comparison to the anti-gun the similarity of pro-gun tweets increases in the aftermath increases by 32% relative to the baseline level of news similarity in the sample.

I assess the robustness of the effects based on the news similarity measure in Table 3.4. The results are insensitive to (i) using heteroskedasticity-adjusted standard errors or (ii) alternative clustering at the state-level, (iii) adding weekday fixed effects to the baseline specification 3.2, (iv) using a mass shooting times politician fixed effect and (v) restraining the set of mass shootings to exceed four deaths as suggested by Yousaf (2021). Further, I present robustness to the choice of number of bigrams and to the choice of the number of news stories in Table A.3.4. The obtained results are largely consistent throughout.

I conclude that based on the direct and indirect method pro-gun politicians seek to lower the salience of the gun policy debate by disproportionately focusing on unrelated issues in the aftermath of the shooting.

3.5 Deception

In this section, I describe the development and application of a neural network model - a deep learning algorithm - to recognize deception. This model detects statement that misrepresent the state of the world, defined as statements about the causes of mass shootings that are highly overrepresented in Fake News (FN) sources relative to Real News (RN) sources. Overall, I establish the finding that pro-gun politicians disseminate misleading information in the aftermath of mass shootings.

3.5.1 Classifying Misrepresentation of the Causes of Mass Shootings

To predict of misrepresentation from newspaper text, I use state-of-the-art neural networks for text analysis from the computer sciences. The algorithm needs to understand (*i*) whether a text makes an association between mass shootings and its causes and (*ii*) whether this association is more predictive of a FN-source relative to a RN-source. In this task, neural networks have strong comparative advantages due to their ability to map complex text-features, like interdependencies of words, to predict the output. Instead, other text-methods that study strategic communication in the political economy literature, such as dictionaries to measure populist rhetoric (Gennaro et al., 2021) or regularized linear regressions to quantify slant (Gentzkow, Shapiro and Taddy (2019), Widmer, Galletta and Ash (2020), are unlikely to excel in modelling the complexity of (*i*) and (*ii*). In addition, neural networks rely on word-embeddings, that represent the meaning of words. Therefore the neural networks' ability to detect misrepresentation is not confined to the *exact* terms it is trained upon (e.g., mental health → mass shooting), as in bag-of-word-approaches, but generalizes (e.g., shooters ← disturbed individuals).[26]

The classification of misleading statements about the causes of mass shootings has three components: (1) I identify statements about causes of mass shootings from news-articles; (2) I train a neural network to distinguish causes predictive for FN-sources versus RN-sources on newspaper data, and (3) I apply the algorithm to Twitter data. Below, I provide a brief summary of these steps and the intuition why my measure captures misleading information, while relegating the detailed description to Appendix A.3.4.

As a first step, I detect statements about the causes of mass shootings in news articles about the shooting events. Initially, I disaggregate news articles into paragraphs and determine the

[26] This is all the more important as the neural network trains on newspaper data but ultimately classifies tweets.

topic of each paragraph by means of a LDA-topicmodel, comparable to the topicmodel in Section 3.4.2. I keep all paragraphs related to causes of mass shootings that are frequently cited in academic research, for instance guns or mental illnesses (Metzl and MacLeish (2015), Duwe (2016)). However, these paragraphs do not necessarily connect the causes to the shooting. For example, a paragraph may discuss mental health issues without a connection to mass shootings. Therefore, I additionally label in a 10% sample of these paragraphs whether an association between causes and shootings is made, or not, and train a neural network to distinguish between the two. Applying the neural network, I remain with 1,050 paragraphs from FN-sources and 4,377 statements from RN-sources that discuss the causes of mass shootings.

In the second step, I train the main neural network to detect statements that misrepresent the causes of mass shootings, that is statements about causes of mass shootings that are predictive of a FN-source relative to a RN-source, in that the likelihood ratio $\Pr(FN|cause)/\Pr(RN|cause)$ is large. The prediction, henceforth $\widehat{FN}$, can be interpreted as a mixture of pure misinformation and misrepresentation of the causes of mass shootings. To see why, note that the predictive power shifts from a RN-source towards the FN-source in only two cases: In the first case, the algorithm detects causes *exclusively* propagated by FN-sources. Second, the predictive value shifts from $\widehat{RN}$ towards $\widehat{FN}$ only if an explanation is heavily overrepresented in the FN-sources compared to RN-sources.[27]

The main neural network takes three inputs: paragraphs related to the causes of mass shootings from FN-sources and RN-sources, plus a random sample of paragraphs that don't discuss the causes of mass shootings, as a large number of tweets will not discuss the causes of mass shootings. These inputs are passed to the deep-learning layers of the neural network that predict if a paragraph makes a false statement about the causes of mass shootings ($\widehat{FN}$), a correct statement about the causes of mass shootings ($\widehat{RN}$) or no statement about the causes of mass shootings ($\widehat{NN}$).

This trained model achieves an F1-score of 0.76 (Recall 0.74, Precision 0.78) in the test-set. The confusion matrix A.3.4 shows that a wide majority of the predictions is located on the diagonal and hence classified correctly. Importantly, the neural network has a particularly low number of statements from RN-sources that are classified as FN-source.

[27] To illustrate this logic consider the following example. I use a simple keyword search to detect explanations related to Islamic terror among the paragraphs from Table A.3.5. According to FBI-investigations, 2.9% (2/70) of the mass shootings have been attributed to Islamic terror. This is closely reflected among RN-sources, where 135 out of 4,258 explanations, i.e. 3.2%, match the keywords. Among the FN-sources, nearly as many, 122 explanations contain those keywords. Given that the total number of causal statements from the FN-sources are only 1,050, this makes up nearly 12% of all explanations.

3.5.2 Pro Gun Politicians' Tweets Misrepresent the Causes of Mass Shootings

Having established that the neural network identifies misleading statements from FN-sources reasonably well on the news data, I apply it now to classify the tweets about mass shootings. The main analysis is conducted on the sample of tweets that uses the location - *or* gun-specific keywords.[28] To further ensure that my results are not driven by false positives, I adopt throughout the rule that a tweet is only assigned to $\widehat{FN}$, if it's probability exceeds 90%.[29]

Descriptive Analysis

Figure 3.4 plots the predictions of the algorithm on the sample of tweets speaking about the mass shooting. Four findings emerge. *(i)* The overwhelming majority of statements from pro-gun and anti-gun politicians make no any claim about the causes of mass shootings ($\widehat{NN}$). Most statements are condolences to victims' families or calls for action. *(ii)* The overall share of statements about the causes of mass shootings made from anti-gun politicians (14%) exceeds that of pro-gun politicians (8%).[30] *(iii)* Third, the share of misrepresenting statements ($\widehat{FN}$) is 3.3% for pro-gun politicians as compared to 1.4% for anti-gun politicians. As discussed in the previous section, the $\widehat{FN}$ contains as well explanations that are heavily distorted towards Fake News sources but not necessarily factually false. A truthful account of the state of the world would therefore require a non-zero share of $\widehat{FN}$. *(iv)* The relative share of the $\widehat{FN}$ with respect to all statements about the the causes of mass shootings is more than 4 times higher for pro-gun politicians (40.8%) relative to anti-gun politicians (9.7%). Hence, the communication of the causal factors underlying mass shootings from pro-gun politicians is disproportionally centered around explanations that are typically propagated by FN-sources to RN-sources.

Event Studies

The event-study analysis is plotted in Figure 3.5. In panel (a) I show the results of regressing the indicator for $\widehat{FN}$ on the dummies for relative-event time. In panel (b) I repeat this regression with the $\widehat{FN}$. Intuitively, for both measures, the spread of narratives about the causes of mass shootings *before* the actual event are precisely estimated zeros. Considering the main outcome, panel (a) demonstrates that pro-gun politicians react immediately to mass killings spreading misleading statements about its causes. Consistent with the findings of Vosoughi, Roy and Aral (2018) that falsehoods diffuse faster and deeper on Twitter, politicians disseminate of

[28] I show in Table 3.5 that my main results hold as well in the more narrow event-and sample as well as in the more broad sample from the topicmodel.

[29] The main results are robust for relaxing this threshold, as demonstrated in Table 3.6.

[30] Broadly these numbers align with the algorithms that were predicting statements in Section A.3.4. The fraction of paragraphs classified as making statements about the causes of mass shootings was between 11% - 15% for the paragraphs from the topics or 3% - 4% on the full sample of paragraphs.

misleading statements about mass shootings as early as possible. The point estimates for the pro-gun politicians at the day of the mass shooting and the subsequent day are statistically different from anti-gun politicians. This difference and the effect for pro-gun in general dissipates around the third day after the mass shooting. In contrast, the propagation of truthful reporting of the state of the world is by large governed by anti-gun politicians. I plot the event-studies that estimate the differential effect in Figure A.3.5. The results are consistent throughout.

Table 3.5 reports the estimates from the pre-post regressions. In all three potential samples the spread of factually misleading explanations of mass shootings by pro-gun politicians is confirmed (columns 1,4, and 7). Additionally, the estimates show this emphasis on distortive messages surrounding the reasons for shootings is significantly stronger for pro-gun compared to anti-gun politicians (columns 2,5 and 8). In my preferred specification, column (2), the point estimate indicates that relative to the sample baseline the spread of false information increases by 54%. On the contrary, narratives that are diagnostic of the major newspapers in the U.S. get substantially less attention from pro-gun politicians, quantitatively and qualitatively (columns 3,6 and 9).

Additionally to the variations in the sample, I show in Table 3.6 that the findings presented above are robust to alternative clustering of standard errors (columns 1 and 2), adding weekday fixed effects (column 3), interacting the fixed effects (column 4), the alternative definition of mass shootings (column 5) and relaxing the threshold for $\widehat{FN}$ to 0.7 (column 5) or 0.8 (column 6).

Taken as a whole, the results in this section reveal that politicians react to bad news events by casting doubt on the interpretation of the event through factually misleading content. The political response takes place immediately to intervene in voters' belief updating as early as possible.

3.6 Heterogeneity

Thus far, the objective of the investigation was to demonstrate the diffusion of distortive messaging after bad news. I now shift the focus to specific groups of events or politicians and test how diffusion of distortive messaging varies between them.

In the first part, I investigate the role of electoral proximity and number of victims, as a proxy for the size of the bad news shock. One would expect to observe the spread of distortive communication to be largest, when the shooting is close to an election or involves many victims. In the second part, I test for heterogeneity across the type of politician and the role of electoral competition. In the third part, I examine the interplay between the media environment and the spread of distraction and deception. In particular, I investigate whether co-occurring natural disasters, representing a jump in the returns of distraction, lead to a increased use of distraction vis-a-vis deception.

Formally, I run Regression 3.1 including an interaction with an indicator for the respective heterogeneity test. Moreover, I repeat also Regression 3.2 to see the difference in behavior for pro-gun politicians and anti-gun politicians separately for different samples.

Heterogeneity across Electoral Cycle

To assess how distortive communication differs over the electoral cycle, I compare shootings that occur 180 days before a general election to shootings that take place in the 180 days thereafter. I find that electoral incentives are important: The results indeed show that the effects are significantly larger in pre-election periods for both, distraction (Table 3.7, Column 1) and deception (Table 3.8, Column 1). Both findings hold also in comparison to anti-gun politicians (Table 3.7, Column 2; Table 3.8, Column 2).

These results are consistent with previous findings that mass shootings have political costs, especially close to elections, through increasing the salience of gun policy (Yousaf, 2021).

Heterogeneity by Number of Victims

Next, I construct an indicator equal to 1 if the number of victims in a shooting exceeded the median number of victims (9). A higher number of victims should necessitate a stronger reaction of politicians, as conceivably media interest and public demands for changing gun policy increase in the number of the victims. I find that pro-gun politicians deceive more when shootings the number of involved persons is large, in absolute (Table 3.8, Column 3) and relative to anti-gun politicians (Table 3.8, Column 4). Looking at distraction offers a more nuanced picture. Unsurprisingly, I observe that pro-gun politicians comment the big shootings more, when drawing on the indirect measure (Table 3.7, Column 3). However, the relationship is significantly weaker than anti-gun politicians (Table 3.7, Column 4). Repeating the analysis with the news-similarity measure, confirms that pro-gun politicians distract more when the number of victims increases (Table A.3.6, Column 3; Table A.3.6, Column 4). On the one hand, the findings suggest that increments in the number of victims force pro-gun politicians to take a stand on the shootings. On the other hand, pro-gun politicians react to this circumstance by increasing the number of deceptive messages as well as referring more to unrelated topics covered in the news.

Heterogeneity by Type of Candidate.

Then, I test for differences between incumbents versus candidates (Tables 3.7 and 3.8, Columns 5 and 6). I observe that pro-gun office holders are less likely to distract than candidates. However, I don't find differences in comparison to anti-gun politicians or with regard to news similarity. Deceptive messages come primarily from elected officials, both absolute and relative to anti-gun politicians. In search for the political responsibility for mass shootings, media frequently cite pro-gun office holders. Given that limelight, incumbents in general are more likely speak about gun-policy and pro-gun office holders respond by misrepresent the causes of mass shootings.

Heterogeneity by Electoral Competition.

Last, I test for differential effects depending on the degree of electoral competition. On the one hand, anticipated closeness of the election increases the returns to distortive messages and one would expect larger coefficients. On the other hand, a common proposition in the persuasion literature is that competition among senders limits the scope for distortion (DellaVigna and Gentzkow, 2010). Through the lens of the persuasion models, heightened electoral competition raises the costs of untruthful messages, for example due to a potential loss in reputation as media attention on close races increases the likelihood of detection of distortive messages. In this case, we would expect smaller coefficients.

Using actual electoral competition exhibits the main disadvantage of being affected by the mass shootings (Yousaf, 2021). Therefore, I use *anticipated* electoral competition. I construct an indicator for high anticipated electoral competition that equals one if the runner-up lost the last election in the same geographic unit by a margin of less than 10 percentage points, which corresponds to the 90th percentile of the vote margin in my data.[31]

For both types of distortive messages, the results for pro-gun politicians alone and in comparison to the anti-gun politicians are insignificant. The point estimates tend to support the explanation that electoral competition decreases the spread of distortive messages (Table 3.7, Columns 7 and 8; Table 3.8, Columns 7 and 8).

3.6.1 News Environment

In this section I want to understand the role of other news events in shaping politicians' distortive communication. I exploit an exogenous shock to the news environment: hurricanes. On the one hand, natural disasters that co-occur with mass shootings can diminish attention devoted to the shooting and reduce the need for distortive communication. On the other

[31] The results are the same using the 80th percentile.

hand, as seen anecdotally in the introduction, natural disasters represent an excellent topic to distract the public. In a conceptual framework where politicians have to decide jointly whether to distort and to choose between distracting or distorting, natural disasters can be interpreted as an exogenous increase in the effectiveness of distraction.

In order to determine the co-occurrence of highly visible mass shootings and natural disasters, I cross-check the EM-DAT database with the Vanderbilt national news. To ensure that I pick up events that are newsworthy for the U.S. media, I focus on hurricanes that occur in the U.S. and that fulfil one criteria: the generation of a bigram with the disaster name in the major news, as defined in Section 3.4.3. Intuitively, this captures natural disasters that occurred within my sample period and were sufficiently newsworthy. In total, this generates 7 incidences of co-occurrences of natural disasters with mass shootings. On average, the number of co-occurring days is 8 and the minimum is 6. The average number of victims in the EM-DAT data is 46.[32]

The results are displayed in Table 3.9. Column 1 and Column 2 use as the dependent variable a topic discovered by the topicmodel, labelled "Hurricane, Storm, Preparation" (topic 45 of Table A.3.3). This represents a direct test for differential emphasis on natural disasters. I observe that pro-gun politicians focus on this topic more so in the aftermath of mass shootings and significantly more so than anti-gun politicians (at the 1% level). In the rest of the table, I interact the standard outcomes for distraction and deception with an dummy that indicates the co-occurrence of a natural disaster with mass shootings. I find that news-similarity from pro-gun politicians gets significantly closer to the national news after shootings when natural disasters take place. This result also holds with respect to anti-gun politicians and further underpins the conjecture that pro-gun politicians actively use natural disasters to divert attention away from mass shootings. Moreover, I find that the gap in coverage of mass shootings between pro-gun and anti-gun politicians widens even further when mass shootings co-occur with hurricanes. Last, the occurrence of natural disasters decreases the spread of deceptive messages (Table 3.9, Columns 7 and 8).

Overall, the results confirm the highly strategic nature of distortive political communication and demonstrate that choice of the distortive strategy reacts to changes in the returns of the effectiveness. While politicians have been found to conceal objectionable behavior such as voting against constituent interests (e.g. Kaplan, Spenkuch and Yuan (2018)) or controversial presidential orders (e.g. Djourelova and Durante (2019)) during times of natural disasters, my

[32]This selection is very comparable to Djourelova and Durante (2019) but differs by choosing hurricanes instead of earthquakes. Employing the selection criteria of Djourelova and Durante (2019) to the EM-DAT data leaves me with 0 incidences of co-occurring earthquakes and mass shootings.

results provide evidence that politicians can use natural disasters to draw attention away from mass shootings.

3.7 Discussion and Conclusion

It has long been argued that elite communication strongly influences which considerations citizens take into account and thus what judgements they reach (Zaller et al. (1992), Kuklinski et al. (2000)). In this paper, I develop novel and broadly applicable text-based measures to unveil two distortive communication strategies: distraction and deception. I apply the measures to provide direct evidence on the use of distortive messaging by politicians that are exposed to a news shock that is detrimental for the politicians' policy platform. To this end, I exploit the random timing of mass shootings in the United States in an event-study design that uses nearly half a million tweets from politicians with different gun policy-agendas.

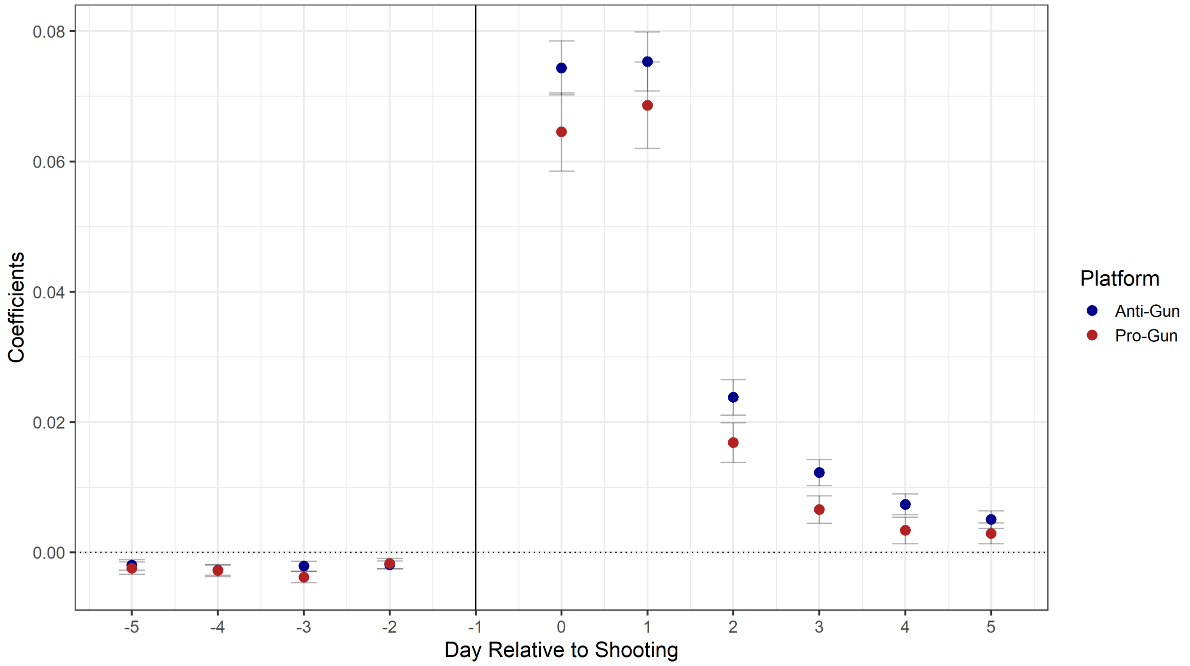
Four main findings emerge. First, pro-gun politicians' tweets cover mass shootings considerably less than politicians that demand more gun control and instead focus on other political topics. Moreover, in the days following the shooting tweets from pro-gun politicians get closer to unrelated topics covered in the national news, in absolute terms and relative to anti-gun politicians. Second, when participating in the gun policy debate, pro-gun politicians spread misleading accounts of mass shootings. Distortive communication by pro-gun politicians is stable before mass shootings, but jumps immediately thereafter. Third, estimated effects tend to be larger when the potential costs of bad news are larger. Fourth, politicians are reactive to sudden shifts in the returns of distortive strategies, that is when shootings co-occur with hurricanes pro-gun politicians focus on distraction rather than deception. All in all, my results highlight that when politicians are exposed to bad news, they use their communicative resources to decrease the salience of the critical event and alter citizens' belief updating.

A natural question to ask is whether distortive statements from a group of politicians affect aggregate beliefs. Grossman and Helpman (2019) show untruthful communication from political opponents can cancel out and the full information benchmark emerges. However, as soon as political parties are able to target audiences that disproportionally contain their supporters, policy platforms diverge and the electorate polarizes. The rise of social media platforms as a primary source of information heightens concerns that the second case prevails (Sunstein, 2018). Indeed, empirical research points out that Republican and Democratic voters update their policy views in *opposite* directions in the aftermath of mass shootings (Yousaf, 2021). While consistent with a belief-updating model where Republicans and Democrats have heterogeneous priors about the usefulness of stricter gun control, my results highlight an alternative mechanism: individuals create their own reality based on the very same *available* facts because they receive different signals on the causes of mass shootings from political elites.

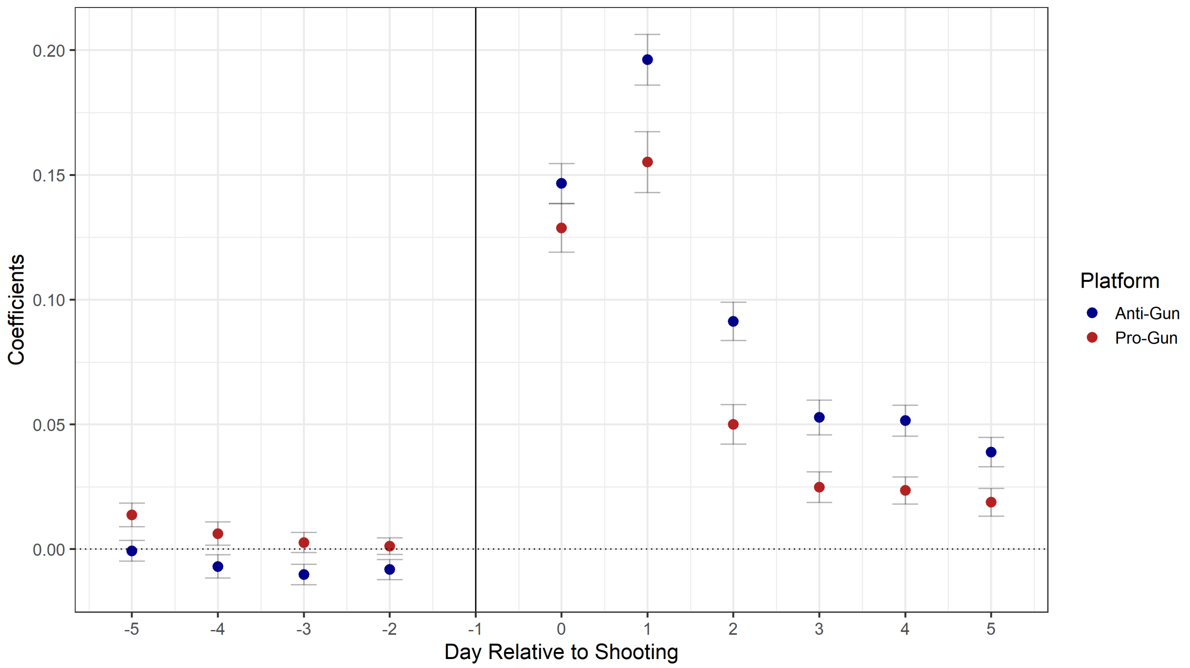
Apart from their implications for polarization, my findings advance our understanding on the emergence of misperceptions and the fight against them. This work sheds light on the timing, source and type of misleading information that is supplied by political elites in the context of mass shootings. This information may foster the creation of misperceptions in the population, as suggested in Flynn, Nyhan and Reifler (2017). Nevertheless, modern democracies are not at the mercy of powerful opinion leaders: citizens can discount information provided by politicians; media organizations can debunk false information. Due to the sheer amount of (political) communication, the design of counter-strategies requires knowledge of type of manipulation techniques and an efficient allocation of the scarce resources of attention and fact-checking.

Last, my paper heightens concerns that politicians can actively evade political accountability. Previous research has documented that politicians use times of inattention by the media, for instance due to natural disasters, to circumvent public scrutiny and act against constituent interests. My results emphasize that politicians employ distraction techniques to reduce the public attention for an issue that reflects negatively on their policy position. It is an open question, with important implications for political accountability, to see whether that attempt to change the trajectory of the public discourse is successful.

Figure 3.1: Event-Study Plot, Distraction, Keywords



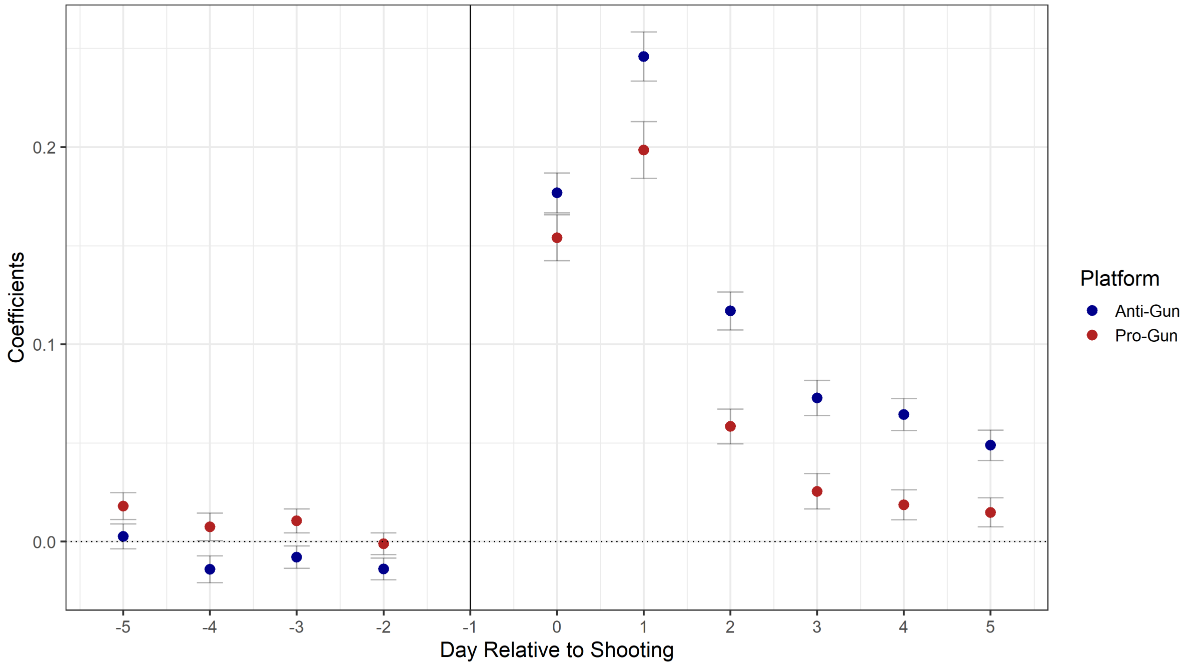
(a) Keywords: Event and Gun



(b) Keywords: Event or Gun

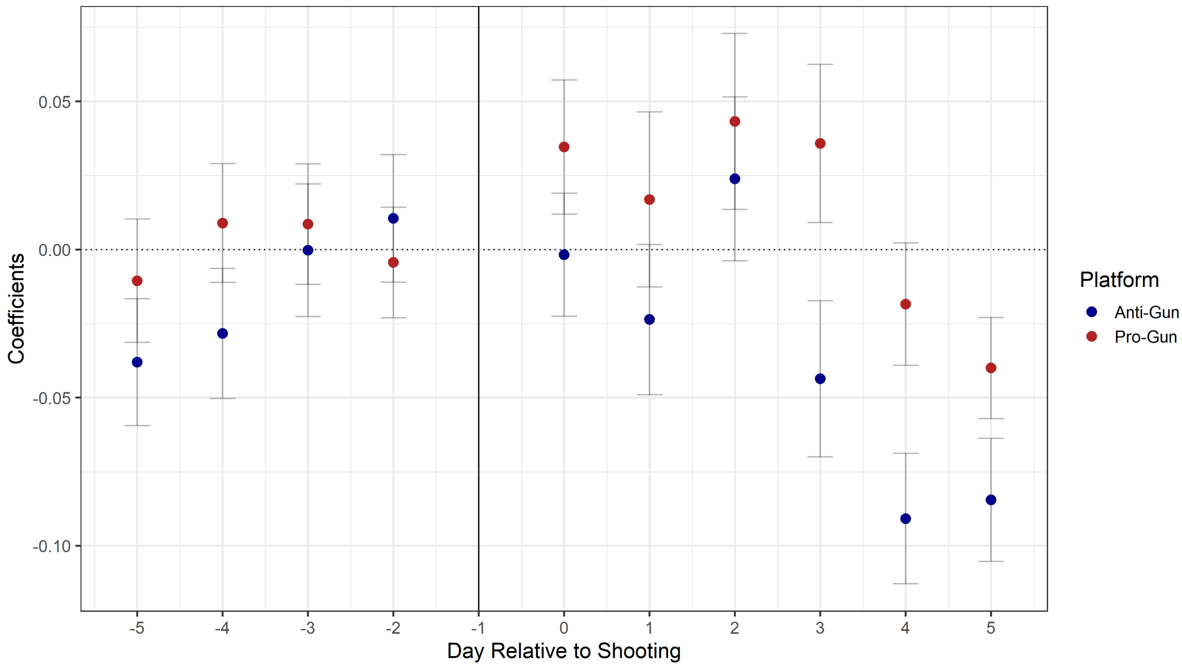
Notes. - The figure plots the estimates of β_k and 95-percent confidence intervals from specification [3.1]. The dependent variable is an indicator that is equal to 1 if tweet j from politician i on day t is contains the respective keywords, and 0 otherwise. The indicator is regressed on a set of leads and lags of days relative to the day before the mass shooting, separately for tweets from pro-gun and anti-gun politicians. The specification controls for politician and mass-shooting FE. Standard errors are clustered by politician. Sample: Full sample.

Figure 3.2: Event-Study Plot, Distraction, Topicmodel



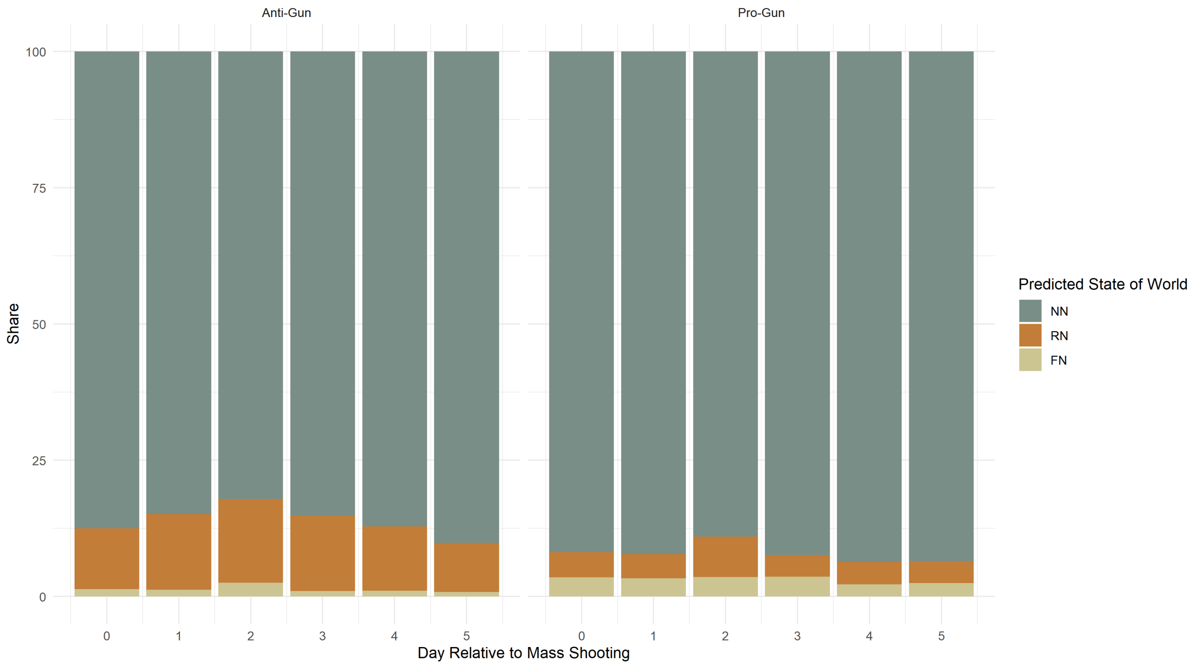
Notes. - The figure plots the estimates of β_k and 95-percent confidence intervals from specification [3.1]. The dependent variable is an indicator that is equal to 1 if the topic of if tweet j from politician i on day t is related to mass shootings, and 0 otherwise. The indicator is regressed on a set of leads and lags of days relative to the day before the mass shooting, separately for tweets from pro-gun and anti-gun politicians. The specification controls for politician and mass-shooting FE. Standard errors are clustered by politician. Sample: Topicmodel sample.

Figure 3.3: Event-Study Plot, Distraction, News Similarity to National News



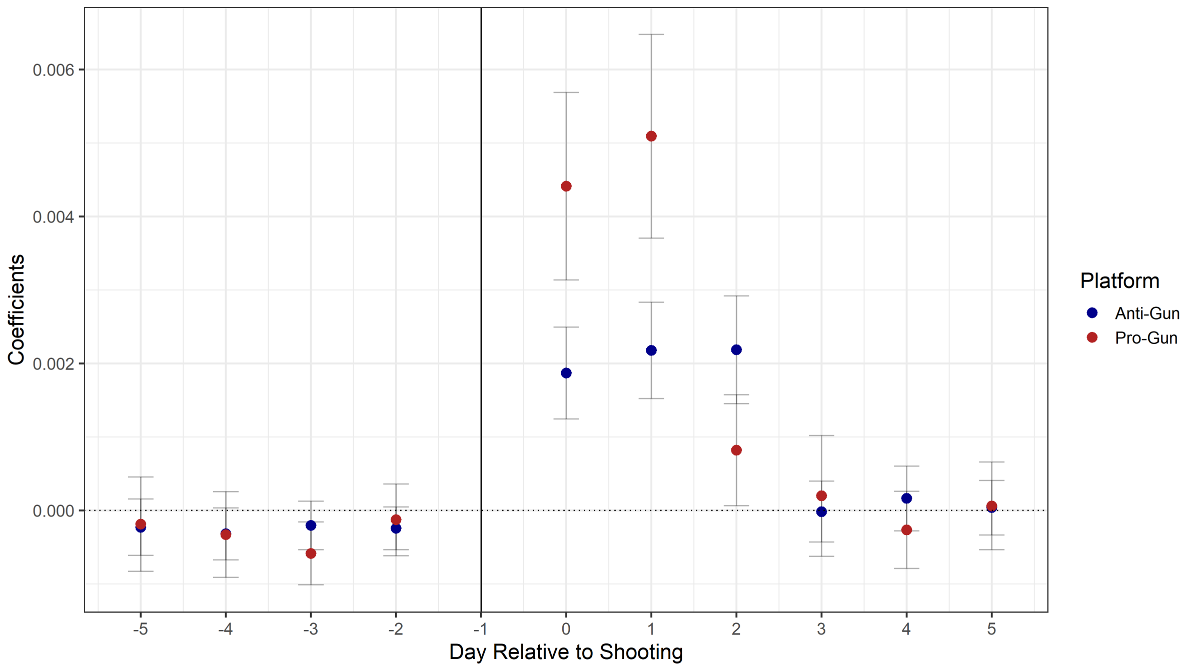
Notes. - The figure plots the estimates of β_k *and 95-percent confidence intervals from specification [3.1]. The dependent variable is the cosine similarity between national news and if tweet* j *from politician* i *on day* t*, multiplied by 100. The outcome is regressed on a set of leads and lags of days relative to the day before the mass shooting, separately for tweets from pro-gun and anti-gun politicians. The specification controls for politician and mass-shooting FE. Standard errors are clustered by politician. Sample: Full sample.*

Figure 3.4: Predicted State of World of Tweets Speaking about Mass Shooting

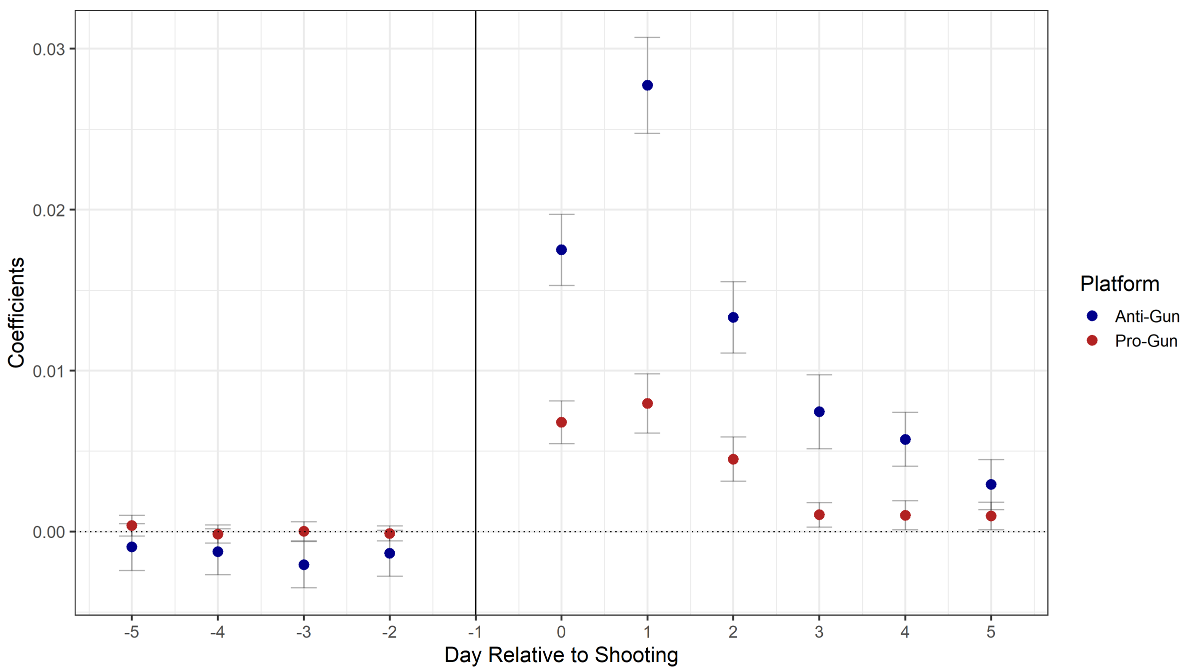


Notes. - The figure plots the shares of assigned label from the neural networks for tweets for the days after mass shootings for pro-gun politicians (right) and anti-gun politicians (left). A tweets gets assigned the fake news label if the neural network predicts it as fake with more than 90% certainty. The other classes are determined by the higher probability.

Figure 3.5: Event-Study Plot, Deception



(a) Sample - Keywords: Event or Gun



(b) Sample - Keywords: Event or Gun

Notes. - The figure plots the estimates of β_k and 95-percent confidence intervals from specification [3.1]. The dependent variable in panel (a) is an indicator for deception that is equal to 1 if tweet j from politician i on day t is labelled by the neural network as fake with more than 90% probability, and 0 otherwise. The dependent variable in panel b is an indicator for accurate information provision that is equal to 1 if tweet j from politician i on day t is labelled by the neural network as real news, and 0 otherwise. The indicator is regressed on a set of leads and lags of days relative to the day before the mass shooting, separately for tweets from pro-gun and anti-gun politicians. The specification controls for politician and mass-shooting FE. Standard errors are clustered by politician. Sample: Full sample.

Table 3.1: Effect of Mass Shootings on Commenting Mass Shootings

	Dependent variable:					
	(Event+Shoot)	(Event+Shoot)	(Event|Shoot)	(Event|Shoot)	Shoot-Topic	Shoot-Topic
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.029***	0.035***	0.061***	0.101***	0.070***	0.128***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)
Pro-Gun		0.004		0.011		0.036***
		(0.003)		(0.007)		(0.011)
Post × Pro-Gun		−0.006***		−0.040***		−0.058***
		(0.001)		(0.003)		(0.004)
Mean Dep Var	0.016	0.018	0.059	0.078	0.091	0.115
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes
MS FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	185,051	494,359	185,051	494,359	146,445	400,205
R^2	0.054	0.046	0.085	0.097	0.094	0.116

Notes. - The table reports the OLS estimates from specifications 3.1 and 3.2. The dependent variable is an indicator that is equal to 1 if tweet j from politician i on day t speaks about the mass shooting, and 0 otherwise. Speaking about a mass shooting is defined as containing mentioning event and shooting-related words in columns (1) and (2), event or shooting-related words in columns (3) and (4), or if the topic with the highest probability from the topicmodel deals with the shooting in columns (5) and (6). The dependent variable is regressed on an indicator for the day of the mass shooting and thereafter, for a pro-gun rating, and their interaction. All columns include politician and mass shooting fixed effects. Columns (1),(3),(5) refer to the sample of pro-gun politicians; columns (2),(4),(6) to the full sample. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.2: Robustness - Distraction

	Dependent variable:				
	1(Shoot-Topic)				
	(1)	(2)	(3)	(4)	(5)
Post	0.128***	0.128***	0.124***	0.125***	0.142***
	(0.001)	(0.008)	(0.003)	(0.003)	(0.003)
Pro-Gun	0.036***	0.036***	0.038***	0.049	0.044***
	(0.006)	(0.009)	(0.011)	(0.059)	(0.011)
Post*Pro-Gun	−0.058***	−0.058***	−0.058***	−0.055***	−0.069***
	(0.002)	(0.006)	(0.004)	(0.004)	(0.004)
Mean Dep Var	0.115	0.115	0.115	0.115	0.124
Pol FE	Yes	Yes	Yes	No	Yes
MS FE	Yes	Yes	Yes	No	Yes
Weekday FE	No	No	Yes	No	No
Pol X MS FE	No	No	No	Yes	No
Observations	400,205	400,205	400,205	400,205	337,633
R^2	0.116	0.116	0.123	0.188	0.123

Notes. - The table reports the OLS estimates from specification 3.2. The dependent variable is an indicator equal to 1 if tweet j from politician i on day t speaks about the mass shooting, that is the topic with the highest probability from the topicmodel deals with the shooting, and 0 otherwise.
Each column replicates the baseline specification (column 6 of Table 1) with the following modifications. Column (1): robust standard errors. Column (2): standard errors clustered at the state level. Column (3): weekday fixed effects. Column (4): mass shooting × politician fixed effects. Column (5): only shootings with more than 4 deaths. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.3: Effect of Mass Shootings on News Similarity

	Dependent variable:			
	News Similarity (Top-5)			
	(1)	(2)	(3)	(4)
Post	0.00893	0.03500***	−0.03067***	−0.01103
	(0.00620)	(0.00887)	(0.00547)	(0.00719)
Pro-Gun			−0.05908	−0.05382
			(0.04458)	(0.04904)
Post*Pro-Gun			0.03938***	0.04117***
			(0.00804)	(0.01074)
Mean Dep Var	0.0999	0.1109	0.1179	0.1329
Pol FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	174,194	109,710	455,905	288,039
R^2	0.02310	0.03496	0.02366	0.03186

Notes. - The table reports the OLS estimates from specification 3.2. The dependent variable is the cosine similarity between evening news and tweet j from politician i on day t, multiplied by 100. The dependent variable is regressed on an indicator for the day of the mass shooting and thereafter, for a pro-gun rating, and their interaction. All columns include politician and mass shooting fixed effects. Columns (1) and (2) refer to the sample of pro-gun politicians; columns (3) and (4) to the full sample. Columns (1) and (3) refer to the full sample of days, columns (2) and (4) shrink the even window to 3 days before and after the mass shooting. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.4: Robustness News Similarity

	Dependent variable:				
	News Similarity (Top-5)				
	(1)	(2)	(3)	(4)	(5)
Post	-0.031^{***}	-0.031^{**}	-0.031^{***}	-0.032^{***}	-0.040^{***}
	(0.004)	(0.012)	(0.005)	(0.005)	(0.006)
Pro-Gun	-0.059^{***}	-0.059	-0.059	-0.082	-0.065
	(0.016)	(0.040)	(0.044)	(0.076)	(0.043)
Post*Pro-Gun	0.039^{***}	0.039^{***}	0.039^{***}	0.039^{***}	0.052^{***}
	(0.006)	(0.010)	(0.008)	(0.008)	(0.009)
Mean Dep Var	0.1179	0.1179	0.1179	0.1179	0.1145
Pol FE	Yes	Yes	Yes	No	Yes
MS FE	Yes	Yes	Yes	No	Yes
Weekday FE	No	No	Yes	No	No
Pol X MS FE	No	No	No	Yes	No
Observations	455,905	455,905	455,905	455,905	383,309
R^2	0.024	0.024	0.024	0.072	0.024

Notes. - The table reports the OLS estimates from specification 3.2. The dependent variable is the cosine similarity between evening news and tweet j from politician i on day t, multiplied by 100.
Each column replicates the baseline specification (column 4 of Table 3) with the following modifications. Column (1): robust standard errors. Column (2): standard errors clustered at the state level. Column (3): weekday fixed effects. Column (4): mass shooting × politician fixed effects. Column (5): only shootings with more than 4 deaths. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.5: Effect Mass Shootings on Deceptive Communication

	Dependent variable:								
	FN	FN	RN	FN	FN	RN	FN	FN	RN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post	0.00167***	0.00060***	0.00480***	0.00192***	0.00122***	0.01345***	0.00191***	0.00142***	0.01751***
	(0.00019)	(0.00007)	(0.00022)	(0.00024)	(0.00014)	(0.00051)	(0.00025)	(0.00016)	(0.00071)
Pro-Gun		0.00010	0.00076		−0.00001	0.00395***		0.00098	0.00603***
		(0.00082)	(0.00075)		(0.00101)	(0.00148)		(0.00109)	(0.00202)
Post × Pro-Gun		0.00112***	−0.00310***		0.00076***	−0.00981***		0.00058*	−0.01271***
		(0.00021)	(0.00026)		(0.00029)	(0.00058)		(0.00031)	(0.00080)
Mean Dep Var	0.00098	0.00059	0.0019	0.00178	0.00136	0.0019	0.00193	0.00151	0.0095
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MS FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	185,051	494,359	494,359	185,051	494,359	494,359	149,592	408,971	408,971
R^2	0.03052	0.02127	0.01911	0.02849	0.01891	0.02780	0.03294	0.02166	0.03372

Notes. - The table reports the OLS estimates from specifications 3.1 and 3.2. The dependent variable is an indicator that is equal to 1 if tweet j from politician i on day t is predicted as fake news (FN) with a probability of more than 90% from the neural network in columns (1),(2),(4),(5),(7),(8), , and 0 otherwise. The dependent variable is an indicator that is equal to 1 if tweet j from politician i on day t as real news (RN) in columns (3),(6),(9), and 0 otherwise. The neural network classifies tweets about mass shootings from the *event-and* method in columns (1)-(3), the *event-or* method in columns (4)-(6) and the *topicmodel* method in columns (7)-(9). The dependent variable is regressed on an indicator for the day of the mass shooting and thereafter, for a pro-gun rating, and their interaction. All columns include politician and mass shooting fixed effects. Columns (1),(4),(7) refer to the sample of pro-gun politicians; columns (2),(3),(5),(6),(8),(9) to the full sample. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.6: Robustness Deception

	Dependent variable:						
	1(FN)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Post	0.00122***	0.00122***	0.00116***	0.00124***	0.00140***	0.00165***	0.00144***
	(0.00013)	(0.00011)	(0.00013)	(0.00014)	(0.00016)	(0.00016)	(0.00016)
Pro-Gun	−0.00001	−0.00001	0.00003	0.01002	0.0000002	0.00060	0.00001
	(0.00059)	(0.00072)	(0.00101)	(0.01074)	(0.00101)	(0.00110)	(0.00105)
Post*Pro-Gun	0.00076***	0.00076***	0.00075***	0.00067**	0.00085***	0.00066**	0.00065**
	(0.00022)	(0.00023)	(0.00029)	(0.00029)	(0.00032)	(0.00032)	(0.00030)
Mean Dep Var	0.00136	0.00136	0.00136	0.00136	0.00153	0.00185	0.00162
Pol FE	Yes	Yes	Yes	No	Yes	Yes	Yes
MS FE	Yes	Yes	Yes	No	Yes	Yes	Yes
Weekday FE	No	No	Yes	No	No	No	No
Pol X MS FE	No	No	No	Yes	No	No	No
Observations	494,359	494,359	494,359	494,359	418,463	494,359	494,359
R^2	0.01891	0.01891	0.01969	0.05491	0.02006	0.02021	0.02016

Notes. - The table reports the OLS estimates from specification 3.2. The dependent variable is an indicator equal to 1 if tweet j from politician i on day t is predicted as fake news (FN) with a probability of more than 90% from the neural network, and 0 otherwise.

Each column replicates the baseline specification (column 5 of Table 3.5) with the following modifications. Column (1): robust standard errors. Column (2): standard errors clustered at the state level. Column (3): weekday fixed effects. Column (4) mass shooting × politician fixed effects. Column (5) only shootings with more than 4 deaths. Column (6) lowers the threshold for being classified as fake to 80%, Column (6) lowers the threshold for being classified as fake to 70%. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.7: Effect of Mass Shootings on Distractive Communication - Heterogeneity

	Dependent variable: Shoot-Topic							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post ×	-0.070^{***}	0.010						
Before Election	(0.018)	(0.008)						
Post ×		-0.085^{***}						
Pro-Gun × Before Election		(0.019)						
Post ×			0.018^{**}	0.073^{***}				
Big Shoot			(0.007)	(0.006)				
Post ×				-0.057^{***}				
Pro-Gun × Big Shoot				(0.009)				
Post ×					0.020^{***}	0.016^{***}		
Incumbent					(0.005)	(0.006)		
Post ×						0.002		
Pro-Gun × Incumbent						(0.008)		
Post ×							0.018	0.006
Elect Comp							(0.014)	(0.011)
Post ×								0.014
Pro-Gun × Elect Comp								(0.018)
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	62,740	174,233	146,573	400,439	146,573	400,439	114,173	330,791
R^2	0.118	0.143	0.095	0.117	0.095	0.116	0.080	0.110

Notes. - The table reports OLS estimates from specifications 3.1 and 3.2. The dependent variable is an indicator that is equal to 1 if a tweet j from a politicians i at a day d speaks about the mass shooting, that is the topic with the highest probability from the topicmodel deals with the shooting, and 0 otherwise. Before Election refers to shootings 180 days before a general election, and for the 180 days after equals 0. Big Shoot refers to shootings with an above median number of victims, and 0 otherwise. Incumbent refers to the sample of incumbents, and 0 otherwise. Elect Comp refers to politicians with high anticipated electoral competition, and 0 otherwise. All columns include politician and mass shooting fixed effects. Columns (1),(3),(5),(7) refer to the sample of pro-gun politicians; columns (2),(4),(6),(8) to the full sample. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.8: Effect of Mass Shootings on Deceptive Communication - Heterogeneity

	Dependent variable: FN							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post ×	0.0047^{***}	0.0019^{***}						
Before Election	(0.0008)	(0.0004)						
Post ×		0.0030^{***}						
Pro-Gun × Before Election		(0.0009)						
Post ×			0.0023^{***}	0.0012^{***}				
Big Shoot			(0.0004)	(0.0002)				
Post ×				0.0011^{**}				
Pro-Gun × Big Shoot				(0.0005)				
Post ×					0.0015^{***}	0.0003		
Incumbent					(0.0005)	(0.0003)		
Post ×						0.0013^{**}		
Pro-Gun × Incumbent						(0.0005)		
Post ×							−0.0004	0.0003
Elect Comp							(0.0008)	(0.0007)
Post ×								−0.0006
Pro-Gun × Elect Comp								(0.0010)
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	80,508	219,741	185,232	494,707	185,232	494,707	143,843	405,687
R^2	0.0401	0.0285	0.0286	0.0190	0.0286	0.0190	0.0207	0.0150

Notes. - The table reports OLS estimates from specifications 3.1 and 3.2. The dependent variable is an indicator equal to 1 if a tweet is predicted as fake news (FN) with a probability of more than 90% from the neural network, and 0 otherwise. Before Election refers to shootings 180 days before a general election, and for the 180 days after equals 0. Big Shoot refers to shootings with an above median number of victims, and 0 otherwise. Incumbent refers to the sample of incumbents, and 0 otherwise. Elect Comp refers to politicians with high anticipated electoral competition, and 0 otherwise. All columns include politician and mass shooting fixed effects. Columns (1),(3),(5),(7) refer to the sample of pro-gun politicians; columns (2),(4),(6),(8) to the full sample. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table 3.9: Effect of Mass Shootings on Distortive Communication - Heterogeneity by News Environment

	Dependent variable:							
	Disaster Topic	Disaster Topic	News Similarity	News Similarity	Shoot-Topic	Shoot-Topic	FN	FN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.0042***	0.0004	−0.0034	−0.0213***	0.0693***	0.1240***	0.0022***	0.0014***
	(0.0013)	(0.0006)	(0.0071)	(0.0058)	(0.0028)	(0.0033)	(0.0003)	(0.0001)
Post × Pro-Gun		0.0038***		0.0164*		−0.0555***		0.0009***
		(0.0014)		(0.0088)		(0.0044)		(0.0003)
Post × Hurricane			0.0270***	0.0026	0.0008	0.0035***	−0.0002***	−0.0001***
			(0.0068)	(0.0048)	(0.0009)	(0.0009)	(0.00005)	(0.00002)
Post × Pro-Gun × Hurricane				0.0210***		−0.0023*		−0.0001**
				(0.0080)		(0.0013)		(0.0001)
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Pro-Gun	Topic	Pro-Gun	Full	Pro-Gun	Topic	Pro-Gun	Full
Observations	146,445	400,205	109,710	288,039	146,445	400,205	185,051	494,359
R^2	0.1057	0.0909	0.0346	0.0314	0.0943	0.1163	0.0286	0.0190

Notes. - The table reports the OLS estimates from specifications 3.1 and 3.2. The dependent variable *Disaster Topic* is an indicator equal to 1 if highest assigned topic of a tweet is labelled "Hurricane, Storm, Prep", i.e. topic 45 from Table A.3.3, and 0 otherwise. The dependent variable *News Similarity* refers to cosine similarity between tweets and evening news. The dependent variable *Shoot-Topic* is an indicator equal to 1 speaks about the mass shooting, that is the topic with the highest probability from the topicmodel deals with the shooting, and 0 otherwise. The dependent variable *FN* is an indicator equal to 1 if a tweet is predicted as fake news (FN) with a probability o more than 90% from the neural network, and 0 otherwise. The variable Hurricane takes a value of 1 if a shooting co-occurs with a hurricane, and 0 otherwise. All columns include politician and mass shooting fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Bibliography

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?" National Bureau of Economic Research.

Abel, Wilhelm, et al. 1966. "Agrarkrisen und Agrarkonjunktur."

Acemoglu, Daron, and Alexander Wolitzky. 2011. "The economics of labor coercion." *Econometrica,* 79(2): 555–600.

Alesina, Alberto, and Nicola Fuchs-Schündeln. 2007. "Goodbye Lenin (or not?): The effect of communism on people's preferences." *American Economic Review,* 97(4): 1507–1528.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2018. "Immigration and redistribution." National Bureau of Economic Research.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2020. "The polarization of reality." Vol. 110, 324–28.

Alesina, Alberto, Edward Glaeser, and Edward Ludwig Glaeser. 2004. *Fighting poverty in the US and Europe: A world of difference.* Oxford University Press.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the origins of gender roles: Women and the plough." *The Quarterly Journal of Economics,* 128(2): 469–530.

Algan, Yann, and Pierre Cahuc. 2010. "Inherited trust and growth." *American Economic Review,* 100(5): 2060–92.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social media and fake news in the 2016 election." *Journal of economic perspectives,* 31(2): 211–36.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. "The welfare effects of social media." *American Economic Review,* 110(3): 629–76.

Ashraf, Quamrul H, Francesco Cinnirella, Oded Galor, Boris Gershman, and Erik Hornung. 2018. "Capital-skill complementarity and the emergence of labor emancipation."

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg. 2018. "Learning about voter rationality." *American Journal of Political Science*, 62(1): 37–54.

Aston, Trevor Henry, Trevor Henry Aston, and Charles HE Philpin. 1987. *The Brenner debate: agrarian class structure and economic development in pre-industrial Europe.* Vol. 1, Cambridge University Press.

Baker, Scott R, Aniket Baksy, Nicholas Bloom, Steven J Davis, and Jonathan A Rodden. 2020. "Elections, political polarization, and economic uncertainty." National Bureau of Economic Research.

Baker, Scott R, Nicholas Bloom, Brandice Canes-Wrone, Steven J Davis, and Jonathan Rodden. 2014. "Why has US policy uncertainty risen since 1960?" *American Economic Review*, 104(5): 56–60.

Banda, Kevin K, and John Cluverius. 2018. "Elite polarization, party extremity, and affective polarization." *Electoral Studies*, 56: 90–101.

Barberá, Pablo, Andreu Casas, Jonathan Nagler, Patrick J Egan, Richard Bonneau, John T Jost, and Joshua A Tucker. 2019. "Who leads? Who follows? Measuring issue attention and agenda setting by legislators and the mass public using social media data." *American Political Science Review*, 113(4): 883–901.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya. 2020. "Facts, alternative facts, and fact checking in times of post-truth politics." *Journal of Public Economics*, 182: 104123.

Becker, Sascha O, Francesco Cinnirella, Erik Hornung, and Ludger Woessmann. 2014. "iPEHD—the ifo Prussian economic history database." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 47(2): 57–66.

Becker, Sascha O, Katrin Boeckh, Christa Hainz, and Ludger Woessmann. 2016. "The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy." *The Economic Journal*, 126(590): 40–74.

Belloc, Marianna, and Samuel Bowles. 2013. "The persistence of inferior cultural-institutional conventions." *American Economic Review*, 103(3): 93–98.

Besley, Timothy. 2005. "Political selection." *Journal of Economic perspectives*, 19(3): 43–60.

Besley, Timothy, and Robin Burgess. 2002. "The political economy of government responsiveness: Theory and evidence from India." *The quarterly journal of economics*, 117(4): 1415–1451.

Bessone, Pedro, Filipe Campante, Claudio Ferraz, and PC Souza. 2020. "Internet, Social Media, and the Behavior of Politicians: Evidence from Facebook in Brazil." Working Paper, Massachusetts Institute of Technology, Cambridge.

Bigoni, Maria, Stefania Bortolotti, Marco Casari, Diego Gambetta, and Francesca Pancotto. 2016. "Amoral familism, social capital, or trust? The behavioural foundations of the Italian North–South divide." *The Economic Journal*, 126(594): 1318–1341.

Binder, Sarah A. 2004. *Stalemate: Causes and consequences of legislative gridlock.* Brookings Institution Press.

Binswanger, Hans P, and Mark R Rosenzweig. 1986. "Behavioural and material determinants of production relations in agriculture." *The Journal of Development Studies*, 22(3): 503–539.

Bloch, Marc. 2015. *Land and Work in Mediaeval Europe (Routledge Revivals): Selected Papers.* Routledge.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "The organization of firms across countries." *The quarterly journal of economics*, 127(4): 1663–1705.

Blum, Jerome. 1957. "The rise of serfdom in Eastern Europe." *The American Historical Review*, 62(4): 807–836.

Bonica, Adam. 2014. "Mapping the ideological marketplace." *American Journal of Political Science*, 58(2): 367–386.

Bowles, Samuel. 1998. "Endogenous preferences: The cultural consequences of markets and other economic institutions." *Journal of economic literature*, 36(1): 75–111.

Bowman, Shearer Davis. 1980. "Antebellum Planters and Vormärz Junkers in Comparative Perspective." *The American Historical Review*, 779–808.

Bowman, Shearer Davis. 1993. *Masters and Lords: Mid-19th-Century US Planters and Prussian Junkers.* Oxford University Press.

Boxell, Levi, Matthew Gentzkow, and Jesse M Shapiro. 2020. "Cross-country trends in affective polarization." National Bureau of Economic Research.

Boyd, Robert, and Peter J Richerson. 1988. *Culture and the evolutionary process.* University of Chicago press.

Boyd, Robert, and Peter J Richerson. 2005. *The origin and evolution of cultures.* Oxford University Press.

Brady, William J, Julian A Wills, John T Jost, Joshua A Tucker, and Jay J Van Bavel. 2017. "Emotion shapes the diffusion of moralized content in social networks." *Proceedings of the National Academy of Sciences,* 114(28): 7313–7318.

Brenner, Robert. 1976. "Agrarian class structure and economic development in pre-industrial Europe." *Past & present,* 70(1): 30–75.

Buggle, Johannes, and Ruben Durante. 2017. "Climate risk, cooperation, and the co-evolution of culture and institutions."

Buggle, Johannes C, and Steven Nafziger. 2018. "The slow road from serfdom: labor coercion and long-run development in the former Russian Empire."

Bullock, John G, and Gabriel Lenz. 2019. "Partisan bias in surveys." *Annual Review of Political Science,* 22: 325–342.

Cagé, Julia, Nicolas Hervé, and Béatrice Mazoyer. 2020. "Social media and newsroom production decisions." *Available at SSRN 3663899.*

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

Canen, Nathan J, Chad Kendall, and Francesco Trebbi. 2021. "Political Parties as Drivers of US Polarization: 1927-2018." National Bureau of Economic Research.

Carsten, Francis L. 1941. "Slavs in North-Eastern Germany." *The Economic History Review,* 11(1): 61–76.

Carsten, Francis Ludwig. 1947. “The origins of the junkers.” *The English Historical Review*, 62(243): 145–178.

Center, Pew Research. 2015. “Americans’ Internet Access: 2000-2015.”

Center, Pew Research. 2019. “Sizing Up Twitter Users.”

Cesarini, David, Christopher T Dawes, James H Fowler, Magnus Johannesson, Paul Lichtenstein, and Björn Wallace. 2008. “Heritability of cooperative behavior in the trust game.” *Proceedings of the National Academy of sciences*, 105(10): 3721–3726.

Cinnirella, Francesco, and Erik Hornung. 2016. “Landownership concentration and the expansion of education.” *Journal of Development Economics*, 121: 135–152.

Clinton, Joshua D, Ira Katznelson, and John S Lapinski. 2016. “Where Measures Meet History.” *Governing in a Polarized Age: Elections, Parties, and Political Representation in America, ed. Alan S. Gerber, and Eric Schickler. Cambridge University Press*, 191–222.

Daniele, Gianmarco, Sergio Galletta, and Benny Geys. 2020. “Abandon ship? Party brands and politicians’ responses to a political scandal.” *Journal of Public Economics*, 184: 104172.

de La Blache, Paul Vidal. 1894. *Atlas général d’histoire et géographie.* Colin.

DellaVigna, Stefano, and Matthew Gentzkow. 2010. “Persuasion: empirical evidence.” *Annu. Rev. Econ.*, 2(1): 643–669.

Demszky, Dorottya, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. “Analyzing polarization in social media: Method and application to tweets on 21 mass shootings.” *arXiv preprint arXiv:1904.01596.*

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “Bert: Pre-training of deep bidirectional transformers for language understanding.” *arXiv preprint arXiv:1810.04805.*

Dicker, Rachel. 2016. “Avoid These Fake News Sites at All Costs.” *https://www.usnews.com/news/national-news/articles/2016-11-14/avoid-these-fake-news-sites-at-all-costs.*

Djourelova, Milena, and Ruben Durante. 2019. “Media attention and strategic timing in politics: Evidence from US presidential executive orders.”

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2011. "The intergenerational transmission of risk and trust attitudes." *The Review of Economic Studies*, 79(2): 645–677.

Domar, Evsey D. 1970. "The Causes of Slavery or Serfdom: A Hypothesis." *The Journal of Economic History*, 18–32.

Druckman, James, and Jeremy Levy. 2021. "Affective Polarization in the American Public."

Druckman, James N, Erik Peterson, and Rune Slothuus. 2013. "How elite partisan polarization affects public opinion formation." *American Political Science Review*, 107(1): 57–79.

Duwe, Grant. 2016. "of Mass Public Shootings in the United States, 1915–2013." *The Wiley handbook of the psychology of mass shootings*, 20.

Eddie, Sean A. 2013. *Freedom's price: serfdom, subjection, and reform in Prussia, 1648-1848.* OUP Oxford.

Eisensee, Thomas, and David Strömberg. 2007. "News droughts, news floods, and US disaster relief." *The Quarterly Journal of Economics*, 122(2): 693–728.

Ekelund, Jr, Robert B, Robert F Hébert, and Robert D Tollison. 2002. "An economic analysis of the protestant reformation." *Journal of Political Economy*, 110(3): 646–671.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global evidence on economic preferences." *The Quarterly Journal of Economics*, 133(4): 1645–1692.

Fernández, Raquel. 2011. "Does culture matter?" In *Handbook of social economics.* Vol. 1, 481–510. Elsevier.

Ferraz, Claudio, and Frederico Finan. 2008. "Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes." *The Quarterly journal of economics*, 123(2): 703–745.

Flynn, DJ, Brendan Nyhan, and Jason Reifler. 2017. "The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics." *Political Psychology*, 38: 127–150.

Franz, Günther. 1979. *Der Dreissigjährige Krieg und das deutsche Volk: Unters. zur Bevölkerungs-u. Agrargeschichte.* Vol. 7, Lucius & Lucius DE.

Galor, Oded, and Ömer Özak. 2016. "The agricultural origins of time preference." *American Economic Review*, 106(10): 3064–3103.

Garz, Marcel, and Jil Sörensen. 2017. "Politicians under investigation: The news Media's effect on the likelihood of resignation." *Journal of Public Economics*, 153: 82–91.

Gennaro, Gloria, and Elliott Ash. 2021. "Emotion and Reason in Political Language." *Center for Law & Economics Working Paper Series*, 2021(02).

Gennaro, Gloria, and Elliott Ash. 2022. "Emotion and reason in political language." *The Economic Journal*, 132(643): 1037–1059.

Gennaro, Gloria, Giampaolo Lecce, Massimo Morelli, et al. 2021. "Mobilization and the Strategy of Populism Theory and Evidence from the United States." CEPR Discussion Papers.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as data." *Journal of Economic Literature*, 57(3): 535–74.

Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy. 2019. "Measuring group differences in high-dimensional choices: method and application to congressional speech." *Econometrica*, 87(4): 1307–1340.

Giavazzi, Francesco, Felix Iglhaut, Giacomo Lemoli, and Gaia Rubera. 2020. "Terrorist attacks, cultural incidents and the vote for radical parties: Analyzing text from twitter." National Bureau of Economic Research.

Glaeser, Edward L, David I Laibson, Jose A Scheinkman, and Christine L Soutter. 2000. "Measuring trust." *The quarterly journal of economics*, 115(3): 811–846.

Goldberg, Yoav. 2016. "A primer on neural network models for natural language processing." *Journal of Artificial Intelligence Research*, 57: 345–420.

Green, Matthew N. 2015. "Underdog Politics." In *Underdog Politics.* Yale University Press.

Groseclose, Tim, and Jeffrey Milyo. 2005. "A measure of media bias." *The Quarterly Journal of Economics*, 120(4): 1191–1237.

Grossman, Gene M, and Elhanan Helpman. 2019. "Electoral Competition with Fake News." National Bureau of Economic Research.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. 2018. "Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign." *European Research Council*, 9(3): 4.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2004. "Does local financial development matter?" *The Quarterly Journal of Economics*, 119(3): 929–969.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural biases in economic exchange?" *The quarterly journal of economics*, 124(3): 1095–1131.

Harnisch, Hartmut. 1972. "Zur Herausbildung und Funktionsweise von Gutswirtschaft und Gutsherrschaft. Eine Klageschrift der Bauern der Herrschaft Neugattersleben aus dem Jahre 1610." *Jahrbuch für Regionalgeschichte*, 4: 179–199.

Harnisch, Hartmut. 1984. "Kapitalistische Agrarreform und industrielle Revolution." *Agrarhistorische Untersuchungen über das ostelbische Preußen*, 43–45.

Harnisch, Hartmut. 1986. "Peasants and markets: the background to the agrarian reforms in feudal Prussia east of the Elbe, 1760–1807." *The German Peasantry: Conflict and Community in Rural Society from the Eighteenth to the Twentieth Centuries*, 37–70.

Helmstetter, Stefan, and Heiko Paulheim. 2018. "Weakly supervised learning for fake news detection on Twitter." 274–277, IEEE.

Henrich, Joseph, Jean Ensminger, Richard McElreath, Abigail Barr, Clark Barrett, Alexander Bolyanatz, Juan Camilo Cardenas, Michael Gurven, Edwins Gwako, Natalie Henrich, et al. 2010. "Markets, religion, community size, and the evolution of fairness and punishment." *science*, 327(5972): 1480–1484.

Hetherington, Marc J, and Thomas J Rudolph. 2015. *Why Washington won't work.* University of Chicago Press.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "LSTM can solve hard long time lag problems." *Advances in neural information processing systems*, 473–479.

Huszár, Ferenc, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. "Algorithmic amplification of politics on Twitter." *Proceedings of the National Academy of Sciences*, 119(1).

Iyengar, Shanto, and Sean J Westwood. 2015. "Fear and loathing across party lines: New evidence on group polarization." *American Journal of Political Science*, 59(3): 690–707.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, not ideologya social identity perspective on polarization." *Public opinion quarterly*, 76(3): 405–431.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J Westwood. 2019. "The origins and consequences of affective polarization in the United States." *Annual Review of Political Science*, 22: 129–146.

Jerit, Jennifer, James H Kuklinski, and Paul J Quirk. 2009. "Strategic politicians, emotional citizens, and the rhetoric of prediction." *The political psychology of democratic citizenship*, 100–124.

Jetter, Michael, and Jay Walker. 2018. "The effect of media coverage on mass shootings."

Jordan-Rozwadowski, Jan v. 1900. "Die Bauern des 18. Jahrhunderts und ihre Herren im Lichte der neuesten deutschen Forschungen." *Jahrbücher für Nationalökonomie und Statistik*, 75(1): 478–514.

Kaak, Heinrich. 2013. *Die Gutsherrschaft: theoriegeschichtliche Untersuchungen zum Agrarwesen im ostelbischen Raum.* Vol. 79, Walter de Gruyter.

Kaplan, Ethan, Jörg L Spenkuch, and Haishan Yuan. 2018. "Natural Disasters, Moral Hazard, and Special Interests in Congress."

Kirkland, Justin H, and Jonathan B Slapin. 2018. *Roll call rebels: strategic dissent in the United States and United Kingdom.* Cambridge University Press.

Klein, Alexander, and Sheilagh Ogilvie. 2017. "Was Domar Right? Serfdom and Factor Endowments in Bohemia."

Knapp, Georg Friedrich. 1887. *Die Bauern-Befreiung und der Ursprung der Landarbeiter in den älteren Theilen Preußens.* Vol. 2, Duncker & Humblot.

Krouse, WJ, and DJ Richardson. 2015. "Mass murder with firearms: Incidents and Victims, 1999 - 2013." *Washington, DC: US Congressional Research Service.*

Kuklinski, James H, Paul J Quirk, Jennifer Jerit, David Schwieder, and Robert F Rich. 2000. "Misinformation and the currency of democratic citizenship." *Journal of Politics*, 62(3): 790–816.

Lau, Richard R, David J Andersen, Tessa M Ditonto, Mona S Kleinberg, and David P Redlawsk. 2017. "Effect of media environment diversity and advertising tone on information search, selective exposure, and affective polarization." *Political Behavior*, 39(1): 231–255.

Lee, Frances E. 2008. "Agreeing to disagree: Agenda content and senate partisanship, 1981–2004." *Legislative Studies Quarterly*, 33(2): 199–222.

Levitsky, Steven, and Daniel Ziblatt. 2018. *How democracies die.* Broadway Books.

Lewandowsky, Stephan, Michael Jetter, and Ullrich KH Ecker. 2020. "Using the president's tweets to understand political diversion in the age of social media." *Nature communications*, 11(1): 1–12.

Luca, Michael, Deepak Malhotra, and Christopher Poliquin. 2020. "The impact of mass shootings on gun policy." *Journal of public economics*, 181: 104083.

Lupu, Noam. 2015. "Party polarization and mass partisanship: A comparative perspective." *Political Behavior*, 37(2): 331–356.

Markevich, Andrei, and Ekaterina Zhuravskaya. 2018. "The economic effects of the abolition of serfdom: Evidence from the Russian Empire." *American Economic Review*, 108(4-5): 1074–1117.

Martin, Gregory J, and Ali Yurukoglu. 2017. "Bias in cable news: Persuasion and polarization." *American Economic Review*, 107(9): 2565–99.

McCarty, Nolan. 2019. *Polarization: What everyone needs to know®.* Oxford University Press.

McCombs, Maxwell E, and Donald L Shaw. 1972. "The agenda-setting function of mass media." *Public opinion quarterly*, 36(2): 176–187.

Meitzen, A. 1866. "Der Boden und die landwirtschaftlichen Ver-hältnisse des preussischen Staates. 4 Bände in 4." *Gebietsumfang vor*, 1868–69.

Melton, Edgar. 2000. "The Transformation of the Rural Economy in East Elbian Prussia, 1750–1830." *The Rise of Prussia*, 1700–1830.

Metzl, Jonathan M, and Kenneth T MacLeish. 2015. "Mental illness, mass shootings, and the politics of American firearms." *American journal of public health*, 105(2): 240–249.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2014. "Resolving debt overhang: Political constraints in the aftermath of financial crises." *American Economic Journal: Macroeconomics*, 6(2): 1–28.

Moskowitz, Daniel J, Jon C Rogowski, and James M Snyder. 2018. "Parsing Party Polarization in Congress."

Müller, Karsten, and Carlo Schwarz. 2020. "From hashtag to hate crime: Twitter and anti-minority sentiment." *Available at SSRN 3149103*.

Naef, Michael, and Jürgen Schupp. 2009. "Measuring trust: Experiments and surveys in contrast and combination."

Nafziger, Steven. 2012. "Serfdom, emancipation, and off-farm labour mobility in Tsarist Russia." *Economic History of Developing Regions*, 27(1): 1–37.

Nunn, Nathan, and Leonard Wantchekon. 2011. "The slave trade and the origins of mistrust in Africa." *American Economic Review*, 101(7): 3221–52.

Ogilvie, Sheilagh. 2014. "Serfdom and the institutional system in early modern Germany." *Slavery and Serfdom in the European Economy from the 11th to the 18th Centuries*, 33–58.

Ondercin, Heather Louise, and Mary Kate Lizotte. 2021. "You've lost that loving feeling: How gender shapes affective polarization." *American Politics Research*, 49(3): 282–292.

Pennington, Jeffrey, Richard Socher, and Christopher D Manning. 2014. "Glove: Global vectors for word representation." 1532–1543.

Persson, Torsten, and Guido Enrico Tabellini. 2002. *Political economics: explaining economic policy.* MIT press.

Petrova, Maria, Ananya Sen, and Pinar Yildirim. 2021. "Social media and political contributions: The impact of new technology on political competition." *Management Science*, 67(5): 2997–3021.

Pierenkemper, Toni, and Richard H Tilly. 2004. *The German economy during the nineteenth century.* Berghahn Books.

Polborn, Mattias K, and James M Snyder Jr. 2017. "Party polarization in legislatures with office-motivated candidates." *The Quarterly Journal of Economics*, 132(3): 1509–1550.

Prior, Markus, Gaurav Sood, Kabir Khanna, et al. 2015. "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions." *Quarterly Journal of Political Science,* 10(4): 489–518.

Proksch, Sven-Oliver, and Jonathan B Slapin. 2012. "Institutional foundations of legislative speech." *American Journal of Political Science,* 56(3): 520–537.

Putnam, Robert D. 2000. "Bowling alone: America's declining social capital." In *Culture and politics.* 223–234. Springer.

Rathje, Steve, Jay J Van Bavel, and Sander van der Linden. 2021. "Out-group animosity drives engagement on social media." *Proceedings of the National Academy of Sciences,* 118(26).

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti. 2013*a*. "Seeds of distrust: Conflict in Uganda." *Journal of Economic Growth,* 18(3): 217–252.

Rohner, Dominic, Mathias Thoenig, and Fabrizio Zilibotti. 2013*b*. "War signals: A theory of trade, trust, and conflict." *Review of Economic Studies,* 80(3): 1114–1147.

Rosenberg, Hans. 1943. "The Rise of the Junkers in Brandenburg-Prussia, 1410-1653: Part 1." *The American Historical Review,* 49(1): 1–22.

Rosenberg, Hans. 1944. "The Rise of the Junkers in Brandenburg-Prussia, 1410-1563: Part II." *The American Historical Review,* 49(2): 228–242.

Schilling, Heinz. 1989. *Höfe und Allianzen: Deutschland 1648-1763.* Vol. 6, Siedler.

Schupp, Jrgen, and Gert G Wagner. 2002. "Maintenance of and innovation in long-term panel studies: The case of the German Socio-Economic Panel (GSOEP)." *AStA Advances in Statistical Analysis,* 2(86): 163–175.

Snyder Jr, James M, and David Strömberg. 2010. "Press coverage and political accountability." *Journal of political Economy,* 118(2): 355–408.

Stock, James H, and Motohiro Yogo. 2002. "Testing for weak instruments in linear IV regression." National Bureau of Economic Research.

Sunstein, Cass. 2018. *# Republic - Divided Democracy in the Age of Social Media.* Princeton University Press.

Tabellini, Guido. 2008*a*. "Institutions and culture." *Journal of the European Economic Association*, 6(2-3): 255–294.

Tabellini, Guido. 2008*b*. "The scope of cooperation: Values and incentives." *The Quarterly Journal of Economics*, 123(3): 905–950.

Taddy, Matt. 2015. "Distributed multinomial regression." *The Annals of Applied Statistics*, 9(3): 1394–1414.

Vasold, Manfred. 1993. "Die deutschen Bevölkerungsverluste während des dreißigjährigen Krieges." *Zeitschrift für bayerische Landesgeschichte*, 56: 147–160.

Voigtländer, Nico, and Hans-Joachim Voth. 2012*a*. "Persecution perpetuated: the medieval origins of anti-Semitic violence in Nazi Germany." *The Quarterly Journal of Economics*, 127(3): 1339–1392.

Voigtländer, Nico, and Hans-Joachim Voth. 2012*b*. "The three horsemen of riches: Plague, war, and urbanization in early modern Europe." *Review of Economic Studies*, 80(2): 774–811.

von Below, Georg. 1923. *Territorium und Stadt.* Verlag R. Oldenbourg.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science*, 359(6380): 1146–1151.

Whaley, Joachim. 2012. *Germany and the Holy Roman Empire: Volume I: Maximilian I to the Peace of Westphalia, 1493-1648.* Vol. 1, Oxford University Press on Demand.

Widmer, Philine, Sergio Galletta, and Elliott Ash. 2020. "Media slant is contagious." *Center for Law & Economics Working Paper Series*, 14.

Yan, Xiaohui, Jiafeng Guo, Yanyan Lan, and Xueqi Cheng. 2013. "A biterm topic model for short texts." 1445–1456.

Yousaf, Hasin. 2021. "Sticking to One's Guns: Mass Shootings and the Political Economy of Gun Control in the United States." *Journal of the European Economic Association.*

Zaller, John R, et al. 1992. *The nature and origins of mass opinion.* Cambridge university press.

Zellers, Rowan, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. "Defending against neural fake news." *arXiv preprint arXiv:1905.12616*.

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political effects of the internet and social media." *Annual Review of Economics*, 12: 415–438.

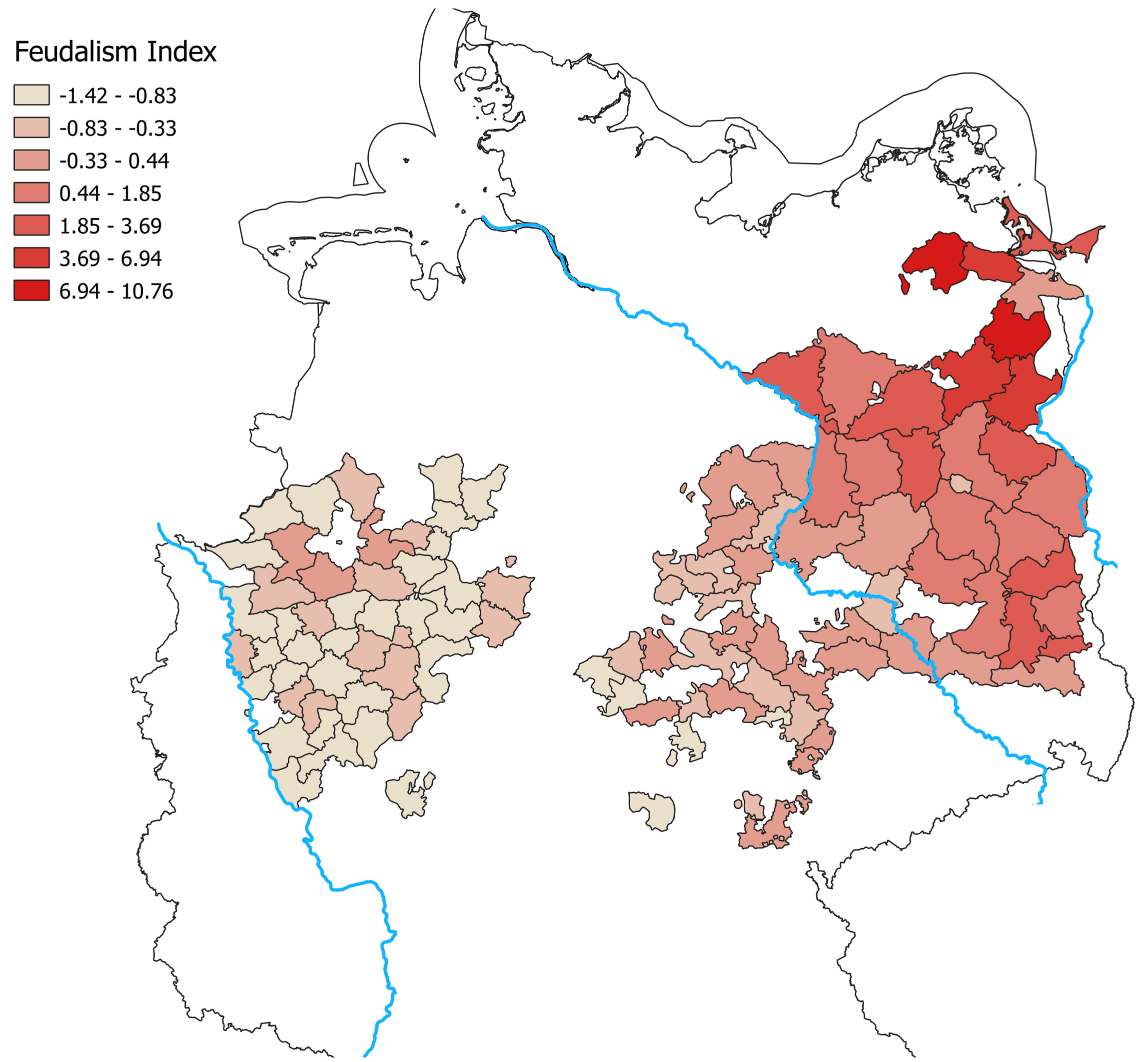
Appendix A

Appendix

A.1 Appendix Chapter 1 - Feudalism and Trust

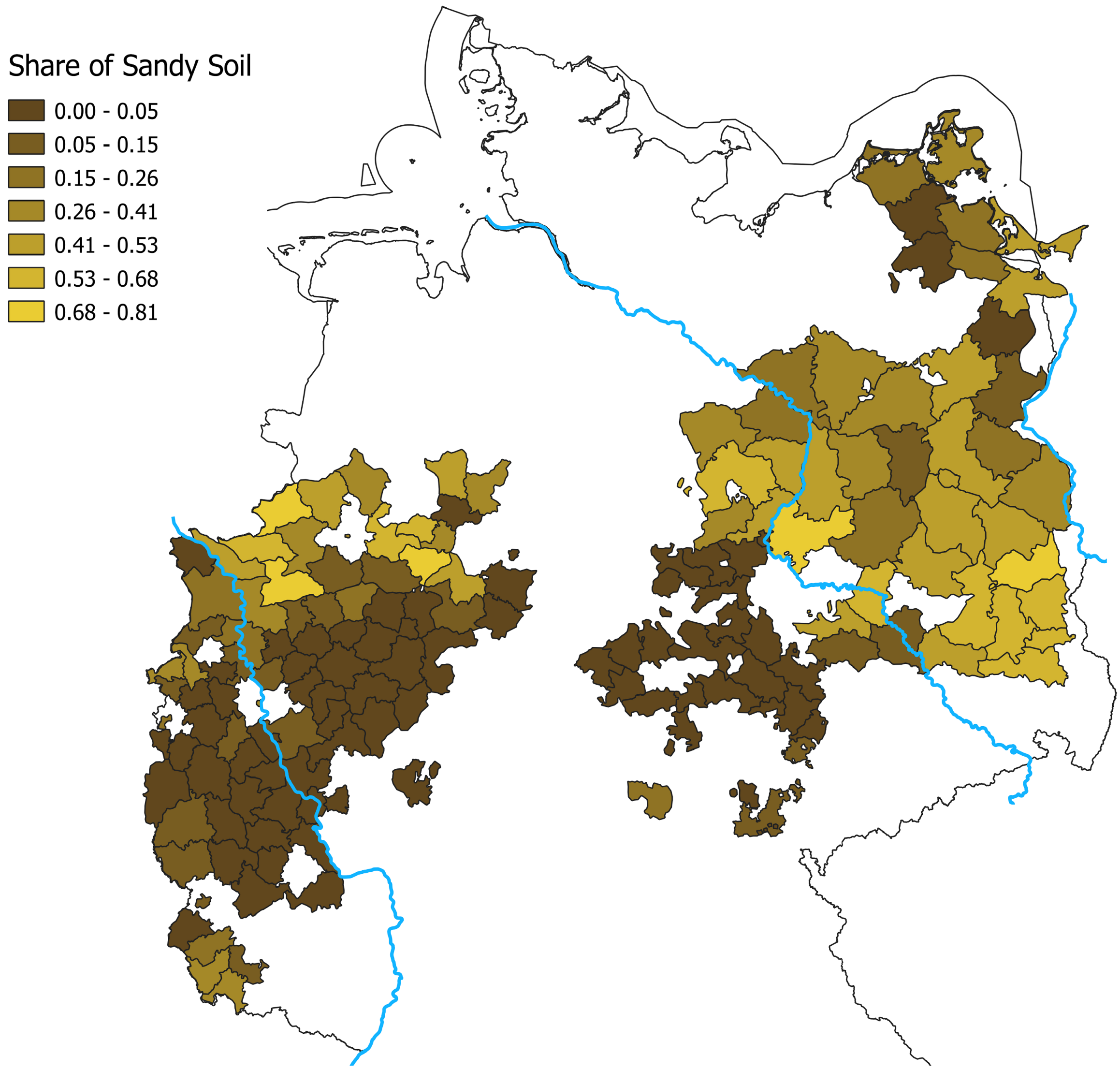
A.1.1 Figure Appendix

Figure A.1.1: Feudalism Index



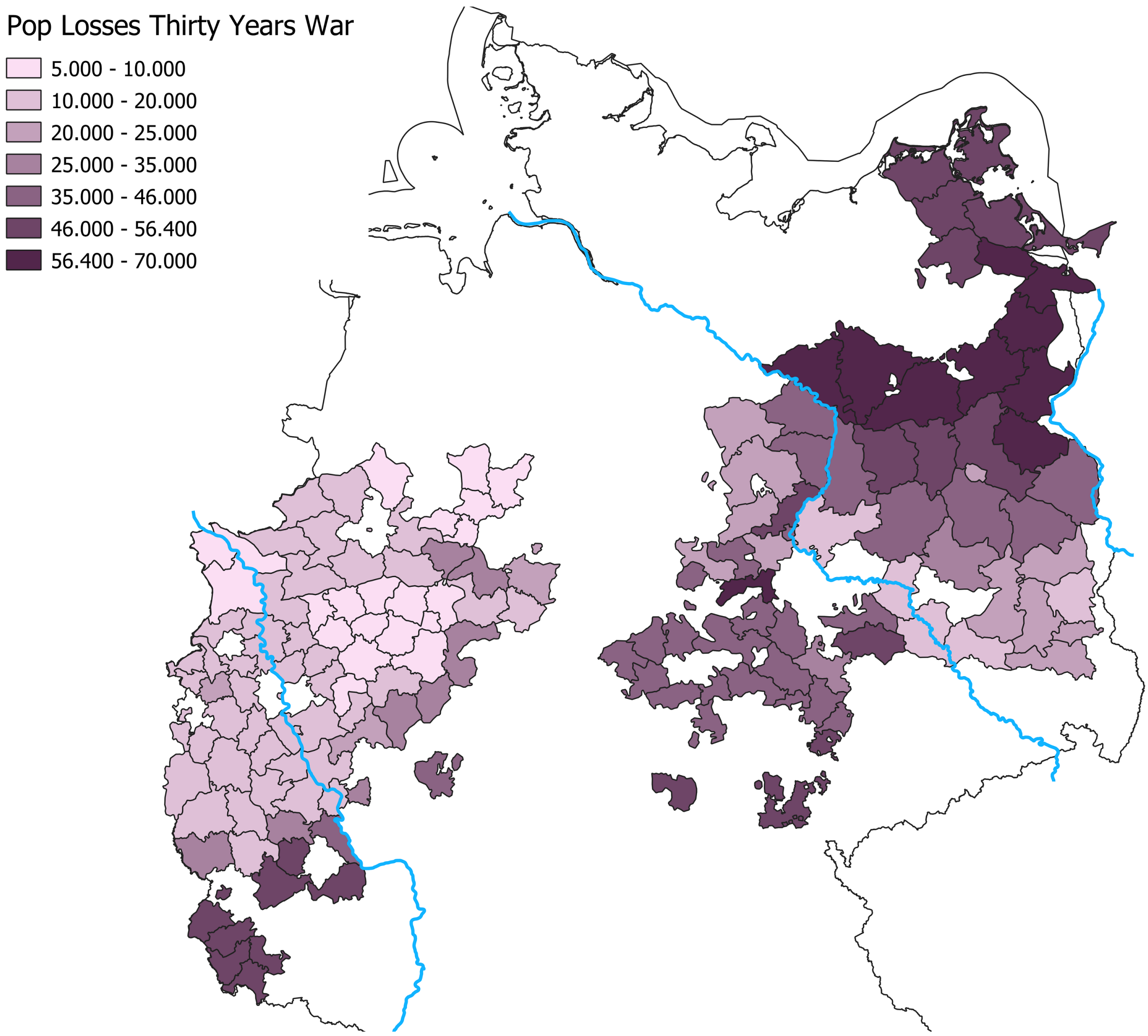
Notes. - Figure plots the spatial distribution of the main regressor, the feudalism index in Prussia in 1849. The indicator is the first principal component of the the share of large landowners, the share of knightestates, servile duties and redemption costs.

Figure A.1.2: Share of Sandy Soil



Notes. - Figure plots the spatial distribution of the share of sandy soil in Prussia in 1866. The indicator is bound between 0 and 1.

Figure A.1.3: Population Losses in the Thirty Years War (1618 - 1648) in Percent



Notes. - Figure plots the spatial distribution of the population losses during the Thirty Years War. The indicator is bound between 0 and 100.

A.1.2 Table Appendix

Table A.1.1: OLS-Estimates - Trusting Nobody

	Dependent variable:		
	Trust Nobody		
	(1)	(2)	(3)
Feudalism	0.032***	0.028***	0.027***
	(0.007)	(0.007)	(0.007)
Household Controls	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes
Fixed Effects	No	Province	State
Observations	11,961	11,961	11,959
R^2	0.071	0.077	0.074
Clusters	113	113	113

Note: The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. Household controls include gender, age, log household income, a dummy for former GDR and a full set of dummies for education, occupation, religion, employment and family status. Geographic controls include latitude, longitude, soil quality, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.

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

Table A.1.2: OLS Estimates - Survey Wave Fixed Effects

	Dependent variable:					
	General.Trust		Distrust Foreigners		Trust Nobody	
	(1)	(2)	(3)	(4)	(5)	(6)
Feudalism	-0.029***	-0.036***	0.031***	0.028***	0.027***	0.022**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.010)	(0.009)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Survey Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	No	Prov.	No	Prov.	No	Prov.
Observations	11,966	11,966	11,961	11,961	11,969	11,969
R^2	0.054	0.065	0.071	0.071	0.061	0.068
Clusters	113	113	113	113	114	114

Note: The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. Household controls include gender, age, log household income, a dummy for former GDR and a full set of dummies for education, occupation, religion, employment and family status. Geographic controls include latitude, longitude, soil quality, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.

*p<0.1; **p<0.05; ***p<0.01

Table A.1.3: OLS-Estimates - Trust Behaviour

	Dependent variable:			
	Freq. Lend Pers. Things	Freq. Lend Money	Freq. Leaving Door Unlocked	
	(1)	(2)	(3)	(4)
Feudalism	-0.040***	-0.015*	-0.014	-0.029*
	(0.007)	(0.008)	(0.014)	(0.016)
Household Controls	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes
Survey Fixed Effects	Yes	Yes	Yes	Yes
Fixed Effects	Province	Province	Province	State
Observations	11,952	11,952	11,949	11,947
R^2	0.18	0.07	0.04	0.04
Clusters	114	114	114	114

Note: The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. Household controls include gender, age, log household income, a dummy for former GDR and a full set of dummies for education, occupation, religion, employment and family status. Geographic controls include latitude, longitude, soil quality, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.

*p<0.1; **p<0.05; ***p<0.01

Table A.1.4: OLS Estimates - Defining Feudalism based on Municipality of Birth

	Dependent variable:					
	Generalized Trust		Distrust in Foreigners		Trust Nobody	
	(1)	(2)	(3)	(4)	(5)	(6)
Feudalism	-0.017***	-0.024***	0.020**	0.021**	-0.004	-0.002
	(0.006)	(0.008)	(0.008)	(0.008)	(0.006)	(0.008)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	No	Prov	No	Prov	No	Prov
Observations	7,474	7,474	7,478	7,478	7,466	7,466
R^2	0.063	0.065	0.057	0.052	0.071	0.075
Clusters	115	115	115	115	115	115

Note: The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. Household controls include gender, age, log household income, a dummy for former GDR and a full set of dummies for education, occupation, religion, employment and family status. Geographic controls include latitude, longitude, soil quality, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.

*p<0.1; **p<0.05; ***p<0.01

Table A.1.5: OLS Estimates - Excluding Religion as Covariate

	Dependent variable:					
	General. Trust		Trust Nobody		Distrust Foreigners	
	(1)	(2)	(3)	(4)	(5)	(6)
Feudalism	-0.033***	-0.037***	0.034***	0.033***	0.025***	0.017**
	(0.007)	(0.007)	(0.009)	(0.007)	(0.009)	(0.007)
Household Controls	Yes	Yes	Yes	Yes	Yes	Yes
Geographic Controls	Yes	Yes	Yes	Yes	Yes	Yes
Historical Controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	No	Prov.	No	Prov.	No	Prov.
Observations	16,228	16,228	16,225	16,225	16,234	16,234
R^2	0.055	0.061	0.079	0.084	0.084	0.089
Clusters	115	115	115	115	115	115

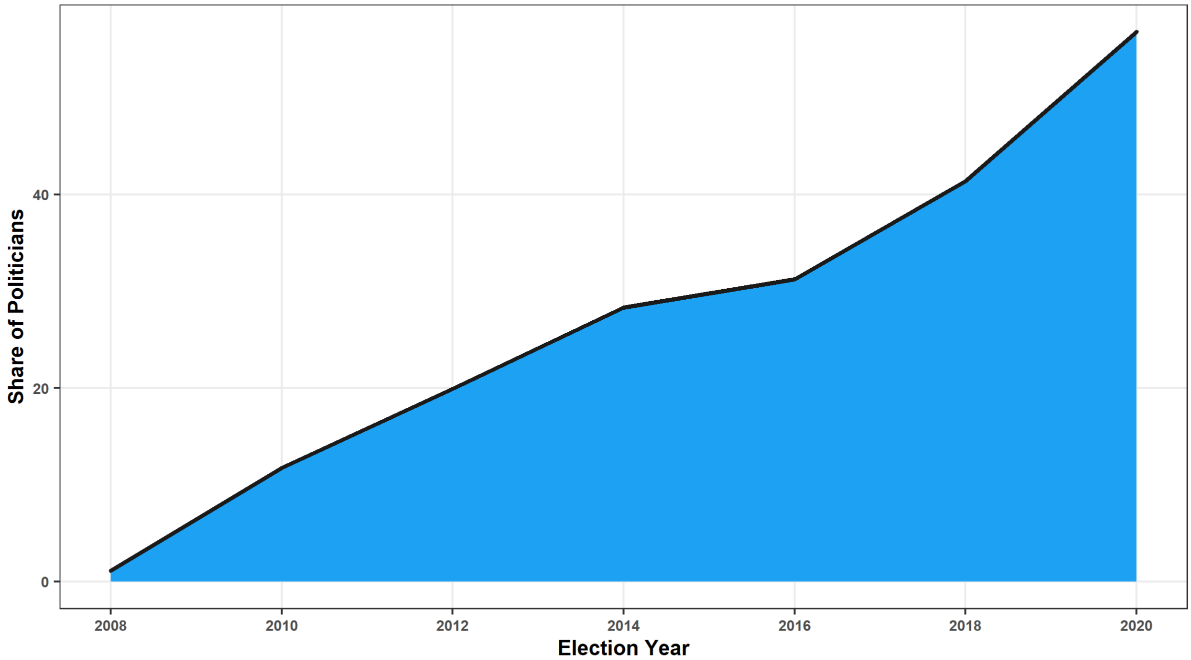
Note: The table reports OLS estimates. The unit of observation is an individual. Standard errors are clustered at the level of the historical county. Household controls include gender, age, log household income, a dummy for former GDR and a full set of dummies for education, occupation, employment and family status. Geographic controls include latitude, longitude, soil quality, % swamps, distance to province capital and indicators for east of river Elbe, having a chaussee or a city. Historical controls include the per-capita counts of non-germans, protestants, steam engines in mining and factory workers as well as population density and an indicator for the common inheritance system. Both, historical and geographical controls refer to 1849.

*p<0.1; **p<0.05; ***p<0.01

A.2 Appendix Chapter 2 - U.S. Politicians on Social Media

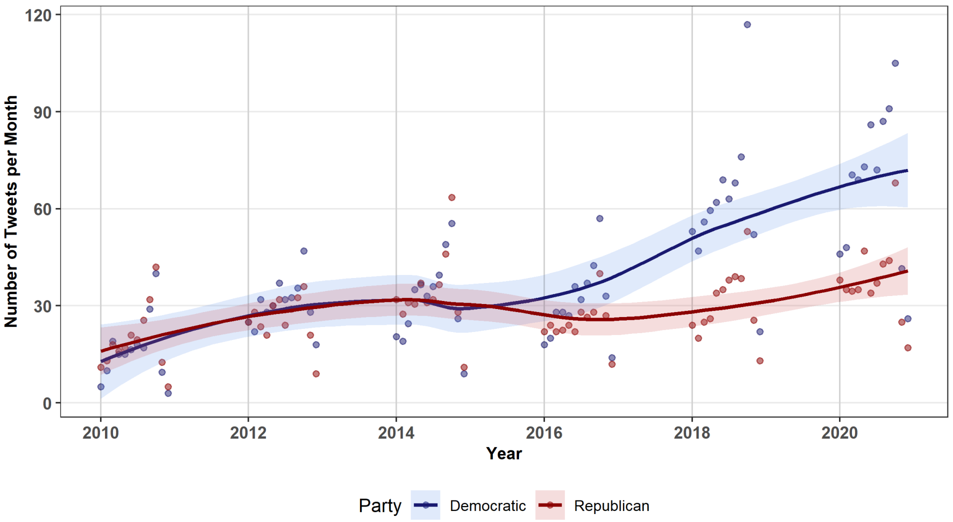
A.2.1 Figure Appendix

Figure A.2.1: Share of Politicians Using Social Media

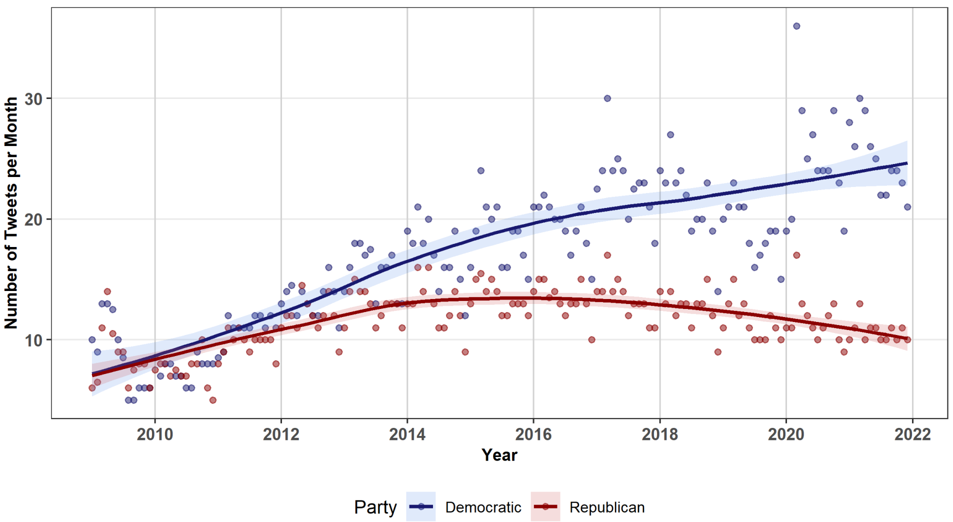


Notes. - The figure plots for federal election years the share of politicians that have a Twitter account in relation to the total number of candidates observed. Sample: Full Sample.

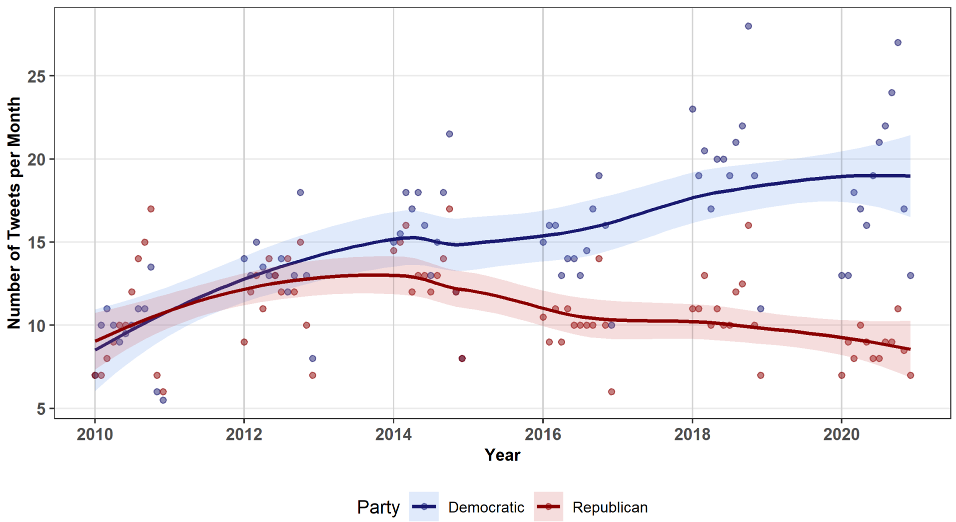
Figure A.2.2: Average Number of Tweets per Month for Median Legislator by Party



(a) Sample: Congressional Candidates



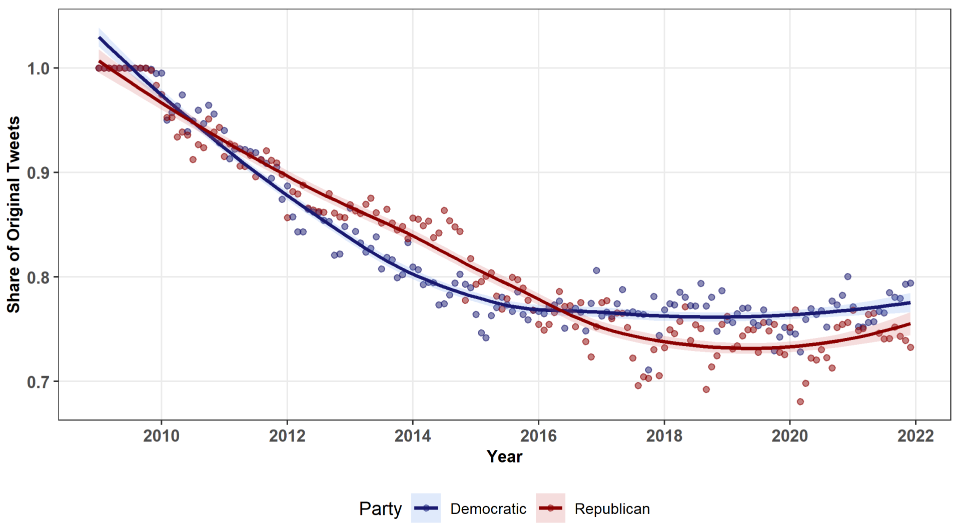
(b) Sample: State Officeholders



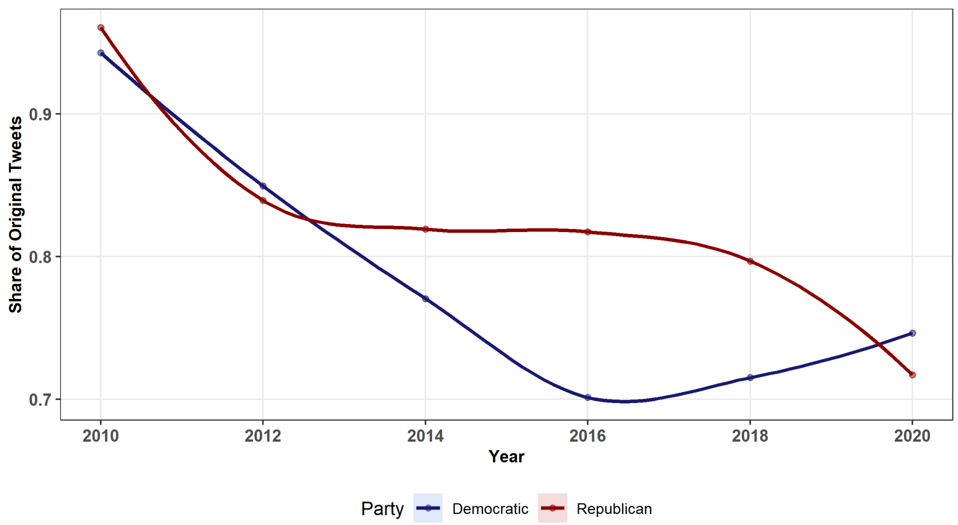
(c) Sample: Candidates for State Office

Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on tweets and retweets.

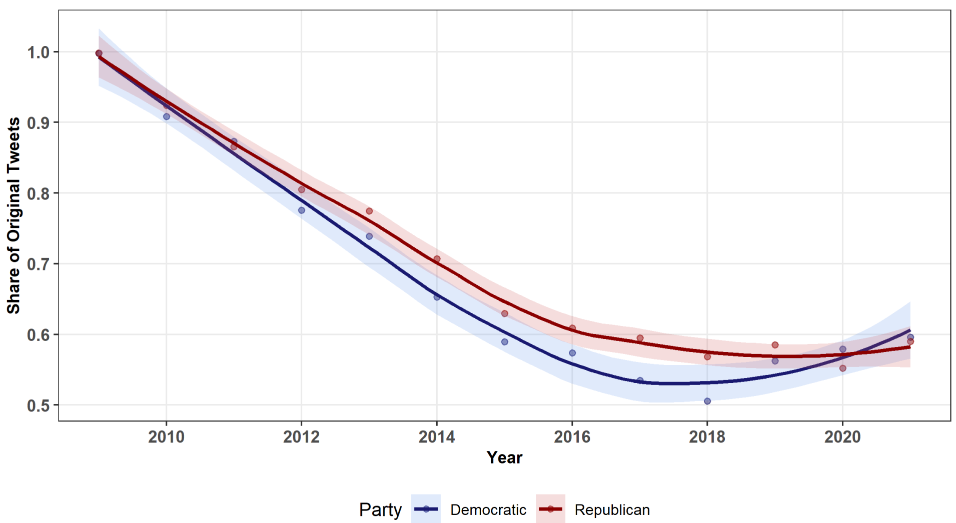
Figure A.2.3: Share of Original Tweets by Party and Year



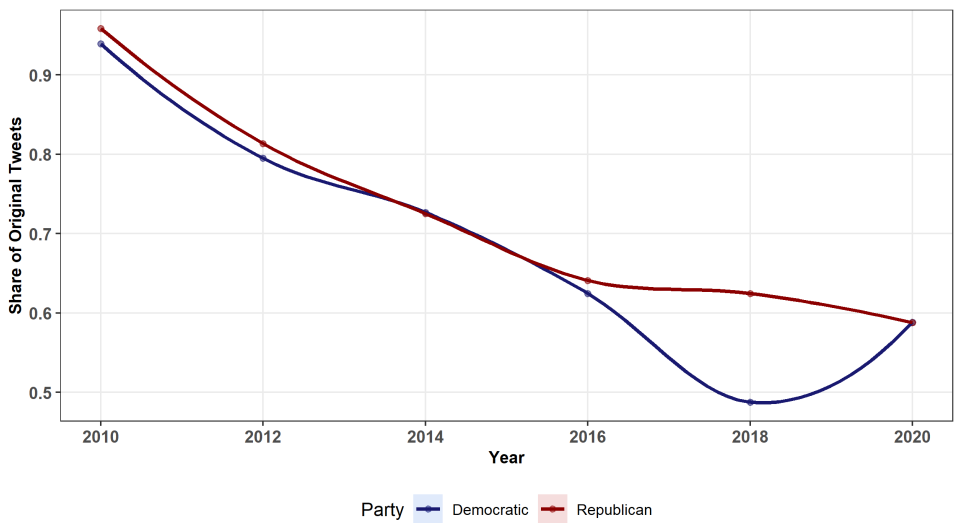
(a) Sample: Members of Congress



(b) Sample:Congressional Candidates



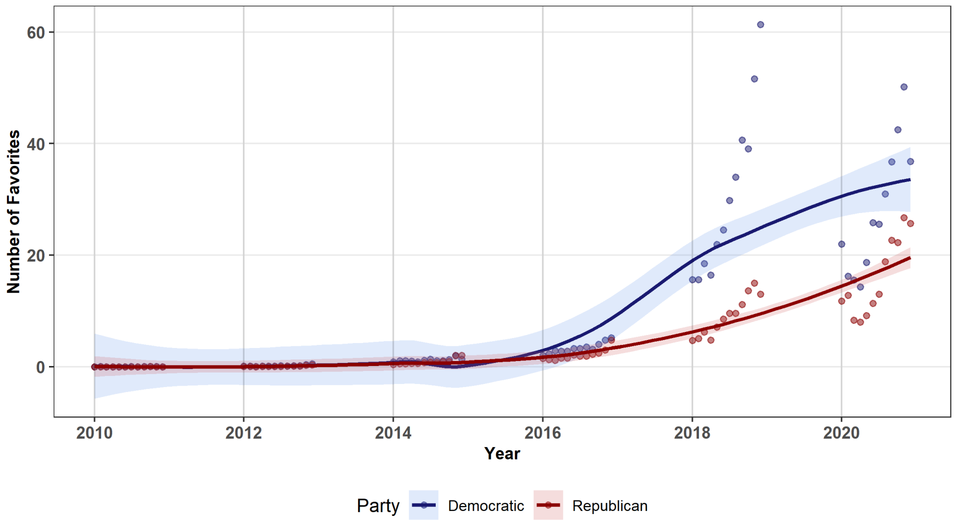
(c) Sample: State Officeholders



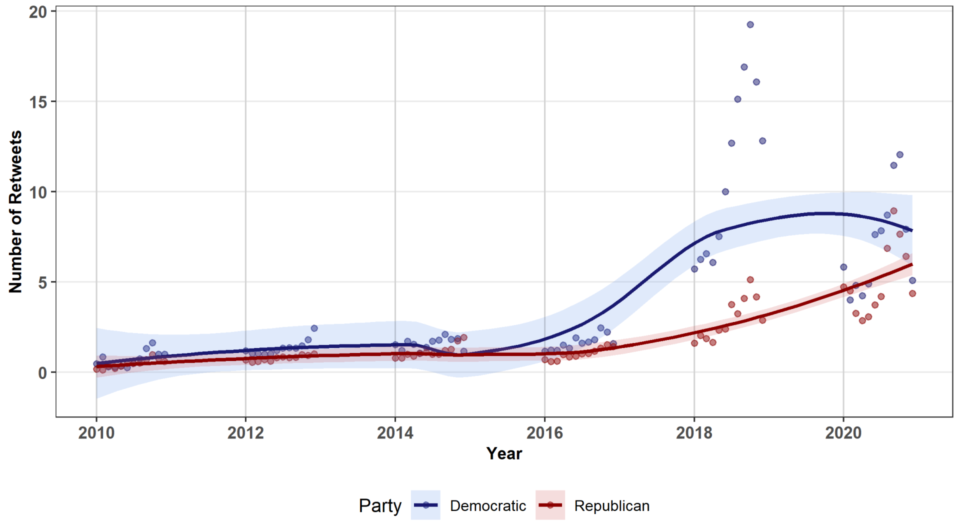
(d) Sample: Candidates for State Office

Notes. - The figure plots the proportion of original tweets to total tweets by party and year. The total tweets comprise original tweets and retweets.

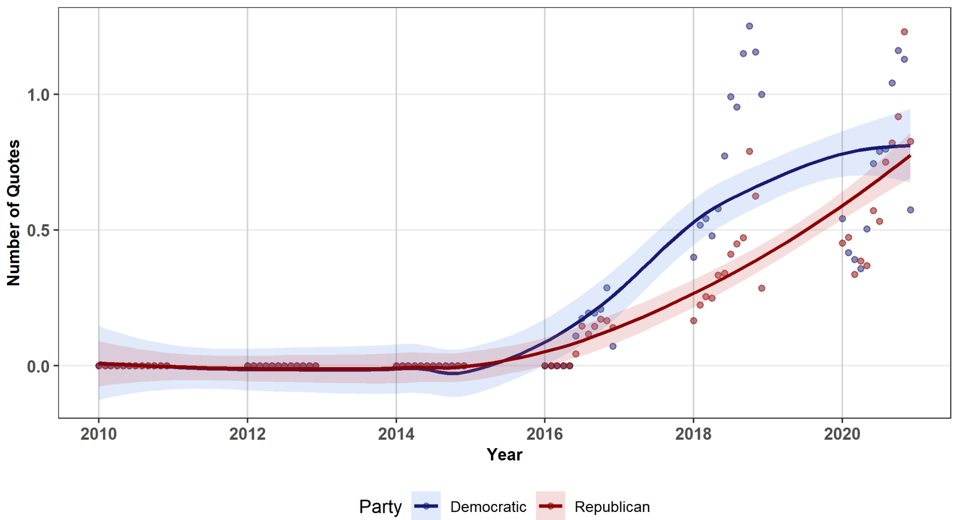
Figure A.2.4: Sample Congressional Candidates: Audience Engagement for Average Tweet of Median Legislator by Party



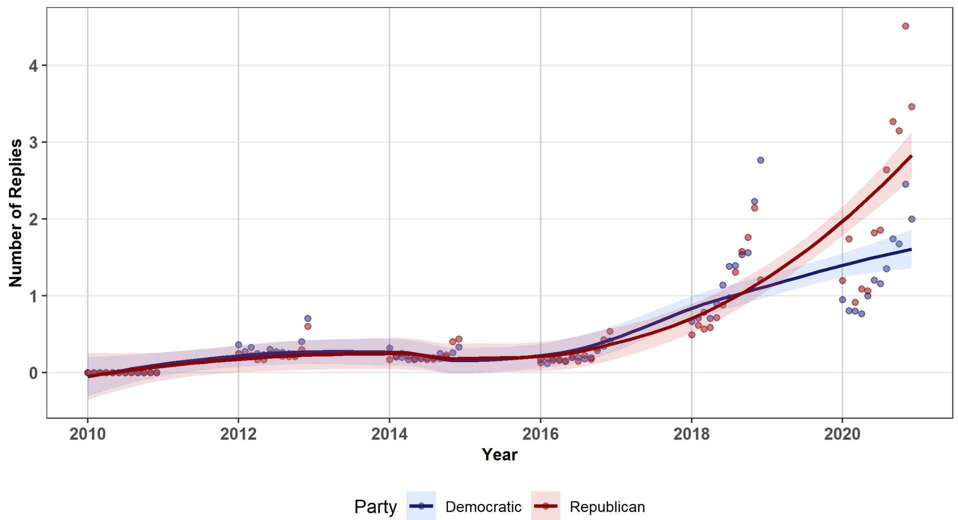
(a) Favorites



(b) Retweets



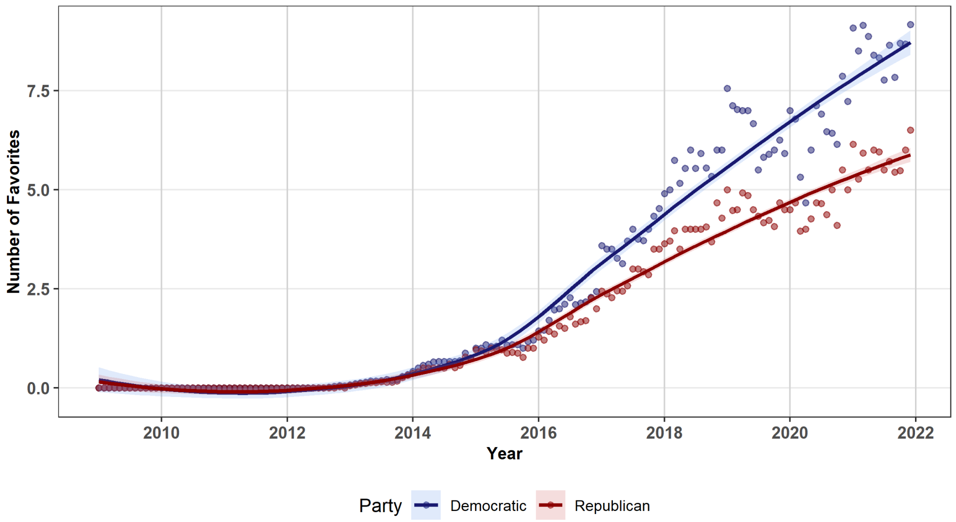
(c) Quotes



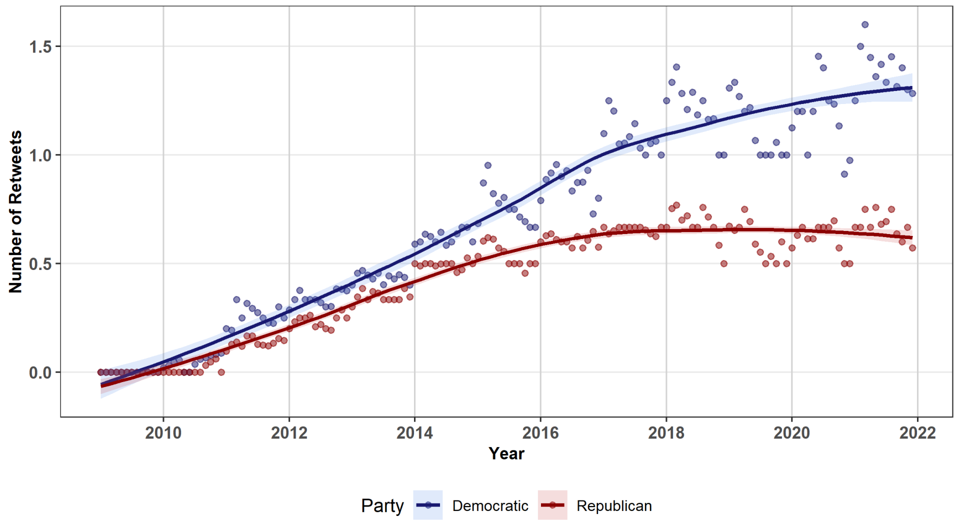
(d) Replies

Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: Congressional Candidates.

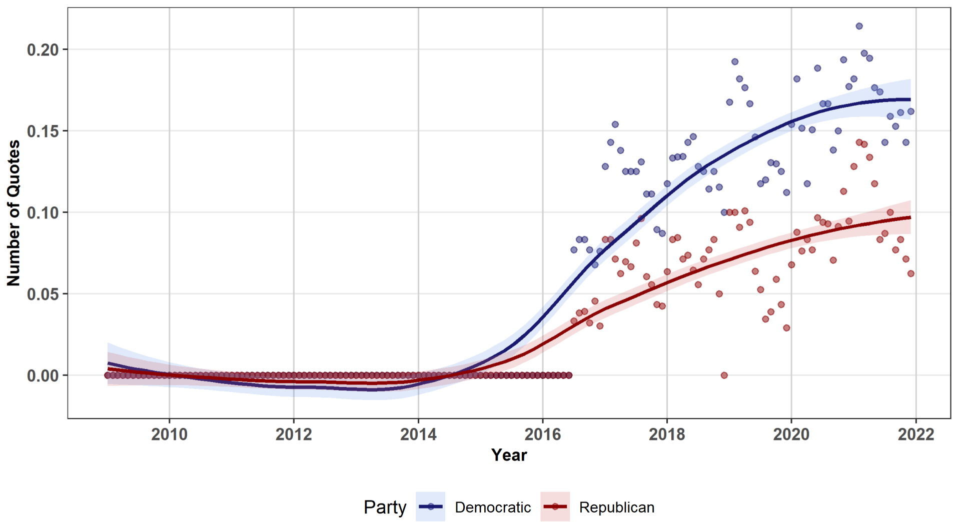
Figure A.2.5: Sample State Officeholders: Audience Engagement for Average Tweet of Median Legislator by Party



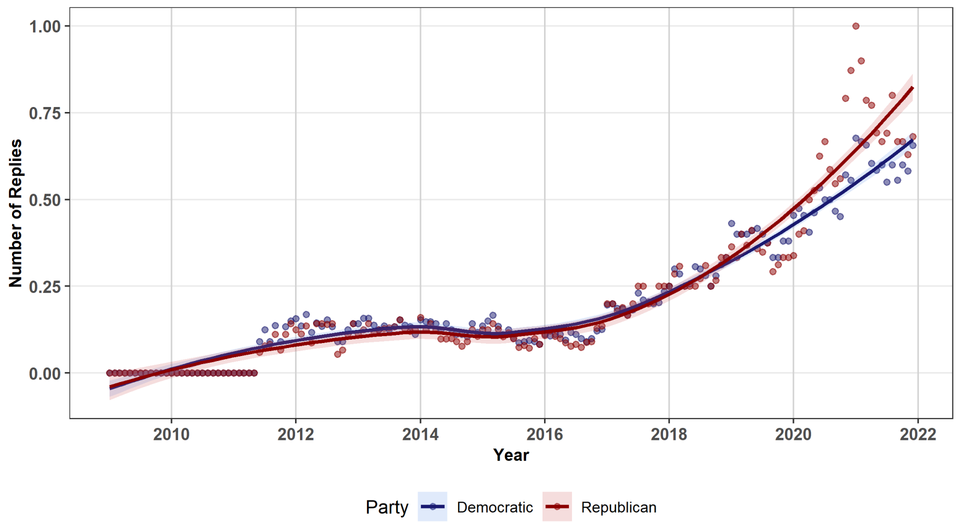
(a) Favorites



(b) Retweets



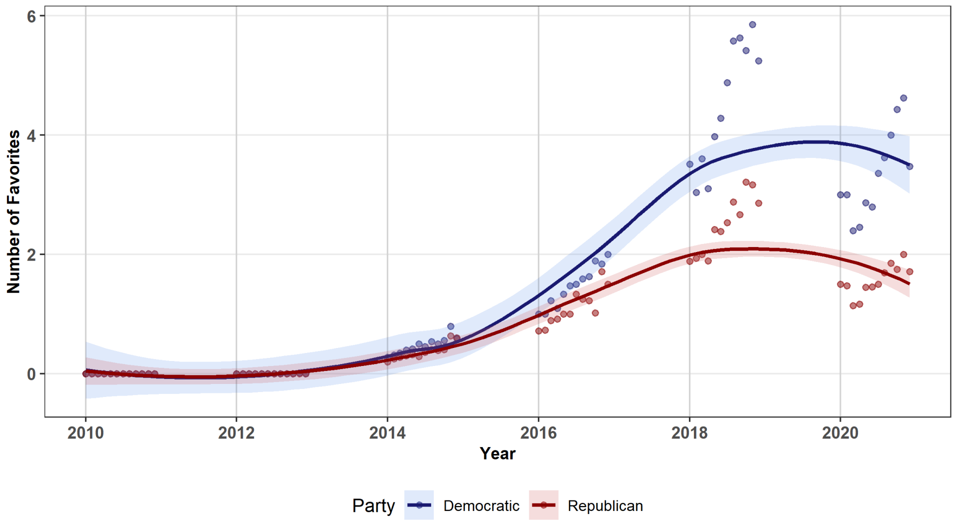
(c) Quotes



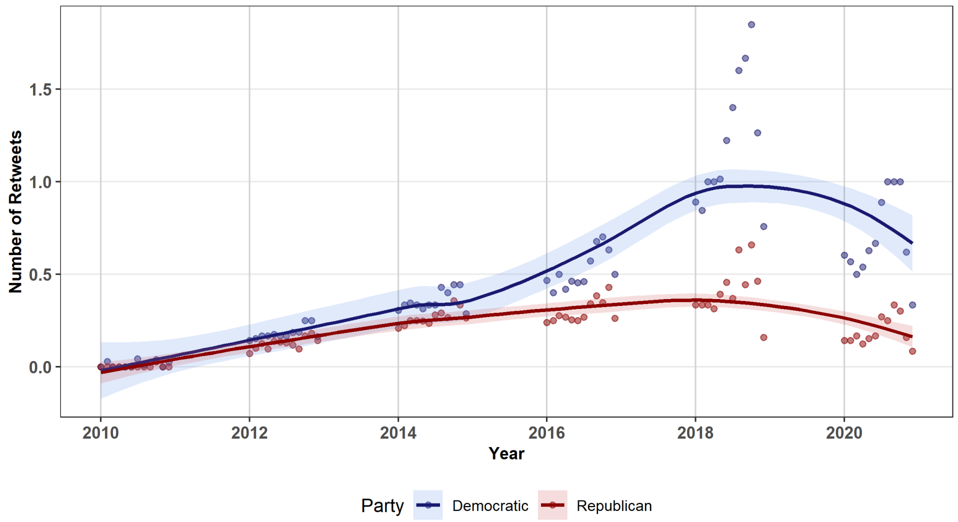
(d) Replies

Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: State Officeholders.

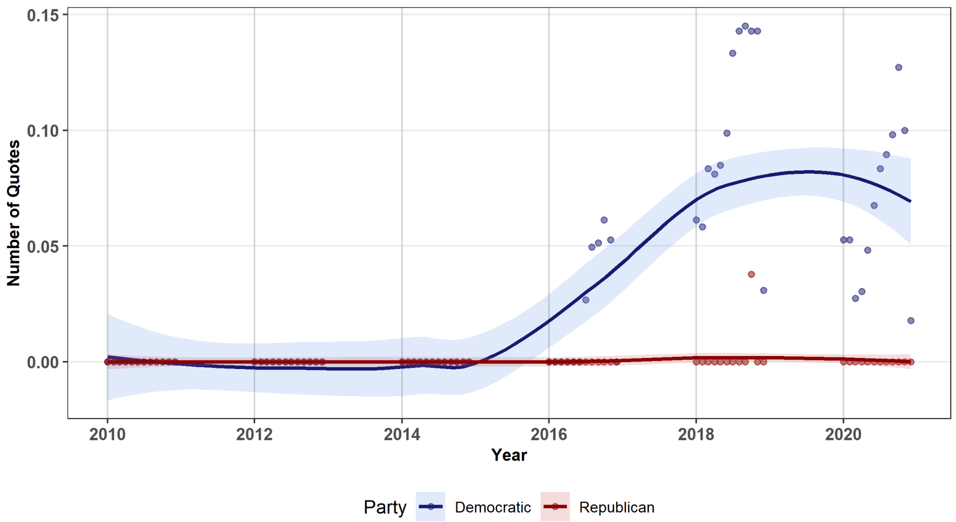
Figure A.2.6: Sample Candidates for State Office: Audience Engagement for Average Tweet of Median Legislator by Party



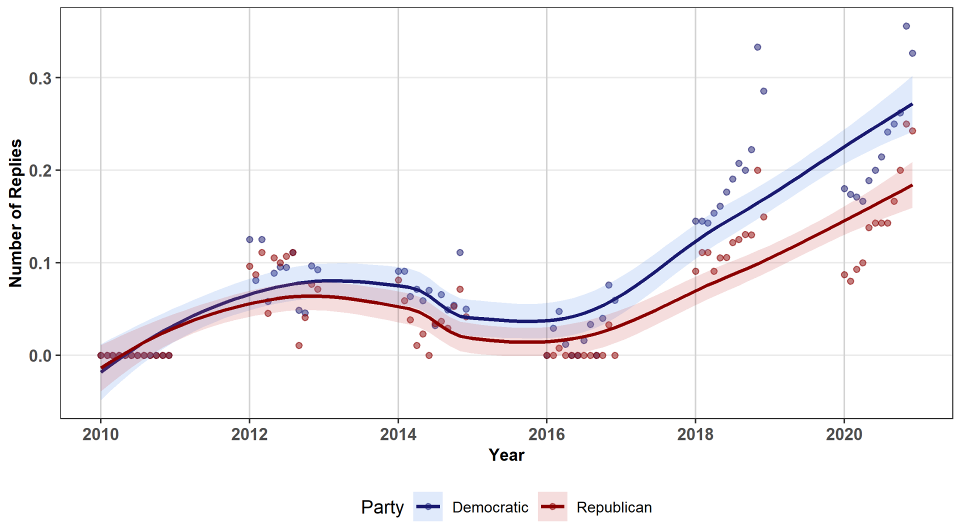
(a) Favorites



(b) Retweets



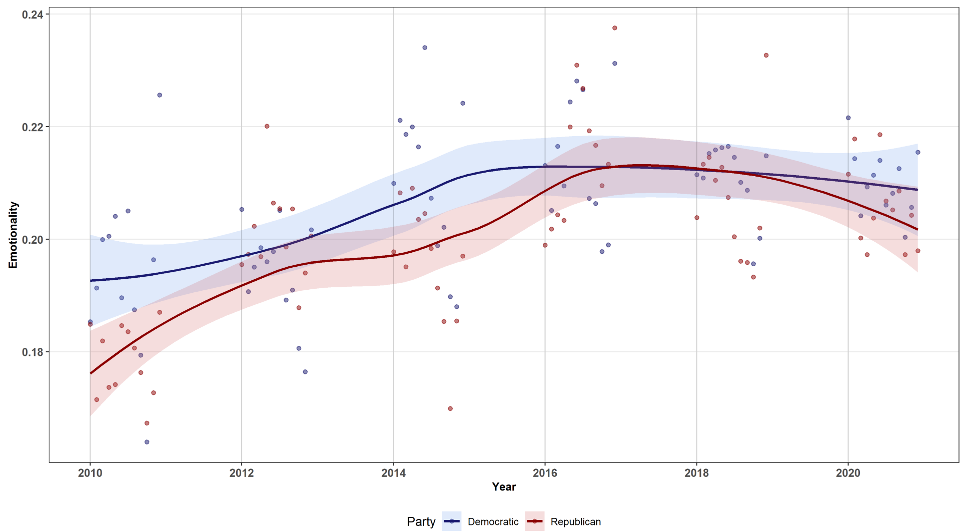
(c) Quotes



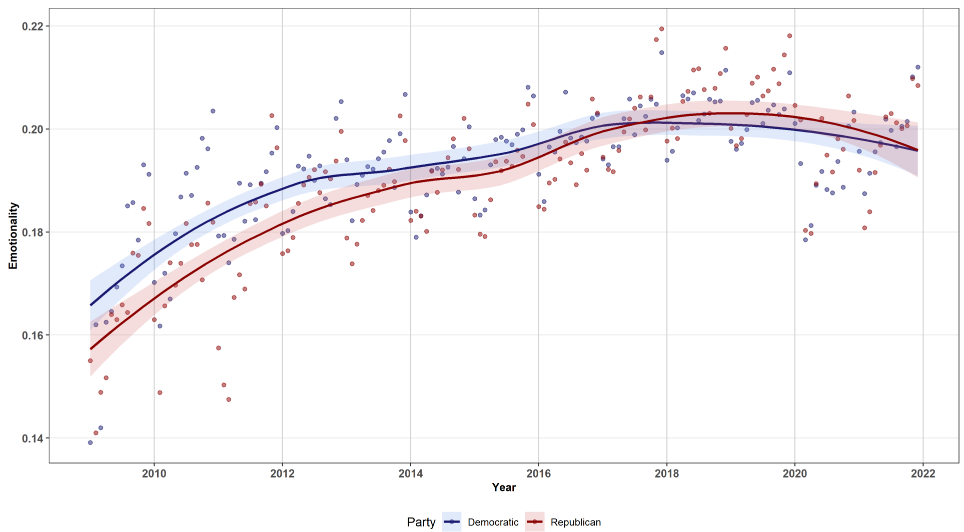
(d) Replies

Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: Candidates for State Office.

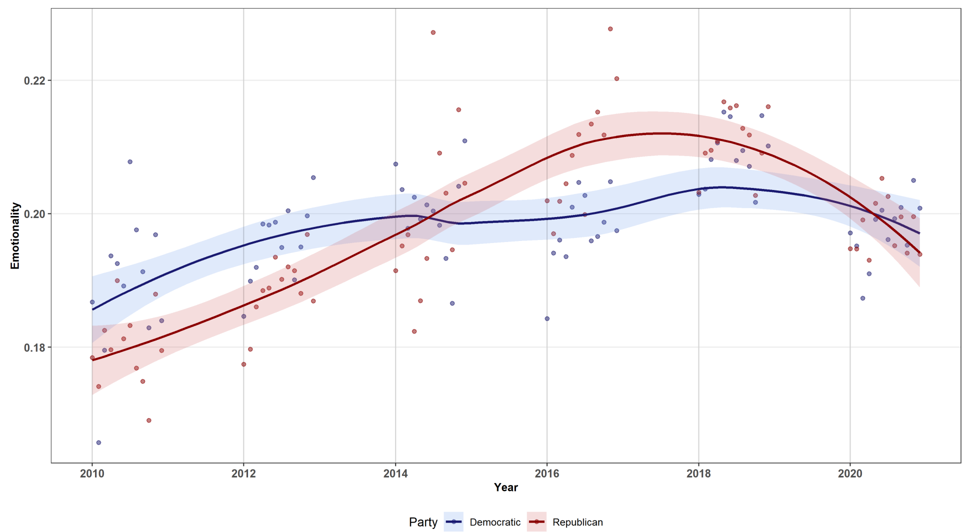
Figure A.2.7: Average Emotionality of Original Tweets per Month for Median Legislator by Party



(a) Sample:Congressional Candidates



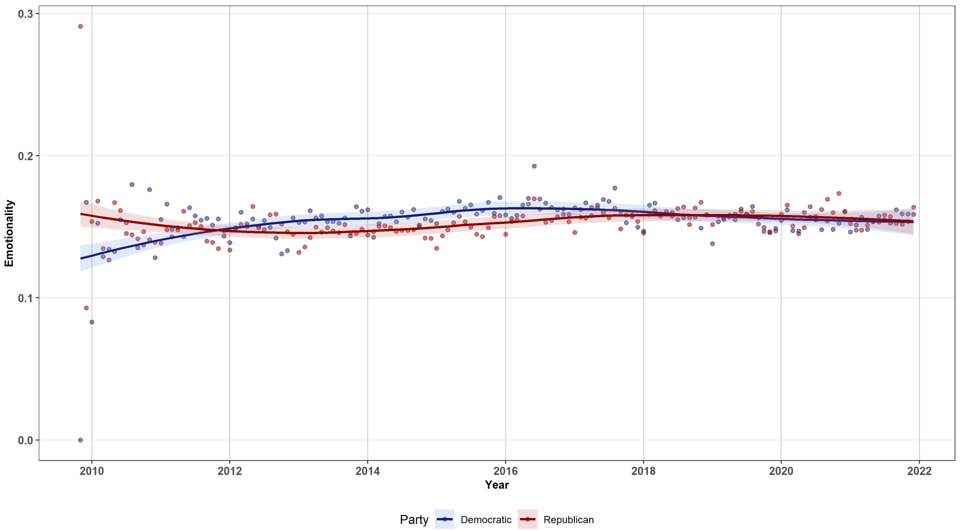
(b) Sample: State Officeholders



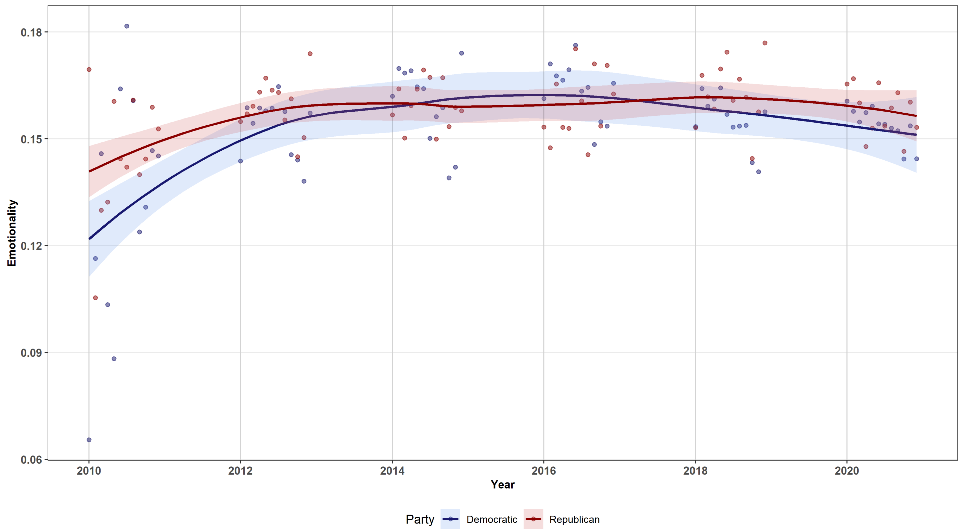
(c) Sample: Candidates for State Office

Notes. - The figure plots the average emotionality per month for the median legislator, by party. I compute emotionality as 1 - share of neutral words at the tweet level and then calculate the average monthly emotionality by politician and display the value for the median legislator in each party. The numbers are based on original tweets.

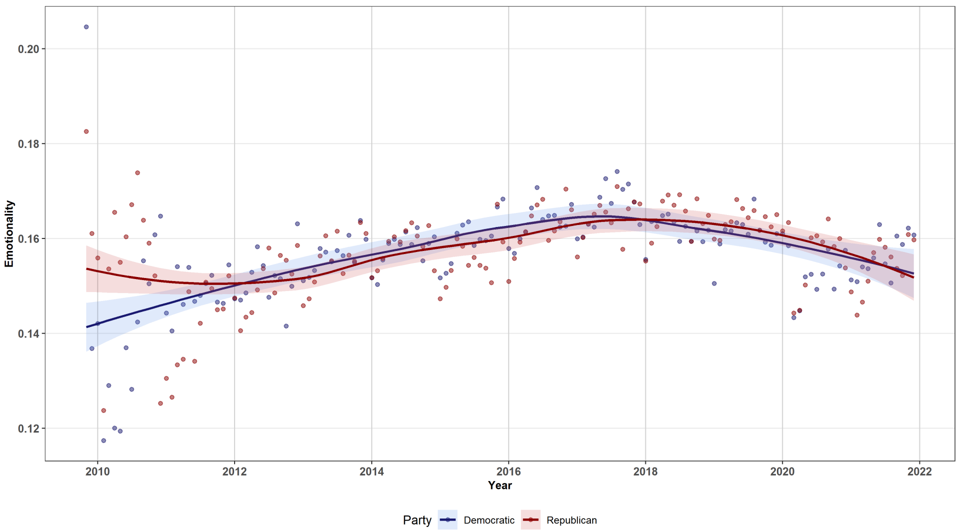
Figure A.2.8: Average Emotionality of Retweets per Month for Median Legislator by Party



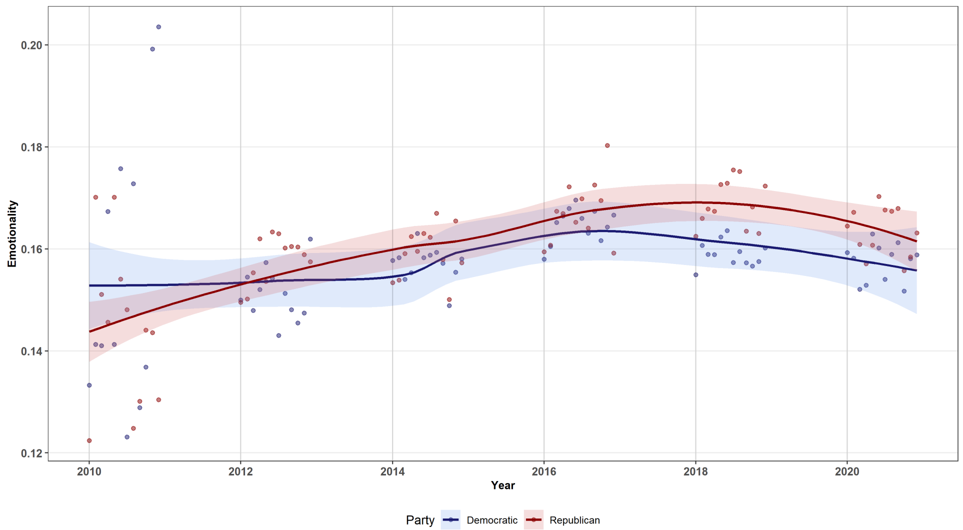
(a) Sample: Members of Congress



(b) Sample:Congressional Candidates



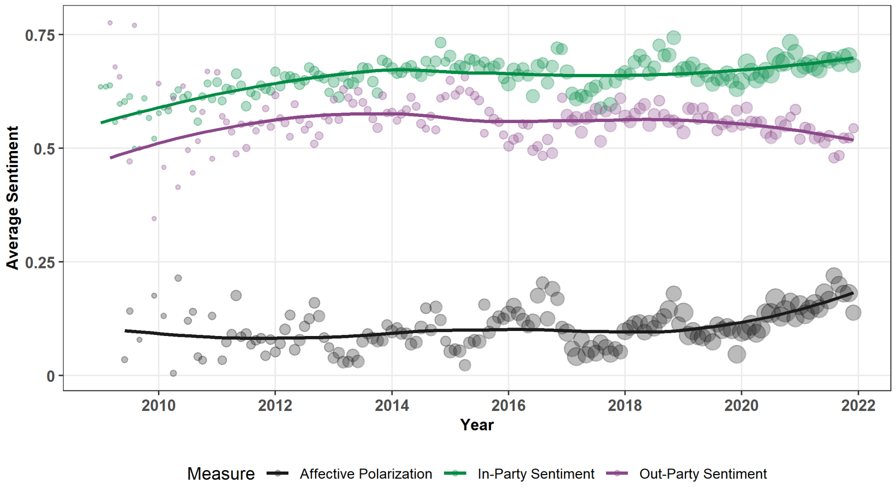
(c) Sample: State Officeholders



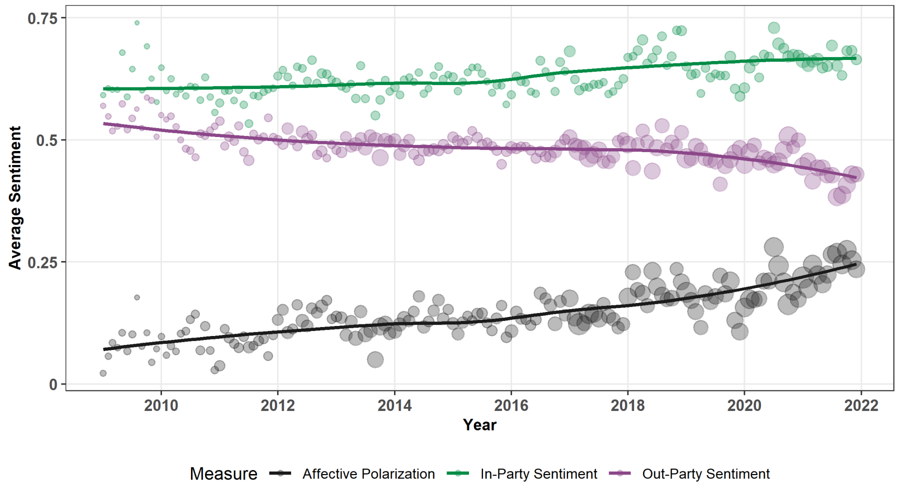
(d) Sample: Candidates for State Office

Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on retweets.

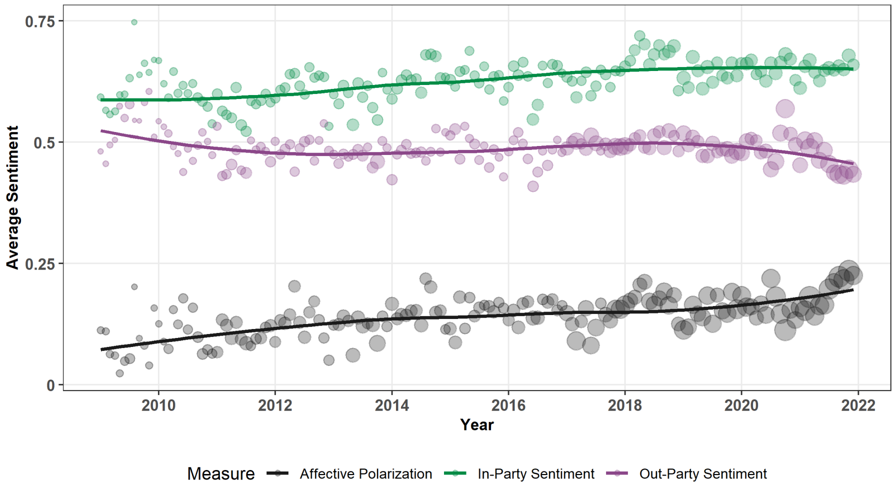
Figure A.2.9: Affective Polarization, In-Party Sentiment and Out-Party Sentiment, 2008 - 2021, by Type of Mention



(a) Sample: Tags (e.g., @mikepence)



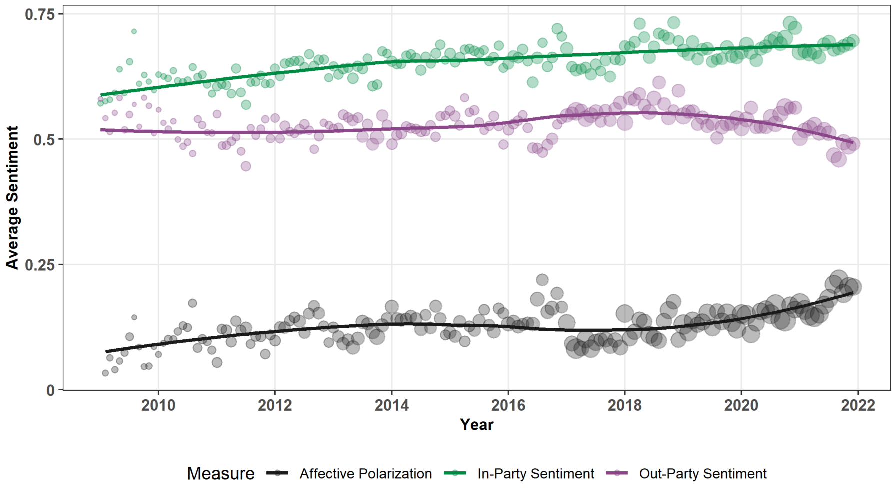
(b) Sample: Name (e.g., Mike Pence)



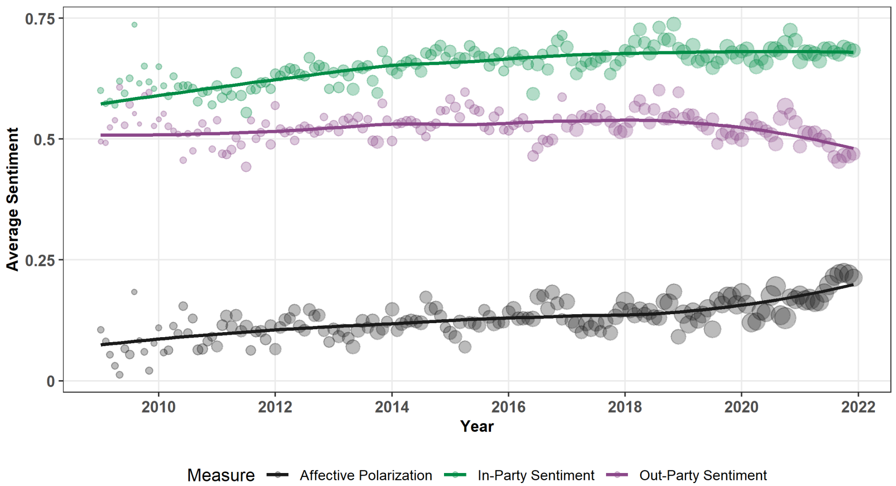
(c) Sample: Ideology (e.g., Democrats)

Notes. - The figure plots the average in-party sentiment, out-party sentiment and affective polarization by month. The measures are computed as described in Section 2.5.1. Panel (a) uses only the subsample of tweets that tag other politicians; Panel (b) uses only the subsample of tweets that mention the names of other politicians; Panel (d) uses only the subsample of tweets that mention keywords related to the ideology of the parties. Sample: Members of Congress.

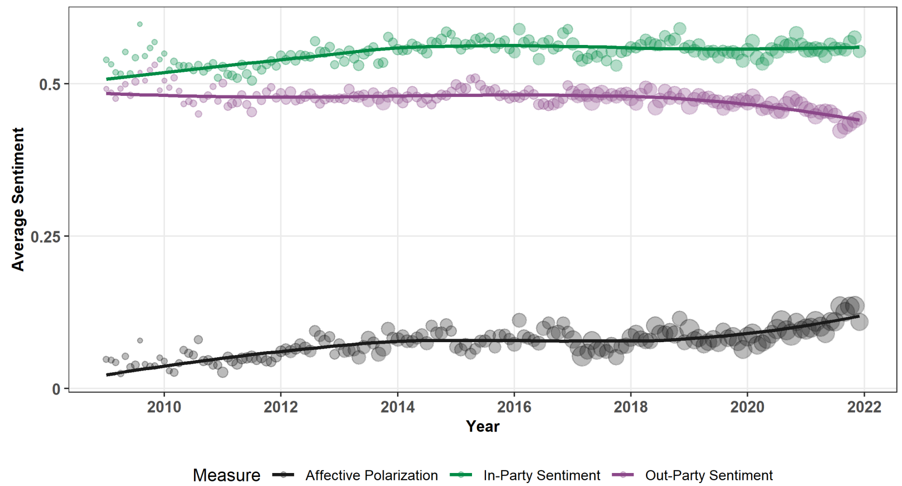
Figure A.2.10: Robustness: Affective Polarization, In-Party Sentiment and Out-Party Sentiment, 2008 - 2021



(a) Average over Individual Aggregation



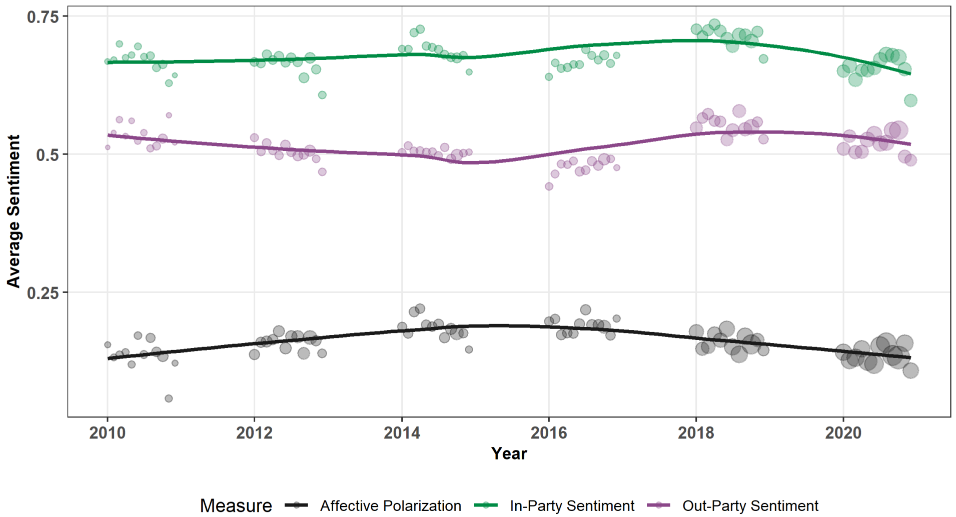
(b) Exclude Tweets about Presidents



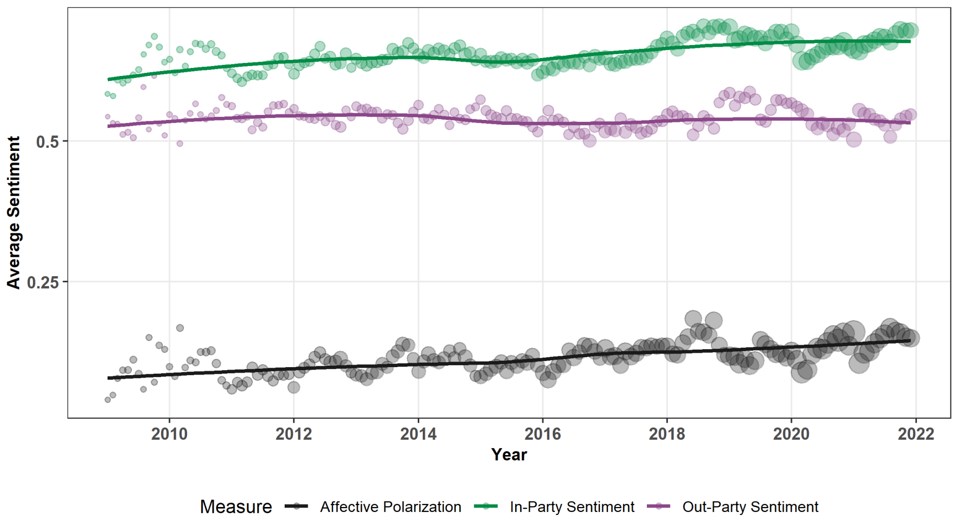
(c) Afinn Dictionary

Notes. - The figure plots the average in-party sentiment, out-party sentiment and affective polarization by month. For Panel (a), (c) and (d) the measures are computed as described in section 2.5.1. In Panel (b) instead, I compute average in- and out-party sentiment within a politician and then average across politicians. Panel (a) uses the full sample of members of Congress; Panel (c) excludes Presidents from the set of targets; Panel (d) uses AFINN, as an alternative sentiment lexicon.

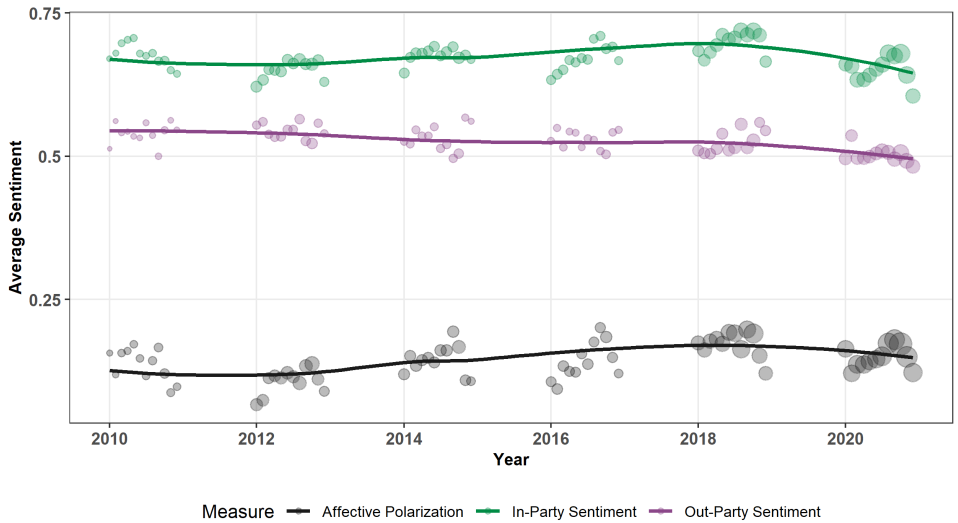
Figure A.2.11: Affective Polarization, In-Party Sentiment and Out-Party Sentiment, 2008 - 2021, by: Type of Office



(a) Sample: Congressional Candidates



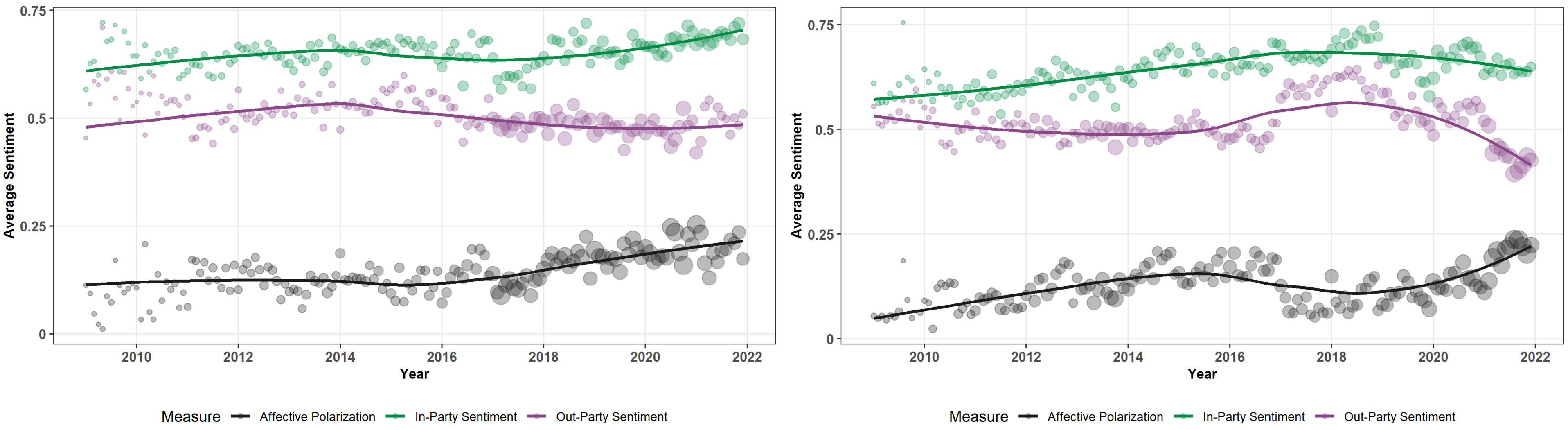
(b) Sample: State Officeholders



(c) Sample: Candidates for State Office

Notes. - The figure plots the average audience engagement for the median legislator by party. For each reaction, I compute the monthly average reaction by politician and display the value for the median legislator in each party. The numbers are based on original tweets. Sample: Members of Congress.

Figure A.2.12: Affective Polarization, In-Party Sentiment and Out-Party Sentiment, 2008 - 2021, by Party

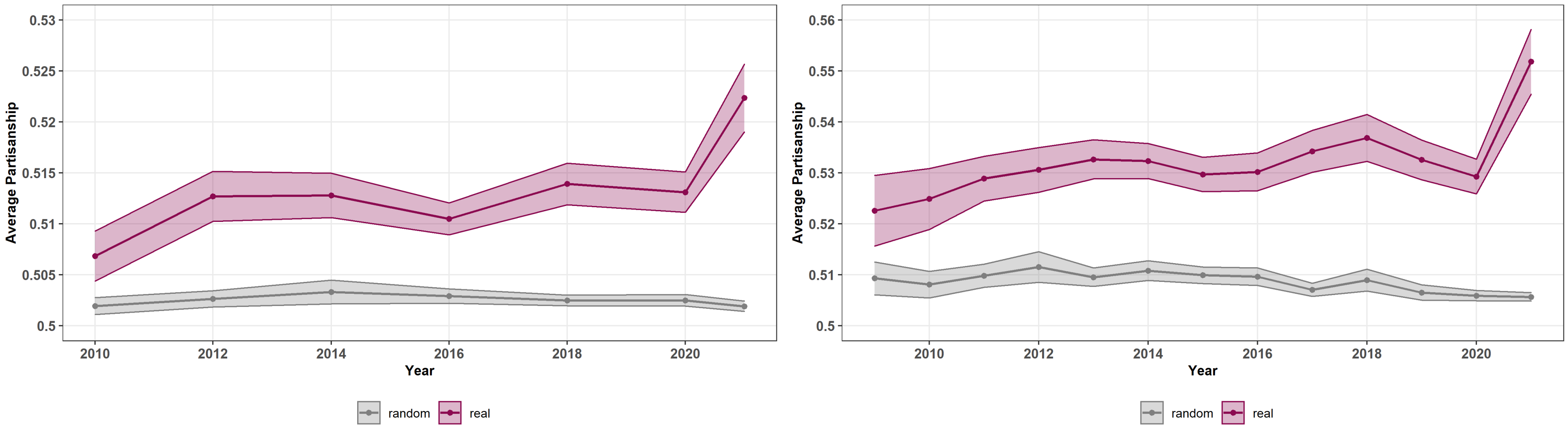


(a) Tweets from Democrats

(b) Tweets from Republicans

Notes. - The figure plots the average in-party sentiment, out-party sentiment and affective polarization by month. The measures are computed as described in Section 2.5.1. Panel (a) uses only the tweets from Democrats. Panel (b) uses only the tweets from Republicans. Sample: Members of Congress.

Figure A.2.13: Partisanship on Social Media, 2009 - 2021



(a) Bi-annual Frequency

(b) Annual Frequency - Lamda = 0.001

Notes. - The figure plots the average partisanship from the penalized estimator $\hat{\pi}_t$. *Panel (a) replicates Figure 2 in Gentzkow, Shapiro and Taddy (2019) and estimates for partisanship bi-annually, that is for the 111th to 117th Congress. Panel (b) plots and estimates partisanship annually and uses a different fixed-cost specification than in the original estimation. In each plot the "real" series is from actual data, while the "random" series assigns politicians' identity at random with the probability that a politician is Republican equal to the average share of Republican politicians in a session of Congress. The shaded areas are pointwise 95% - confidence intervals based on 100 subsamples of the data. For each subsample, I randomly draw 20 percent of all members of congress active on social media and compute the penalized estimate.*

A.2.2 Table Appendix

Table A.2.1: Effect of Jump in Likes on Future Attacks on Opponents

	Dependent variable:			
	N-Attacks	1(Attack)	N-Tweets	N-Attacks/N-Tweets
	(1)	(2)	(3)	(4)
	Panel A. Period T +- 3			
Post-Like	0.073***	0.096***	0.042***	0.068***
	(0.006)	(0.005)	(0.006)	(0.004)
Observations	98,294	98,294	98,294	98,294
	Panel B. Period T +- 5			
Post-Like	0.078***	0.096***	0.040***	0.069***
	(0.007)	(0.005)	(0.006)	(0.005)
Observations	130,947	130,947	130,947	130,947
	Panel C. Period T +- 7			
Post-Like	0.072***	0.095***	0.039***	0.066***
	(0.006)	(0.005)	(0.005)	(0.005)
Observations	159,362	159,362	159,362	159,362
	Panel D. Period T +- 9			
Post-Like	0.073***	0.096***	0.042***	0.068***
	(0.006)	(0.005)	(0.006)	(0.004)
Observations	179,954	179,954	179,954	179,954
Pol FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	Yes	Yes
Calender-Day FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full
R^2	0.181	0.208	0.319	0.140

Notes. - The table reports the OLS estimates from specification 2.6. The unit of observation is at the politician-day level. Column (1) uses a dependent variable the number of attacks (per politician-day), Column (2) uses an indicator whether an attack occured, Column (3) uses the number of tweets and Column (4) uses the number of attacks divided by the number of tweets. The explanatory variables is an indicator that takes the value of 1 in the t days after a politician attacked an opponent and this attack was 'liked' by 100% more than the two-week rolling average of the politician, and 0 in the t days before this attack. Panel (a.) - Panel (d.) present time-windows for $t \in (3, 5, 7, 9)$ days before and after the attack. All regressions include full sets of candidate, state, office, calender-year and -day fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table A.2.2: Effect of Jump in Backlash on Future Attacks on Opponents

	Dependent variable:			
	N-Attacks	1(Attack)	N-Tweets	N-Attacks/N-Tweets
	(1)	(2)	(3)	(4)
	Panel A. Period T +- 3			
Post-Backlash	0.060^{***}	0.065^{***}	0.028^{***}	0.043^{***}
	(0.009)	(0.009)	(0.009)	(0.008)
Observations	47,736	47,736	47,736	47,736
	Panel B. Period T +- 5			
Backlash	0.059^{***}	0.062^{***}	0.032^{***}	0.046^{***}
	(0.008)	(0.007)	(0.007)	(0.006)
Observations	68,367	68,367	68,367	68,367
	Panel C. Period T +- 7			
Backlash	0.059^{***}	0.064^{***}	0.032^{***}	0.046^{***}
	(0.007)	(0.006)	(0.006)	(0.006)
Observations	88,291	88,291	88,291	88,291
	Panel D. Period T +- 9			
Backlash	0.062^{***}	0.067^{***}	0.038^{***}	0.048^{***}
	(0.007)	(0.006)	(0.006)	(0.006)
Observations	104,464	104,464	104,464	104,464
Pol FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	Yes	Yes
Calender-Day FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full
R^2	0.181	0.208	0.318	0.140

Notes. - The table reports the OLS estimates from specification 2.6. The unit of observation is at the politician-day level. Column (1) uses a dependent variable the number of attacks (per politician-day), Column (2) uses an indicator whether an attack occured, Column (3) uses the number of tweets and Column (4) uses the number of attacks divided by the number of tweets. The explanatory variables is an indicator that takes the value of 1 in the *t* days after a politician attacked an opponent and this attack was 'ratioed' by 100% more than the two-week rolling average of the politician, and 0 in the *t* days before this attack. Panel (a.) - Panel (d.) present time-windows for $t \in (3, 5, 7, 9)$ days before and after the attack. All regressions include full sets of candidate, state, office, calender-year and -day fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table A.2.3: Permutation Test: Effect of Random Jump in Likes on Future Attacks on Opponents

	Dependent variable:			
	N-Attacks	1(Attack)	N-Tweets	N-Attacks/N-Tweets
	(1)	(2)	(3)	(4)
	Panel A. Period T +- 3			
Post-Like	0.002	0.010	−0.003	0.005
	(0.006)	(0.007)	(0.005)	(0.006)
Observations	97,899	97,899	97,899	97,899
	Panel B. Period T +- 5			
Post-Like	−0.001	0.003	0.004	0.002
	(0.005)	(0.005)	(0.005)	(0.005)
Observations	126,186	126,186	126,186	126,186
	Panel C. Period T +- 7			
Post-Like	0.001	0.004	0.005	0.002
	(0.005)	(0.005)	(0.005)	(0.005)
Observations	150,221	150,221	150,221	150,221
	Panel D. Period T +- 9			
Post-Like	0.007	0.004	0.013***	0.006
	(0.005)	(0.004)	(0.005)	(0.004)
Observations	166,418	166,418	166,418	166,418
Pol FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	Yes	Yes
Calender-Day FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full
R^2	0.166	0.166	0.327	0.122

Notes. - The table reports the OLS estimates from specification 2.6. The unit of observation is at the politician-day level. Column (1) uses a dependent variable the number of attacks (per politician-day), Column (2) uses an indicator whether an attack occured, Column (3) uses the number of tweets and Column (4) uses the number of attacks divided by the number of tweets. The explanatory variables is an indicator that takes the value of 1 in the t days after a politician attacked an opponent, and 0 in the t days before this attack. Panel (a.) - Panel (d.) present time-windows for $t \in (3,5,7,9)$ days before and after the attack. All regressions include full sets of candidate, state, office, calender-year and -day fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table A.2.4: Permutation Test: Effect of Random Jump in Backlash on Future Attacks on Opponents

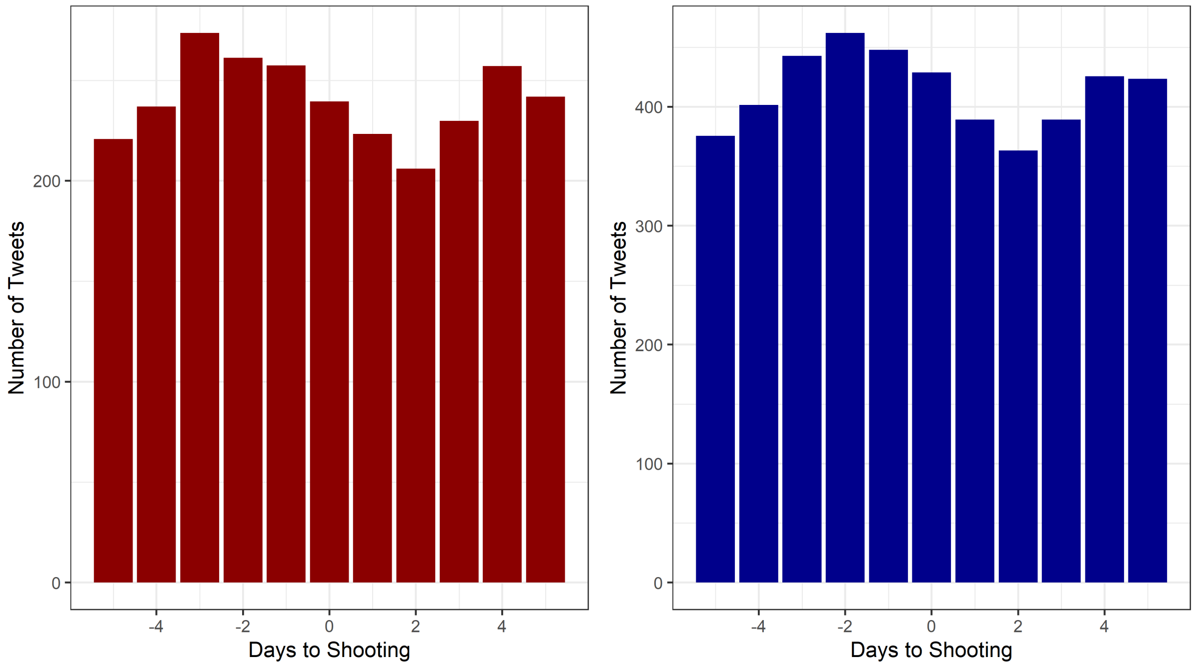
	Dependent variable:			
	N-Attacks	1(Attack)	N-Tweets	N-Attacks/N-Tweets
	(1)	(2)	(3)	(4)
	Panel A. Period T +- 3			
Post-Backlash	0.002	0.012**	0.002	0.005
	(0.001)	(0.006)	(0.001)	(0.006)
Observations	81,226	81,226	81,226	81,226
	Panel B. Period T +- 5			
Post-Backlash	0.001	0.007	0.001	0.010*
	(0.001)	(0.005)	(0.001)	(0.005)
Observations	108,273	108,273	108,273	108,273
	Panel C. Period T +- 7			
Post-Backlash	0.0003	0.003	0.001	0.006
	(0.001)	(0.005)	(0.001)	(0.005)
Observations	131,891	131,891	131,891	131,891
	Panel D. Period T +- 9			
Post-Backlash	−0.002	0.009*	0.001	0.007
	(0.002)	(0.005)	(0.001)	(0.005)
Observations	149,207	149,207	149,207	149,207
Pol FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Office FE	Yes	Yes	Yes	Yes
Calender-Day FE	Yes	Yes	Yes	Yes
Sample	Full	Full	Full	Full
R^2	0.181	0.208	0.318	0.140

Notes. - The table reports the OLS estimates from specification 2.6. The unit of observation is at the politician-day level. Column (1) uses a dependent variable the number of attacks (per politician-day), Column (2) uses an indicator whether an attack occured, Column (3) uses the number of tweets and Column (4) uses the number of attacks divided by the number of tweets. The explanatory variables is an indicator that takes the value of 1 in the t days after a politician attacked an opponent, and 0 in the t days before this attack. Panel (a.) - Panel (d.) present time-windows for $t \in (3,5,7,9)$ days before and after the attack. All regressions include full sets of candidate, state, office, calender-year and -day fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

A.3 Appendix Chapter 3 - Bad News and Political Deception

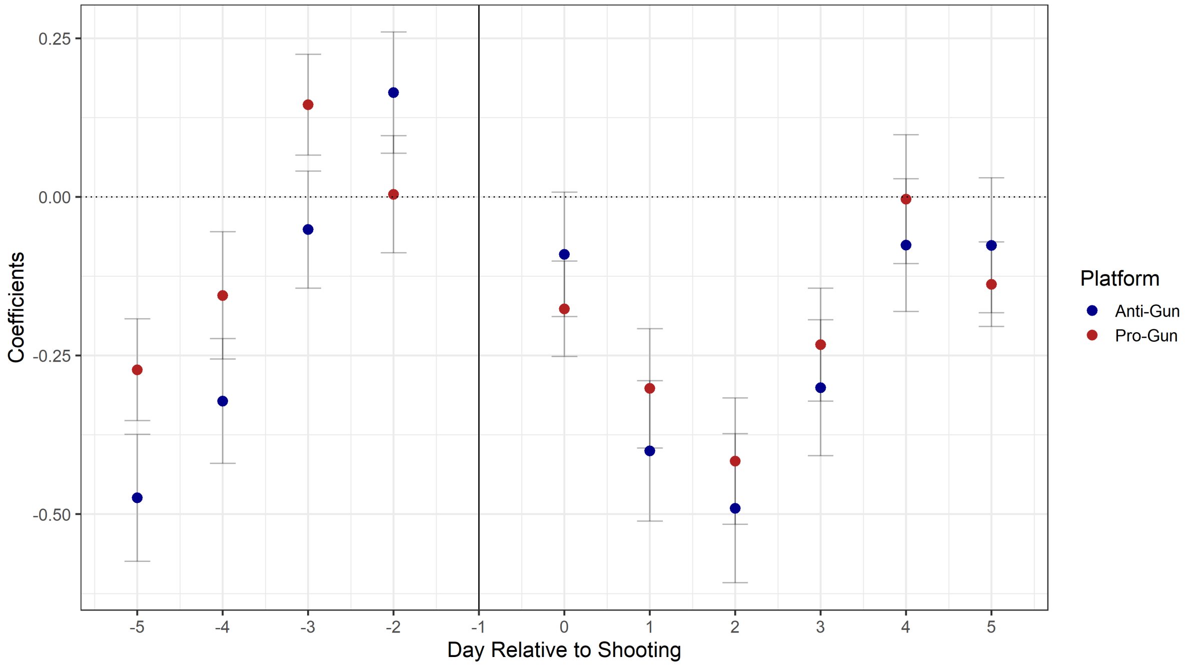
A.3.1 Figure Appendix

Figure A.3.1: Number of Tweets around Mass Shooting Events by Gun-Stance

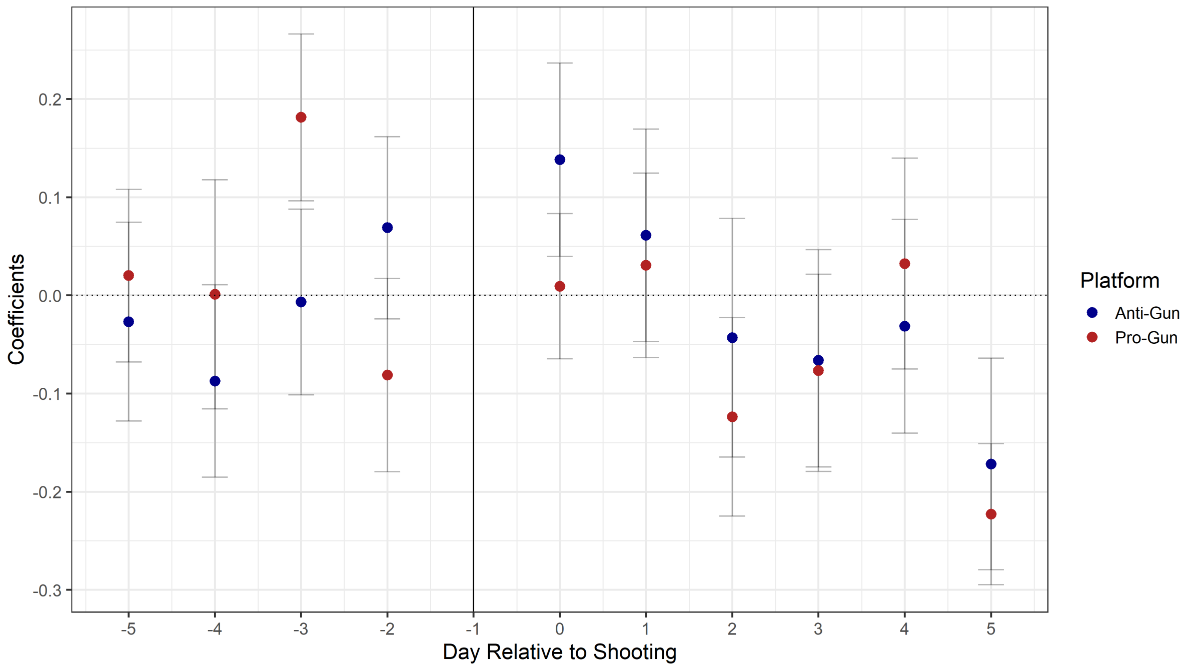


Notes. - The figure plots the average number of tweets for each day relative to the mass shooting event for pro-gun politicians (red) and anti-gun politicians (blue).

Figure A.3.2: Effect of Mass Shootings on Twitter Activity



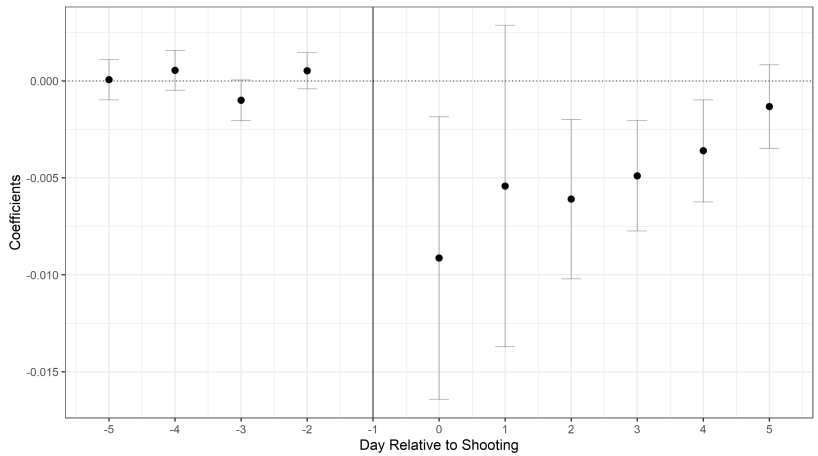
(a) Specification 1



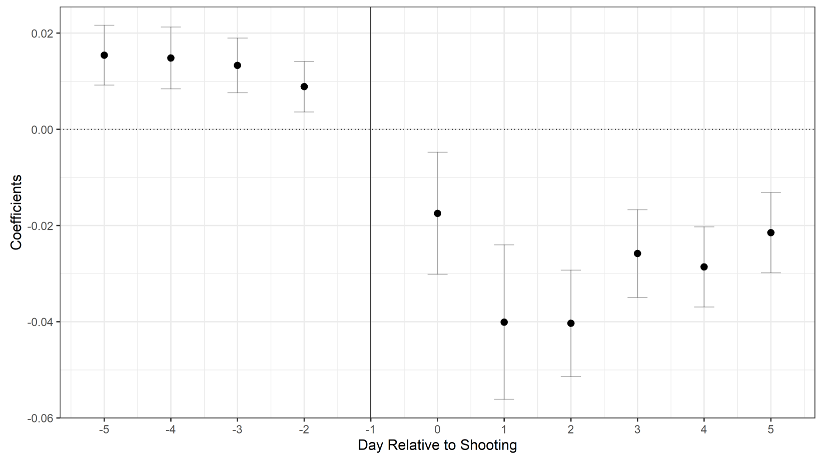
(b) Specification 1 plus Weekday Fixed Effects

Notes. - The figure plots the estimates of β_k and 95-percent confidence intervals from specification [3.1]. The dependent variable is the number of tweets from politician i on day t. The outcome is regressed on a set of leads and lags of days relative to the day before the mass shooting, separately for tweets from pro-gun and anti-gun politicians. The specification in Panel (a) controls for politician and mass-shooting FE. The specification in Panel (b) adds weekday fixed effects. Standard errors are clustered by politician. Sample: Full sample.

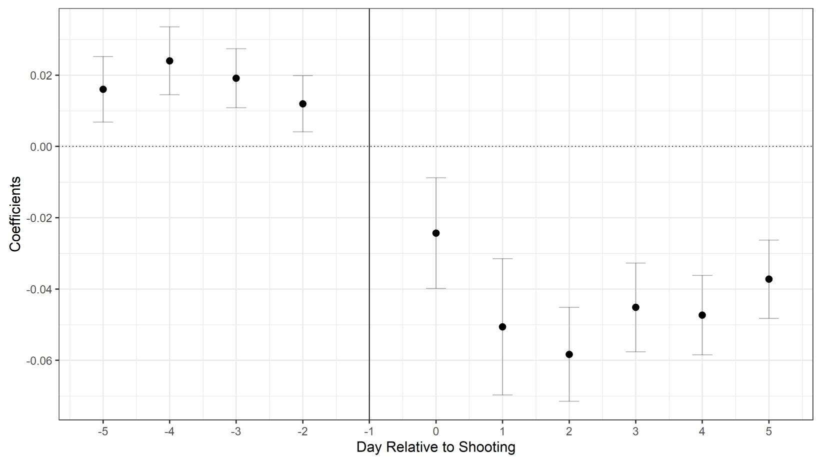
Figure A.3.3: Event-Study Plot, Distraction, Differences between Pro-Gun and Anti-Gun Coefficients



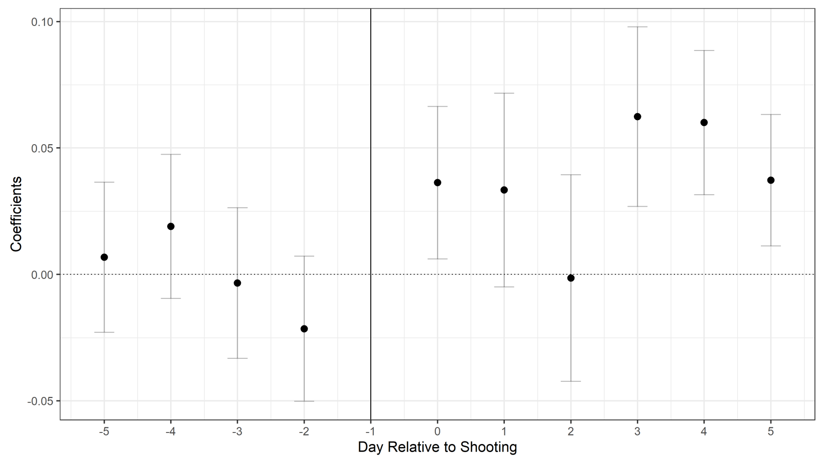
(a) Outcome - Indicator(Event and Gun)



(b) Outcome - Indicator(Event or Gun)



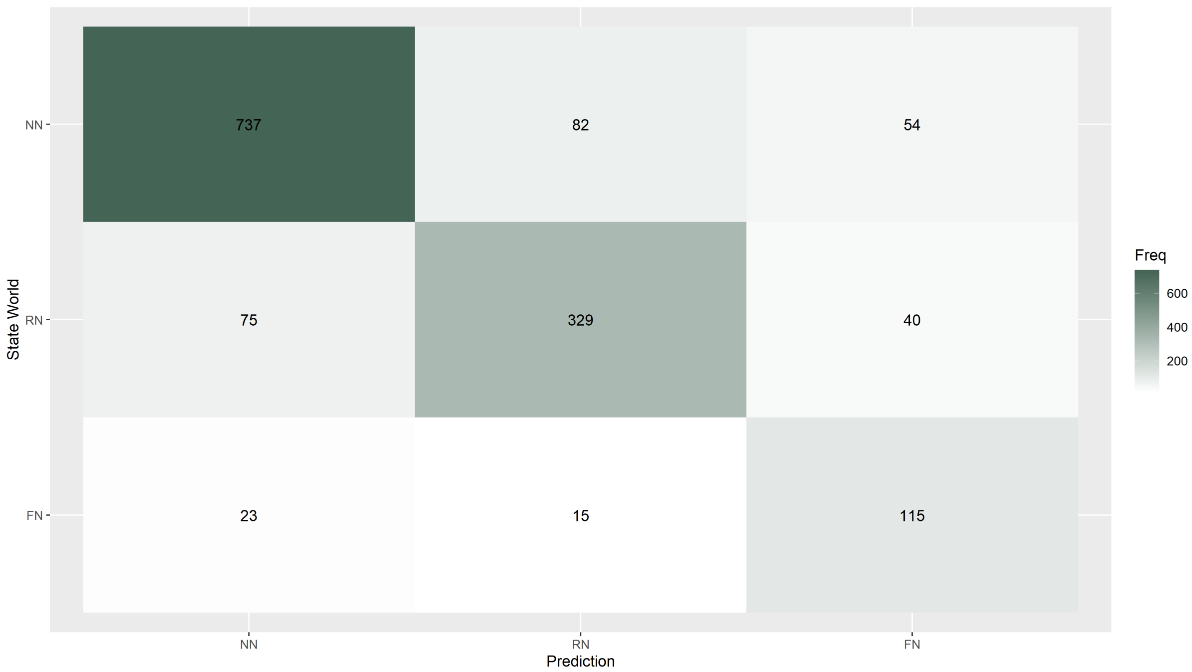
(c) Outcome - Indicator(Shooting-Topic)



(d) Outcome - News Similarity

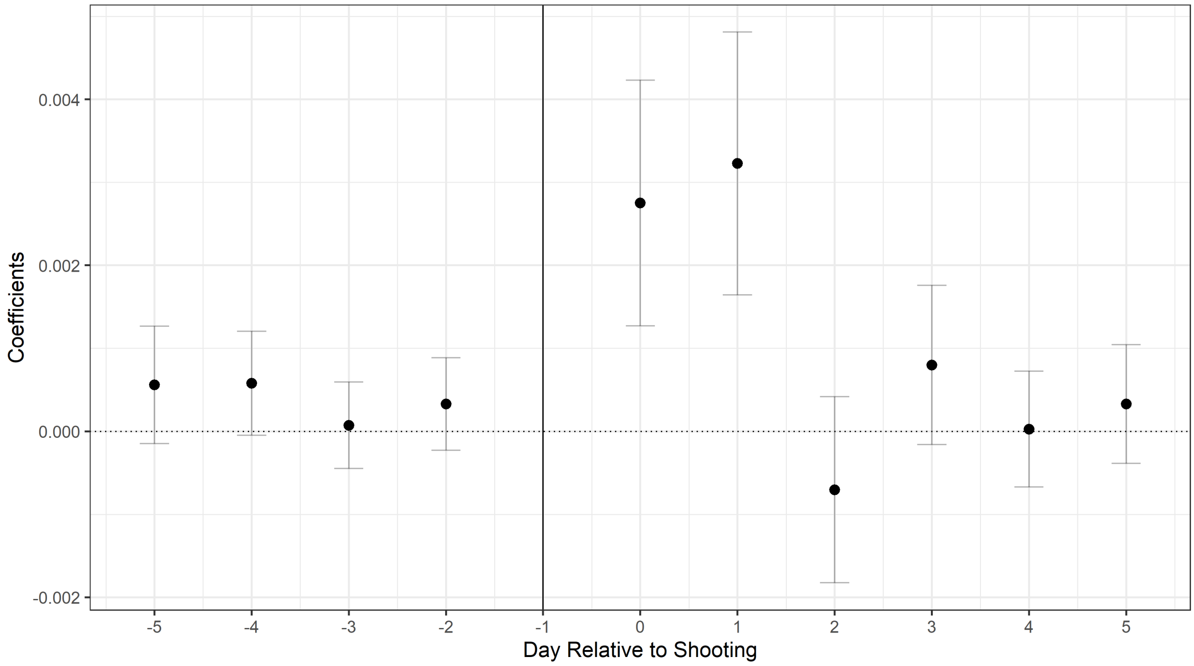
Notes. - The figure plots the estimates of $\beta_k\times$ pro-gun and 95-percent confidence intervals from specification [3.1]. The dependent variable is an indicator that is equal to 1 if tweet j from politician i on day t contains event and gun-related keywords (Panel (a)); event or gun-related keywords (Panel (b)), and 0 otherwise. The dependent variable in panel (c) is an indicator that is equal to 1 if the topic of tweet j from politician i on day t is related to mass shootings, and 0 otherwise. The dependent variable in panel (d) is the cosine similarity between national news and if tweet j from politician i on day t, multiplied by 100. The indicator is regressed on a set of leads and lags of days relative to the day before the mass shooting interacted with an indicator for pro-gun. The specification controls for politician and mass-shooting FE. Standard errors are clustered by politician. Sample: Full sample.

Figure A.3.4: Test-Set Prediction Performance for Main Neural Network on News Data

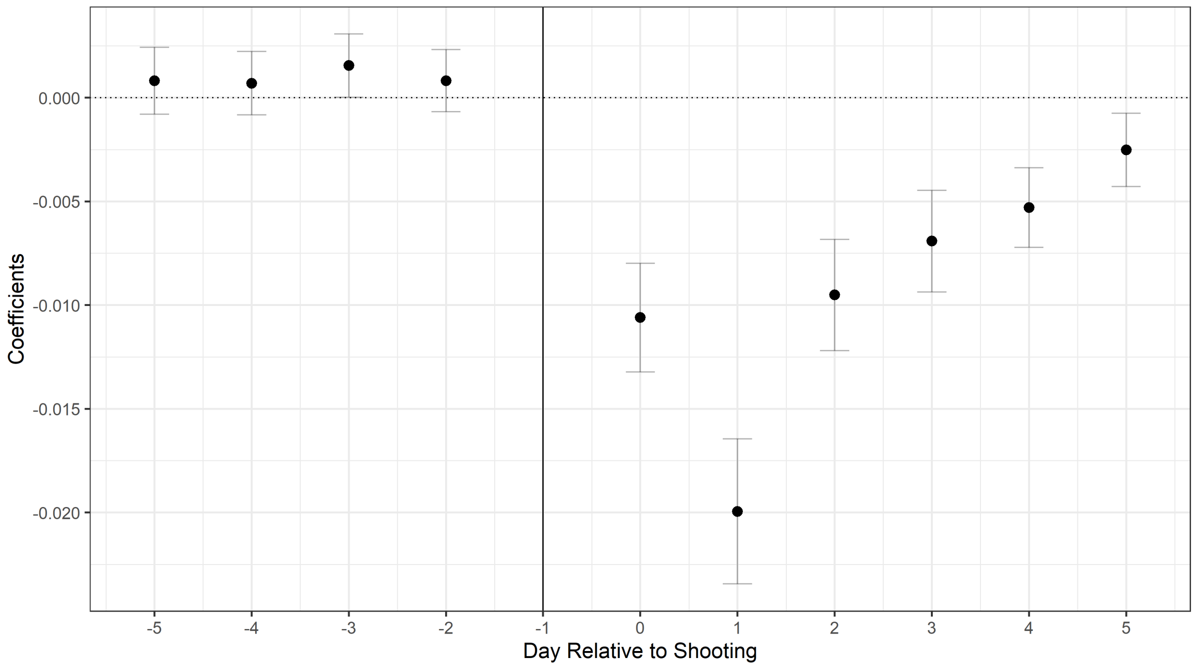


Notes. - The figure plots the confusion matrix for test-set predictions on the newspaper-data. The y-axis contains the true state of the world; the x-axis the predicted state of the world. Hence, the diagonal from the top left to the bottom right contains correct predictions. The elements to the left of the diagonal are false positives and the elements to the right of the diagonal are false negatives.

Figure A.3.5: Event-Study Plot, Deception, Differences between Pro-Gun and Anti-Gun Coefficients



(a) Outcome: FN



(b) Outcome: RN

Notes. - The figure plots the estimates of $\beta_k \times$ pro-gun and 95-percent confidence intervals from specification [3.1]. The dependent variable in panel a is an indicator for deception that is equal to 1 if tweet j from politician i on day t is labelled by the neural network as fake with more than 90% probability, and 0 otherwise. The dependent variable in panel b is an indicator for accurate information provision that is equal to 1 if if tweet j from politician i on day t is labelled by the neural network as real news, and 0 otherwise. The indicator is regressed on a set of leads and lags of days relative to the day before the mass shooting interacted with an indicator for pro-gun. The specification controls for politician and mass-shooting FE. Standard errors are clustered by politician. Sample: Full sample.

A.3.2 Table Appendix

Table A.3.1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
pro_gun	494,359	0.374	0.484	0	1
event_and	494,359	0.018	0.133	0	1
event_or	494,359	0.078	0.268	0	1
gun_topic	400,205	0.115	0.319	0.000	1.000
news_similarity	352,737	0.002	0.011	0.000	0.482
fn	494,359	0.001	0.037	0	1
rn	494,359	0.007	0.083	0	1
incumbent	494,359	0.579	0.494	0	1
hurricane	494,359	0.126	0.332	0	1
disaster_topic	400,205	0.012	0.107	0.000	1.000
news_pressure	481,317	0.516	0.156	0.019	1.000
elect_comp_lag	407,028	0.500	0.304	0.000	0.999

Table A.3.2: Effect of Mass Shootings on Twitter Activity

	Dependent variable:			
	Number Tweets	Number Tweets	1(Tweeted)	1(Tweeted)
	(1)	(2)	(3)	(4)
Post	−0.103***	−0.021	−0.008***	0.004*
	(0.031)	(0.031)	(0.002)	(0.002)
Pro-Gun	−0.198	−0.200	0.012	0.012
	(0.198)	(0.198)	(0.023)	(0.023)
Post*Pro-Gun	−0.053	−0.051	−0.014***	−0.013***
	(0.040)	(0.040)	(0.003)	(0.003)
Mean Dep Var	2.296	2.296	0.533	0.533
Pol FE	Yes	Yes	Yes	Yes
MS FE	Yes	Yes	Yes	Yes
Weekday FE	No	Yes	No	Yes
Observations	347,424	347,424	347,424	347,424
R^2	0.381	0.400	0.352	0.389

Notes. - The table reports OLS estimates from specification 3.2. The dependent variable in column (1) and (2) is the number of tweets sent from a politicians i at a day d. The dependent variable in column (3) and (4) is an indicator equal to 1 if a politician i has sent a tweet at a day d, and 0 otherwise. The dependent variables are regressed on an indicator for the day of the mass shooting and thereafter, for a pro-gun rating, and their interaction. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table A.3.3: Topicmodel - Relevant Topics

topic	top terms	political	shooting
1	vote,dai,support,famili,peopl	Yes	No
2	trump,presid,administr,american,obama	Yes	No
3	water,protect,land,nation,public	Yes	No
4	health,care,veteran,access,mental	Yes	No
5	sexual,justic,abus,assault,victim	Yes	No
6	women,health,care,right,access	Yes	No
7	school,student,teacher,kid,children	Yes	No
8	investig,mueller,trump,presid,elect	Yes	No
9	gun,weapon,ban,check,background	No	Yes
10	texa,san,#txlege,tx,houston	Yes	No
11	virginia,west,miner,joe,coal	Yes	No
12	worker,wage,pai,labor,union	Yes	No
13	food,famili,program,million,cut	Yes	No
14	vote,poll,elect,dai,ballot	Yes	No
15	shoot,kill,peopl,polic,new	No	Yes
16	north,korea,nuclear,iran,deal	Yes	No
17	vote,senat,bob,cruz,#txsen	Yes	No
18	evacu,fire,counti,updat,close	Yes	No
19	children,famili,immigr,trump,administr	Yes	No
20	law,feder,public,govern,bill	Yes	No
21	gun,violenc,action,shoot,congress	No	Yes
22	drug,prescript,price,cost,lower	Yes	No
23	compani,million,pai,american,peopl	Yes	No
24	right,protect,vote,equal,law	Yes	No
25	govern,hous,republican,congress,american	Yes	No
26	project,citi,road,bridg,new	Yes	No
27	peopl,right,democraci,american,china	Yes	No
28	women,equal,woman,american,congress	Yes	No
29	donat,campaign,debat,win,chip	Yes	No
30	rule,trump,join,letter,administr	Yes	No

31	secur,social,medicar,cut,senior	Yes	No
32	fund,bill,hous,secur,committe	Yes	No
33	opioid,crisi,epidem,drug,commun	Yes	No
34	peopl,vote,time,right,american	Yes	No
35	puerto,disast,rico,relief,hurrican	Yes	No
36	trump,call,attack,presid,republican	Yes	No
37	famili,victim,shoot,prayer,commun	No	Yes
38	court,senat,confirm,judg,suprem	Yes	No
39	famili,friend,pass,miss,senat	Yes	No
40	cancer,research,awar,diseas,support	Yes	No
41	tax,cut,famili,class,middl	Yes	No
42	protect,vote,#netneutr,life,internet	Yes	No
43	student,educ,colleg,school,program	Yes	No
44	honor,veteran,dai,servic,nation	Yes	No
45	hurrican,storm,prepar,stai,florida	Yes	No
46	tax,job,busi,american,#taxreform	Yes	No
47	fund,govern,vote,senat,democrat	Yes	No
48	energi,clean,ga,oil,job	Yes	No
49	democrat,candid,elect,win,vote	Yes	No
50	israel,war,saudi,attack,terrorist	Yes	No
51	fbi,clinton,secur,memo,hous	Yes	No
52	hous,senat,pass,vote,bill	Yes	No
53	commun,infrastructur,invest,rural,access	Yes	No
54	protect,care,condit,health,pre	Yes	No
55	hate,commun,violenc,stand,attack	No	Yes
56	job,busi,economi,program,innov	Yes	No
57	climat,chang,action,trump,crisi	Yes	No
58	offic,law,enforc,polic,commun	No	Yes
59	job,rate,economi,wage,growth	Yes	No
60	bill,act,pass,bipartisan,support	Yes	No
61	militari,forc,nation,servic,air	Yes	No
62	dreamer,#dreamer,immigr,protect,#daca	Yes	No
63	governor,gov,#mdpolit,#mdgovdeb,maryland	Yes	No

64	#maga,trump,congress,michael,pro	Yes	No
65	health,enrol,insur,plan,#getcov	Yes	No
66	senat,vote,#mtpol,support,#mtsen	Yes	No
67	tax,cut,spend,budget,billion	Yes	No
68	monei,peopl,elect,polit,campaign	Yes	No
69	border,immigr,secur,illeg,wall	Yes	No
70	farmer,trade,farm,economi,agricultur	Yes	No

Table A.3.4: Robustness News Similarity

	Dependent variable:							
	News Similarity (Top-5)				News Similarity (Top-3)		News Similarity (Top-7)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.035***	−0.034***	0.023***	−0.035***	0.010	−0.064***	0.029***	−0.004
	(0.009)	(0.006)	(0.008)	(0.005)	(0.011)	(0.008)	(0.007)	(0.004)
Pro-Gun		−0.065		−0.050		−0.097		−0.037
		(0.049)		(0.039)		(0.061)		(0.035)
Post*Pro-Gun		0.042***		0.036***		0.056***		0.011*
		(0.008)		(0.008)		(0.011)		(0.006)
Mean Dep Var	0.1162	0.1225	0.1138	0.1201	0.1486	0.1618	0.090345	0.09083
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	109,710	455,905	109,710	455,905	109,710	455,905	109,710	455,905
R^2	0.035	0.024	0.033	0.023	0.028	0.023	0.035	0.020

Notes. - The table reports the OLS estimates from specification 3.2. The dependent variable is the cosine similarity between evening news and tweet j from politician i on day t, multiplied by 100.
Each column replicates the baseline specifications (column 2 and column 4 of Table 3) with the following modifications: Columns (1) and (2) lower the number of bigrams-considered for the classification to 3,000; Columns (3) and (4) extend the number of bigrams-considered for the classification to 5,000. Columns (5) and (6) lower the number of news stories considered for the classification per day and channel to 3; Columns (7) and (8) extend lower the number of news stories considered for the classification per day and channel to 7. Columns (1),(3),(5) and (7) refer to the sample of pro-gun politicians; Columns (2),(4),(6),(8) to the full sample of politicians. All columns include politician and mass shooting fixed effects. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

Table A.3.5: Distribution of Causes of Mass Shootings by News Source

	Real News Sources		**Fake News Sources**	
		N		N
Guns	0.34	1570	0.09	95
Mental Illness	0.23	988	0.37	392
Video Games	0.04	190	0.06	69
Terror Attack	0.39	1729	0.47	497
$\sum$	1	4,377	1	1,050

Table A.3.6: Tweets about Causes of Shootings Predicted with 99% Probability from Fake News Source/ Real News Source

1	@JasonWhitely: #BREAKING: US Army says the shooter had mental health issues, served in Iraq in 2011. Treated for PTSD.
2	@spo1981 guns & uncontrolled immigration & attitude down playing Islamic terrorist threat = slaughter in Orlando
3	We need to end the pattern: Islamic terrorists attack after being radicalized online. #sayfie
4	@mpchc1 bad guys R criminals who would love 2 see their victims disarmed. Also on list: mentally disturbed people taking psych related drugs
5	Ron – "I believe, this is again another attack, a terrorist attack inspired by ISIS, and that is the root cause."
6	The killer's name is Omar Saddiqui Mateen. He's suspected to have ties to the Islamic State. This was an act of radical Islamic terrorism.
7	@StockMonsterUSA: Sutherland Springs, Texas Killer Devin Patrick Kelley is being said to be a Radical Alt-Left Antifa member.
8	@DiamondBar8 @atticascott: I never heard a word from you about taking guns & violence out of movies &; off violent TV programs & video games
9	What happened on Sunday in #Orlando was a horrific terrorist attack and was inspired by radical Islamic jihadism.
10	All shooters had prescription psychotropics in common. I never liked these drugs. We are looking for more established articles, but this is buried pretty deep.
11	Whether it turns out to be "terror" related or not, ALL mass shootings are terrorism. And yes,easy access to guns made it possible. #Orlando
12	If we look at the data one thing stands out: mass shootings have roots in domestic violence. Close the #DomesticViolenceLoophole
13	Yet again, a horrific mass shooting was carried out by a gunman with a history of domestic violence.
14	This tragedy underscores critical issues-from the gunman's ability 2 access guns despite clear danger signs 2 the need 4 better mental health care. Our troops put their lives on the line 2 defend our freedoms, but when they need help we don't protect them.
15	Praying for the victims of the Tree of Life Synagogue in Pittsburgh & for our nation to summon the moral courage to stop hate-filled mass shooting We stand with the Jewish community. Good people of all faiths must do more to combat anti-Semitism &; religious intolerance.
16	Antisemitism, hate & violence have no place in our society. The senseless murder of 11 people at the #TreeofLifeSynagogue is a heinous act of #domesticterrorism. We must stand together against the evil forces of hate and fear that seek to divide us.
17	Why are 100 Americans killed by guns every day? Hint: It's not mental health. It's not video games. It's guns. We have a moral responsibility to take action NOW. I've called for change & for @realDonaldTrump, @senatemajldr, & @housegop, to #EndGunViolence.
18	We just observed a moment of silence for the #Orlando victims. Then @SpeakerRyan refused to act to keep guns out of the hands of terrorists.
19	After the worst mass shooting in US history, we must be vigilant against terrorism. And ask how assault rifles make mass killings so easy?
20	The El Paso murderer published a horrific anti-immigrant manifesto just before the shooting. FBI rightly treating this as domestic terrorism. White extremism is on the rise, fueled by xenophobic hateful rhetoric–and directly connected to shootings.

Notes. - Sample of 10 tweets predicted with more than 99% probability as $\widehat{FN}$, that is representative for a Fake News Source (top), and $\widehat{RN}$, that is representative for a R News Source tweets (bottom).

Table A.3.7: Effect of Mass Shootings on News Similarity to Evening News - Heterogeneity

	Dependent variable: News Similarity							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post ×	0.193***	0.027						
Before Election	(0.030)	(0.022)						
Post ×		0.171***						
Pro-Gun × Before Election		(0.037)						
Post ×			0.071***	0.006				
Big Shoot			(0.027)	(0.017)				
Post ×				0.054*				
Pro-Gun × Big Shoot				(0.032)				
Post ×					−0.025	−0.015		
Incumbent					(0.017)	(0.014)		
Post ×						−0.016		
Pro-Gun × Incumbent						(0.022)		
Post ×							0.106*	0.064**
Elect Comp							(0.054)	(0.026)
Post ×								0.048
Pro-Gun × Elect Comp								(0.062)
Pol FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	47,720	128,216	109,809	288,241	109,809	288,241	85,435	235,664
R^2	0.052	0.039	0.033	0.031	0.033	0.031	0.039	0.033

Notes. - The table reports OLS estimates from specifications 3.1 and 3.2. The dependent variable is the cosine similarity between evening news and tweet j from politician i on day t, multiplied by 100. Before Election refers to shootings 180 days before a general election, and for the 180 days after equals 0. Big Shoot refers to shootings with an above median number of victims, and 0 otherwise. Incumbent refers to the sample of incumbents, and 0 otherwise. Elect Comp refers to politicians with high anticipated electoral competition, and 0 otherwise. All columns include politician and mass shooting fixed effects. Columns (1),(3),(5),(7) refer to the sample of pro-gun politicians; columns (2),(4),(6),(8) to the full sample. Standard errors are clustered at the politician level (in parenthesis): *p<0.1; **p<0.05; ***p<0.01

A.3.3 Technical Appendix

Conceptual Framework

Highly visible news event, such as natural disasters, economic shocks or malfeasance of elected officials, commonly unveil factual information about policy alternatives or politicians. Voters can base electoral decisions on these events in a broad class of political economy models.

First, in a probabilistic voting model, the news event updates voters beliefs on the state of the world, that in turn can influence the support for policies (Grossman and Helpman, 2019). For example, in the advent of natural disasters, citizens learn about climate change and support more climate-friendly policies. Second, information is the key mechanism for political accountability in monitoring models of political agency (Persson and Tabellini (2002), Besley (2005)). In a retrospective voting setting, news events reveal the latent variable of a politicians performance in office and thereby enter the voting decision - for instance natural disasters enable voters to learn about an incumbents ability through emergency preparedness (Ashworth, Bueno de Mesquita and Friedenberg, 2018). Last, according to agenda setting theory, the news coverage devoted to an issue will influence its salience and therefore the weight that voters attach to the issue in their voting decision (McCombs and Shaw, 1972). If one party has a relative advantage among voters in that issue, the other party faces potential electoral costs.

A stylized example how visible news events affect political outcomes goes as follows:

1. There are two possible states of the world $\theta \in \{\overline{\theta}, \underline{\theta}\}$ and a sequence of news events s signals $\{\overline{\theta}, \underline{\theta}, \overline{\theta}\}$.
2. Voters have the prior $\theta = 1/2$ and after observing s update their beliefs that $P(\theta = \overline{\theta}) > 1/2$ and respectively $P(\theta = \underline{\theta}) < 1/2$.
3. Voters re-evaluate policy preferences or political performance.

This stylized example rests on the implicit assumption that the political system disseminates the relevant facts. But, "those best positioned to provide relevant facts, elected officials [...] lack the incentive to do so" (Kuklinski et al. (2000), p.791). In the sequence from above a politician of type $\underline{\theta}$ or running on a policy platform $\underline{\theta}$, faces electoral costs and therefore the news event is bad news.

The $\underline{\theta}$ politician decides then on her communication strategy, that is between silence, distraction or deception. A politician that stays silent leaves the belief updating of voters after the events uninterrupted.

While it is an option to bear the potential costs of lost votes, this politician can also seek to manipulate the belief formation process of voters. For one, the $\underline{\theta}$ politician can distract the general public. The politician then communicates for instance the following sequence $Distract$: $\{\overline{\theta}, \underline{\theta}, \emptyset\}$. That is, at $t = 3$ the politician seeks to deflect citizens' attention and media reporting away from the original event. Voters receiving signal do not update their priors after the second period, and conclude that $P(\theta = \overline{\theta}) = 1/2$ and $P(\theta = \underline{\theta}) = 1/2$.

Alternatively, the $\underline{\theta}$ politician can misreport the state of the world. In this case, the politician signals $Deceive$: $\{\overline{\theta}, \underline{\theta}, \underline{\theta}\}$. A politician could for example deny the link between climate change and the frequency of natural disasters. Upon receiving this sequence, the voter supposes that $P(\theta = \overline{\theta}) < 1/2$ and $P(\theta = \underline{\theta}) > 1/2$ and potentially support policies that follow from $\underline{\theta}$.

A.3.4 Measuring Misrepresentation of the State of the World

In this section I describe my measure for misrepresentation of the state of the world. I start by clarifying my conceptual approach and place it in the literature for fake news (FN) detection. Then, I explain the data selection, specifically the isolation from statements about the causes of mass shootings in relevant news articles. Next, I outline how I trained the main neural network model and report it's performance on the news outlet data. In a last step, I validate my measure on the tweets. I provide a stylized overview of this training process in Section A.3.4.

Conceptual Approach

I use a variant of the source-based approach of FN-classification that has been adopted in the economics (Allcott and Gentzkow (2017), Guess, Nyhan and Reifler (2018)) and computer science literature (Vosoughi, Roy and Aral (2018), Helmstetter and Paulheim (2018), Zellers et al. (2019)). One part in this literature (e.g., Allcott and Gentzkow (2017)) relies on small-scale, hand-labelled datasets from fact-checkers. These datasets consist of single news articles from FN-Outlets labelled as fake from journalists from fact-checking organisations. This enables and confines the researcher to track the spread of fake news only through the respective articles on social media.

The other part in this literature (e.g., Helmstetter and Paulheim (2018)) collects large-scale text-datasets from sources that are classified as FN-sources and RN-sources. The researchers then train text-classification models that learn to distinguish between RN and FN based on characteristic text-features. This source-based approach treats all statements from FN-outlets as factually false and all statements by RN-outlets as factually true. The main conceptual concern for classifying misleading information based on the *sources* is that FN-outlets usually spread a mix of RN-*content* and FN-*content*.

The proponents of the text-based approach (e.g., Helmstetter and Paulheim (2018)) argue that the machine learning model is able to successfully identify RN-*content* from a FN-*source* under two conditions: *(i)* the news samples should contain information from both outlets for the same time period and *(ii)* the number of documents from RN-sources exceeds the number of documents from FN-sources by a multitude. If a FN-source then shares factually correct information, it is likely that the same information is as well reported in RN-sources. Given that RN statements are larger by a multitude, an algorithm improves the accuracy of its predictions by classifying it as RN.

This argument is particularly likely to hold for my case, where the text-inputs for the machine learning model are *(i)* statements about the causes of mass shootings from the same period of time within the same *narrow topic* and *(ii)* the number of documents from RN-outlets outnumbers the FN-outlets by a factor of four. Hence, I expect the classifier to label RN-content from a FN-source as real news. The tight focus on statements about the causes of mass shootings also helps to improve on interpretability. While the text-algorithms in Helmstetter and Paulheim (2018) achieve a high performance in separating statements from FN-sources from RN-sources, it's unclear if the machine learning model picks up differences in factual accounts, differences in topics, or differences in stylistic elements.

Identifying Statements about the Causes of Mass Shootings

Data Selection

Before identifying statements about the causes of mass shootings from newspaper articles, I apply three pre-processing steps. First, I disaggregate news articles into paragraphs, as the unit of analysis. While providing more context than a single sentence, the commonly assumed practice of mixing truthful with untruthful content of fake news outlets seems plausibly more severe at the article-level rather than in a paragraph. Thereby I end up with around 151,000 paragraphs (RN: 120,000; FN: 31,000).

Second, I exclude from the analysis all paragraphs and tweets, that contain the words "President", "Trump", "Obama", "Republican", "Democrat", "GOP". Newspapers and politicians tend to accuse the political opponent of making false claims. Consider the sentence "Trump falsely claimed the shooter had mental illnesses." This sentence occurs relatively frequent in tweets from anti-gun politicians rather than in pro-gun tweets as well as in RN-sources rather than FN-sources. The inclusion of those type of sentences harms the classification of misleading content. For one, the distribution of causes of mass shootings under the RN-label will be incorrectly skewed towards (false) statements from the political parties. For another, the algorithm will learn to associate the accusation "Trump falsely claimed" with the RN-label, find it as well in anti-gun tweets and thus induces a mechanical correlation.

Third, I replace the name of the perpetrators with the phrase "shooter". Thereby, the neural network learns to generalize from statements about the reasons why a *specific* perpetrator conducted a mass shooting to the causes of mass shootings in general.[1]

Identifying Statements about the Causes of Mass Shootings

In order to isolate paragraphs that discuss the causes of mass shootings, I first run a topicmodel. Given that the texts are not tweets but longer news paragraphs, I adopt a LDA-topicmodel with 80 topics. Out of these 80 topics, I link 9 topics to the following causes of mass shootings: Guns, Mental Illnesses, Violent Video Games, Terror (Domestic, Foreign). The categories accessibility to guns, mental health conditions and violent video games are often discussed explanations for shootings in academic research (Metzl and MacLeish (2015), Duwe (2016)). The terror topic arises more recently and connects shootings to either foreign terror, that is to Islamist motives or domestic terror, that is concerned with (hate) attacks against racial or religious minorities. Keeping all paragraphs whose most likely topic is associated with these causes of mass shootings reduces the sample to 38,000 paragraphs (RN: 28,000; FN: 10,000).

However, often the topic-paragraphs don't make associations between mass shootings and the causes. For instance, a topic paragraph related to mental illnesses doesn't necessarily discuss mental health *in relation to* mass shootings. Therefore, I apply a supervised machine learning approach to identify statements about the causes of mass shootings. I draw a 10% sample of the topic paragraphs and label paragraphs that makes an association between mass shootings/shooters and one of the causes as 1, and 0 otherwise. I train for each of the two news sources a neural network algorithm, comparable to the one described in Section A.3.4. The respective neural networks reach an F1-Score of 0.81 on the FN-sources and 0.78 on the RN-sources on a hold-out. Applying these neural networks to the full set of topic paragraphs yields 1,050 statements from FN-sources and 4,377 statements from RN-sources.

I list the distribution of causes of mass shootings by type of news outlet in Table A.3.5. Two main

[1] This choice is without loss of generality as I found that in the Twitter-corpus the shooters' names are hardly ever mentioned.

patterns emerge. Intuitively, among RN-sources explanations centering around the availability of guns are the mode. Note, that substantial shares are also devoted to mental illnesses, hate speech and terror attacks, lending support to that my RN-measure relies on a broad ideological base of news sources. Second, among FN-sources terror attacks and mental illnesses feature relatively more prominent.

Training the Neural Network

After the identification of statements about the causes of mass shootings from RN-sources and FN-sources, I describe how I train the main neural network model.

Neural networks are the state-of-the-art tool for text analysis in the computer sciences. At a very general level, neural networks parametrically map complex relationships from an input to an output. The inputs for the my neural network a twofold: The text inputs are the paragraphs related to the causes of mass shootings from RN and FN sources, plus a random sample that is drawn from the paragraphs that don't deal with the topics associated with the causes of mass shootings. The inclusion of the third category is necessary as a large number of tweets that speaks about the mass shootings will not discuss its causes. Experimenting with different versions of input data, I chose a version of the data that includes the 1,050 paragraphs about the causes from FN-sources, 2,100 paragraphs about the causes from RN-sources and 4,200 paragraphs that speak about mass shootings but not it's causes.[2] Moreover, I include two-numerical features to the neural network: the number of words in the paragraph as well as the relative event time when the paragraph was written. As demonstrated in Helmstetter and Paulheim (2018), the predictive power of the neural networks increases upon the inclusion of numerical features. I include the paragraph length to account for potential differences between only-online FN-Sources and partially-print RN-sources. Additionally, I include the relative event-time, as previous research has shown that FN spreads substantially faster than real-news (Vosoughi, Roy and Aral, 2018). I use 80% of the data as the training sample and then evaluate the performance of the trained neural network on a hold-out set of the remaining 20%.

The architecture of a neural network model comprises three types of layers where each layer is build from different neurons. The first layer is the input layer that takes the tokenized

[2] I tested versions of the model that use 2,3 or 4 times the number of RN-sources as compared to FN-paragraphs about the causes. In my sample, the 2:1 relation achieved the highest predictive power. This not much different from the algorithm in Helmstetter and Paulheim (2018) that uses a ratio of 2.6. Importantly, I ensured that the relative weight of causal paragraphs from Table A.3.5 remains unchanged.

and preprocessed news-statements and transforms the words into a numerical feature representation. One of the main advantages of neural networks is that they take into account the meaning of words (Gentzkow, Kelly and Taddy, 2019). For each word in a sentence, the word-embedding representation stores in the background an internal dictionary of the closest associated words. I use the Twitter GloVe embeddings (Pennington, Socher and Manning, 2014), as the ultimate goal is classify tweets.
The transformed inputs are then passed to the intermediate layer, that consists of a recurrent neural network (RNN) with bidirectional Long Short-Term Memory (LSTM) units (Hochreiter and Schmidhuber, 1997).[3] Their main building block is an internal state that can capture context information for each word and "memory cells" that store information about past features that is not determined a priori. Consider the sentence "The main driver of mass shootings in the United States are mental illnesses". LSTMs can infer the connection between "mass shootings" and "mental illnesses" over fully flexible context windows. Moreover, the bidirectional representation of features allows the neural network to utilize both the previous and future context, i.e. the LSTM recognizes the connection of causes and shootings irrespective of the order in which they appear.
Last, the output layer of the neural network assigns to each paragraph a probability score for each label.

To set up a neural network the researcher needs to choose a host of hyperparameters, such as the type of layers, combinations and depth of layers, activation functions. Following best practices in the computer science literature and experimentation I obtain the following model (for more details see A.3.4 and A.3.6). In the text-input layer, I use the 4,000 most frequent tokens, pad paragraphs after 50 tokens and use the Twitter-version of the Stanford GloVe word embeddings with a dimensionality of 200. This text-representation is passed to a bidirectional LSTM with 128 units. I concatenate the output of the LSTM with the numerical data, that is the length of the news paragraphs and the relative-event time. I pass the joint input to a fully connected (dense) layer with 32 units. The output layer then combines this information for its prediction. I tackle overfitting by choosing a low number of parameters and including various dropout and max-pooling layers. Last, I account for the different class sizes through a reweighting of the loss function, i.e. the internal optimizer of the neural network.

This trained model achieves an F1-score of 0.76 (Recall 0.74, Precision 0.78) in the test-set.

[3] LSTMs are widely employed by computer scientists and "responsible for many state-of-the-art sequence modelling results" (Goldberg (2016), p.400), including Googles' language representation model BERT (Devlin et al., 2018).

The confusion matrix A.3.4 shows that a wide majority of the predictions is located on the diagonal and hence classified correctly. The precision of the model is higher than its' recall, that is the mass in the cells in the lower-left part from the diagonal is lower than the counterparts in the upper-right part. In particular, the neural network exhibits a particularly low number of false positives from RN-sources. This tendency strongly limits concerns that downstream effects on the Twitter data are driven by false positives. Moreover, the confusion matrix displays relatively higher number in the false negatives from FN-sources. This pattern is assertive that the neural network detects the RN-*content* from FN-sources. Note also, that the majority of errors accrues *within* RN-sources, between statements that make associations between causes and shootings ($\widehat{RN}$) and statements that just deal with the shootings ($\widehat{NN}$). These mistakes are relatively inexpensive, as the main focus of the neural network is to detect misleading statements.

I apply this trained neural network to predict the label of tweets. This may raise concerns about the ability of the algorithm to detect misleading statements on Twitter. I account for the cross-domain prediction in three ways. First, I used in the training process of the neural network the GloVe Twitter embeddings. Second, I transform hashtags into words (#Orlando → Orlando; #StonemanShooting → Stoneman Shooting). Third, I manually harmonize major idiosyncrasies in my Twitter corpus with the embeddings. For example, I replace the word "ISIL" with "ISIS", as in the GloVe embeddings "ISIL" doesn't occur.

Validation

This section reports a validation exercise. Following Gennaro and Ash (2021), I provide a list of example tweets that speak about mass shootings and get predicted as from a FN-source ($\widehat{FN}$) or from a RN-source ($\widehat{RN}$). Specifically, Table A.3.6 samples tweets that get predicted with more than 99% $\widehat{FN}$ or $\widehat{RN}$.

The examples illustrate clear and intuitive differences between which causes are representative for FN-sources or RN-sources. Tweets predicted from FN-sources ($\widehat{FN}$) connect mass shootings with radical islamic terror, use of antidepressants, uncontrolled immigration, radical left ideas (Antifa) or violent video games. Instead, tweets predicted RN-sources ($\widehat{RN}$) emphasize the role of access to guns, a history of domestic violence, anti-Semitism, and hate crime motives, such as xenophobia. Note that the algorithm also captures subtle differences within specific topics: while mental health issues occur in both samples, they get discussed differently: ($\widehat{FN}$)-tweets emphasize mental health problems per se or in connection to psychotropic drugs,

($\widehat{RN}$)-tweets focus on mental health issues in connection to access to guns. The same holds true for the terror-topic. Tweets related to radical islamic terror get assigned to ($\widehat{FN}$), while tweets related to domestic terror are predicted ($\widehat{RN}$). Overall, the patterns found in the Tweets strengthen the notion that the neural network models picks up meaningful differences in the coverage of explanations of mass shootings provided by different news sources.

Overview of Neural Network Algorithm

1. Data Selection → Corpus of news paragraphs about mass shootings (RN: 120k; FN: 31k)

2. Identify paragraphs that contain statements about causes of mass shootings

   (I) Run Topicmodel → RN: 28k; FN: 10k

      (a) Specification: LDA, 80 topics

      (b) Select paragraphs with highest probability topic coinciding to: guns, health, video games, domestic or foreign terror (9/80 topics)

   (II) Train Neural Network → RN: 4,4k; FN: 1k

      (a) Handlabeled 10% sample whether paragraph connects one of topics to mass shooting, or not

3. Main Neural Network: Detect Misleading Statements about causes of shootings

   (a) Text Inputs: FN:1k; RN:2,1k; NN: 4,2k

   (b) Numerical Inputs: Paragraph Length, Time Difference to Shooting

   (c) Input Parameters: Number of tokens 4000; Sequence Length 50 (Padding); GloVe Twitter Embeddings (Dimensionality: 200)

   (d) Network Structure → see Figure A.3.6

   (e) Train Neural Network using 65% as training data, 15% as validation data and 20% as hold-out set

   (f) Evaluate neural network: F1-score, Recall, Precision on hold-out set

Figure A.3.6: Structure of the Neural Network