DISCUSSION PAPER SERIES

DP15938

# Reconciliation Narratives: The Birth of a Nation after the US Civil War

Elena Esposito, Tiziano Rotesi, Alessandro Saia and Mathias Thoenig

**ECONOMIC HISTORY**

**PUBLIC ECONOMICS**

CEPR

*ISSN 0265-8003*

# Reconciliation Narratives: The Birth of a Nation after the US Civil War

*Elena Esposito, Tiziano Rotesi, Alessandro Saia and Mathias Thoenig*

Discussion Paper DP15938
Published 18 March 2021
Submitted 16 March 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Economic History
- Public Economics

# Reconciliation Narratives: The Birth of a Nation after the US Civil War


## Abstract

We study how the spread of the Lost Cause narrative -a revisionist and racist retelling of the US Civil War- shifted both opinions and behaviors toward reunifying the country and racially alienating African Americans. Drawing on a large set of archival data from between 1910 and 1920, we reconstitute a monthly record of the staggered screenings across US counties of The Birth of a Nation, a blockbuster movie that popularized the Lost Cause narrative across large segments of the population. Our empirical analysis shows that the movie induced (i) a semantic shift in the public discourse toward more patriotic and less divisive language on post-Civil War nation building; (ii) a surge in patriotism with an increased enlistment rate in the US military; and (iii) a cultural convergence between former Confederate and Unionist states with a rise in the adoption of first names traditionally associated with the former enemy's regional identity. We go on to document how the racist content of the narrative helped foster reconciliation through a common enemy rhetorical argument. While we find that the movie strengthened discrimination against African Americans in public discourse and the labor market, our quantitative estimates suggest that 55% of the total effect of the movie on reconciliation was indirectly mediated precisely through this rise in discrimination. All of our findings are detected within both former Confederate and Unionist states.



JEL Classification: D74, N92, Z1

Keywords: Reconciliation, narratives, US Civil War, Segregation, African Americans



Elena Esposito - elena.esposito.1@unil.ch
*University of Lausanne and CEPR*

Tiziano Rotesi - tiziano.rotesi@gmail.com
*University of Lausanne*

Alessandro Saia - a.saia9585@gmail.com
*University of Bologna*

Mathias Thoenig - mathias.thoenig@unil.ch
*Université de Lausanne and CEPR*

# RECONCILIATION NARRATIVES

## –"The Birth of a Nation" after the US Civil War– *

Elena Esposito† Tiziano Rotesi ‡ Alessandro Saia § Mathias Thoenig¶

March 16, 2021

**Abstract**

We study how the spread of the Lost Cause narrative – a revisionist and racist retelling of the US Civil War – shifted both opinions and behaviors toward reunifying the country and racially alienating African Americans. Drawing on a large set of archival data from between 1910 and 1920, we reconstitute a monthly record of the staggered screenings across US counties of *The Birth of a Nation*, a blockbuster movie that popularized the Lost Cause narrative across large segments of the population. Our empirical analysis shows that the movie induced (i) a semantic shift in the public discourse toward more patriotic and less divisive language on post-Civil War nation building; (ii) a surge in patriotism with an increased enlistment rate in the US military; and (iii) a cultural convergence between former Confederate and Unionist states with a rise in the adoption of first names traditionally associated with the former enemy's regional identity. We go on to document how the racist content of the narrative helped foster reconciliation through a "common enemy" rhetorical argument. While we find that the movie strengthened discrimination against African Americans in public discourse and the labor market, our quantitative estimates suggest that 55% of the total effect of the movie on reconciliation was indirectly mediated precisely through this rise in discrimination. All of our findings are detected within both former Confederate and Unionist states.

**Keywords**: Reconciliation, Narratives, US Civil War, Segregation, African Americans

**JEL Codes**: D74, N92, Z1

*We thank Ruben Durante and Hans-Joachim Voth for their helpful comments. We also thank the participants of seminars and conferences at Zurich (May 2019), Cambridge (January 2020), Policy for Peace Conference (February 2020), King's College (February 2021), Queen's University Belfast (February 2021), University of Torino (February 2021), Universitat de Barcelona (February 2021), and University of Bologna (March 2021). We thank Daniel Rosenlehner for excellent research assistance. Elena Esposito, Tiziano Rotesi, and Alessandro Saia acknowledge financial support from the SNF-Grant (185167). Mathias Thoenig acknowledges financial support from the SNF-Grant (182242).

†Department of Economics, University of Lausanne.

‡Department of Economics, University of Lausanne.

§Department of Economics, University of Bologna.

¶Department of Economics, University of Lausanne; and CEPR.

"History is a set of lies agreed upon"
–Napoleon Bonaparte–

# 1 Introduction

Although recurrent wars and persistent hatred are common throughout history,[1] former enemies have also been known to reconcile, sometimes quite suddenly. Famously, such a reconciliation occurred between France and Germany, who fought three wars in less than a century, but went on to become founding nations of the European Union. What are the cultural drivers of these reversals of distrust and animosity?

Historians and social scientists have long emphasized how narratives of past wars can reactivate wounds or, on the contrary, heal minds and foster reconciliation (see Dube et al. 2016). Such views hold that the framing of memories and selective recollection of facts about the causes of conflicts, the deployment of violence, and the settlement of disputes can profoundly influence beliefs and representations. These narratives can take many different forms, from founding myths to divisive expressions of hatred. When it comes to nation building, one frequently observed narrative relates to the existence, real or imagined, of a common enemy.[2] But what is the actual impact of these historical narratives? Do they causally and meaningfully change opinions and behaviors? Or are they rather ex-post rationalizations and window-dressing explanations of economic and political processes that involve deeper stakes and special interests? This paper sheds light on these questions using quantitative empirical methods.

Specifically, we study how the spread of the Lost Cause narrative – a revisionist retelling of the history of the American Civil War (1861-1865) – contributed to national reunification in both the South and the North. With its unparalleled level of destruction and violence, the conflict left American society deeply divided. In the post-war decades, the reconstruction process was slow and fragile.[3] In 1915, an extreme version of the Lost Cause narrative, which had formerly been confined to educated circles, was popularized across large segments of the population by the Hollywood blockbuster *The Birth of a Nation*. The narrative and movie plot are about reconciliation and alienation, structured by a common-enemy logic whereby in the aftermath of the conflict former enemies from Unionist and Confederate states must unite to restore white supremacy, threatened by the enfranchisement of African Americans. While the racist legacy of the movie and its impact on anti-Black violence has long been acknowledged by scholars in American history (see Ang

[1] Collier and Hoeffler (2004) find that more than two thirds of conflict outbreaks took place in countries where multiple conflicts have been recorded. DeRouen and Bercovitch (2008) demonstrate that a large majority of civil wars stem from enduring rivalries. More broadly, Voigtländer and Voth (2012) and Voth (2020) discuss the persistence of hatred, attitudes, and beliefs.

[2] In the recent period, such common enemy rhetoric has been observed in the context of the consolidation of the French Republic after 1870, the Italian Risorgimento, the gradual building of Switzerland and the building of modern Lebanon, among many other examples.

[3] Considerable animus persisted on both sides after the end of the war (see Buck, 1973; Foster, 1988; Silber, 1997 and Blight, 2009, among others). In Appendix Figure A7, we illustrate this fact using newspaper extracts referring to ongoing sectional tensions and debates about the Civil War and its causes and consequences five decades after the end of the conflict.

(2020) for compelling evidence), our aim here is to analyze another, often overlooked, facet of the Lost Cause narrative, namely its impact on reconciliation. In particular, we explore the hypothesis, put forth by David Blight (2009), that the societal cost of reunification was the perpetuated alienation of Black Americans. In Blight's (2009) words: "In the end this is a story of how the forces of reconciliation overwhelmed the emancipationist vision in the national culture, how the inexorable drive for reunion both used and trumped race."

Exploiting longitudinal variation, we explore a variety of attitudes towards reconciliation and alienation in the US between 1910 and 1920 and estimate the impact of the screening of *The Birth of a Nation* (BoaN) on these attitudes at the county level after its national release in 1915. The core of our analysis focuses on reconciliation. Our main outcome variable is a new monthly county-level measure of opinions related to national unity derived from a text analysis of a large dataset of local newspaper articles, containing more than 25 million pages from around 3760 newspapers located across more than 1000 US counties. We also consider two outcome variables that relate to behavioral changes. One measures patriotism by looking at enrollment in the US Navy. The other examines the names of newborn babies, comparing the prevalence of names that evoke the Confederacy and the Union. Indeed, naming decisions have important consequences for children's socio-economic trajectories (Algan et al. 2021) and, in our context, are interpreted as behavioral markers of reconciliation. Our baseline finding is that the screening of the movie induced (i) a semantic shift in public discourse towards a more nationalistic and less divisive rhetoric; (ii) a surge in patriotism with an increased enlistment rate; and (iii) a cultural convergence between former Confederate and Unionist states with an increased adoption of first names traditionally associated with the former enemy's regional identity. We then investigate how the racist content of the narrative contributed to reconciliation between the North and South. While we find that the movie strengthened discrimination against African Americans in public discourse and the labor market, our quantitative estimates suggest that 55% of the movie's impact on reconciliation was indirectly mediated precisely through this rise in discrimination.[4] These empirical results all appear to be stable and statistically robust to a large battery of sensitivity checks (alternative definitions of the variables, estimation samples, econometric models, etc.).

All in all, our findings show that the exposure to the Lost Cause narrative via screenings of *The Birth of a Nation* shifted both opinions and behaviors toward reunifying the country. The logic of a common enemy (namely, African Americans), inherent in the movie's narrative partially served the purposes of reconciliation. Importantly, we detect an impact of this narrative within both former Confederate states and former Unionist states. We interpret this evidence as a quantitative confirmation of just how influential and structuring the Lost Cause has been for American society in its attempt to cope with the long-run legacy of the US Civil War.

The study of reconciliation narratives raises several methodological challenges that our context, data and empirical design enable us to address. First, a general issue concerns their measurement. Here, the fact that the main vector of the large-scale diffusion of the Lost Cause was a movie of-

[4]Our approach is to supplement existing evidence on racial hatred with less violent but potentially more pervasive elements of alienation, since most of the violent expressions of racism (crimes, lynchings and KKK presence) were largely confined to former Confederate states in the 1910s.

fers a unique opportunity to trace its spread within the population in a fine-grained and accurate manner. The 1910s, which immediately preceded the Golden Age of Radio, are the last period in US history in which the media environment consisted of a limited amount of mostly local media sources. Screenings of BoaN in town theaters, with the film's astonishing visual effects (see section 2), may have made long-lasting impressions on audiences who had few alternative sources to inform their beliefs regarding the merits of the Lost Cause narrative. It should be noted that in spite of its tremendous commercial success, there is no existing record of the places and dates of BoaN screenings. However, in the 1910s, movie screenings were publicized in local newspapers, and thus we were able to retrieve this information from digitized ads preserved in the online archive, newspapers.com. The impressive coverage of this archive allows us to reconstruct the path of the movie's diffusion at the county-month level.

A second challenge concerns endogenous exposure to a narrative. With social media, for example, it is now well established that users tend to self-select into the news and informational content to which they are exposed (Bakshy et al. 2016, Halberstam and Knight 2016). In our case, airing the film is likely to be endogenous due to unobserved shifters of the movie demand and supply that correlate with attitudes. We propose a two-way fixed effect estimation strategy, where time and county fixed effects allow us to remove all time-invariant shifters, combined with an instrumental variable setup that accounts for *time-varying* shifters. To this end, we make use of quasi-random variation in screening locations and dates generated by the logistics of movie distribution. Specifically, we instrument screenings of BoaN with screenings of *The Million Dollar Mystery* (MDM), a comedy that was released 231 days before BoaN, fictitiously postponing its spatio-temporal diffusion as if MDM was released on the same date as BoaN. The plot of MDM – a secret society that attempts to gain control of a lost million dollars – is politically neutral and radically different from that of BoaN. The statistical power of the instrument relies on the fact that the spatio-temporal distribution patterns of MDM and BoaN are highly correlated. Our key identifying assumption is that the demand for MDM is exogenous to local opinions and attitudes, particularly those related to reconciliation and alienation.

**Related Literature** Our research contributes to a rich pre-existing literature on the role of memory in US history, including the well-known studies of Buck (1973), Foster (1988) and Blight (2009). By devising strategies to measure attitudes on reconciliation and identify the effect, we contribute to this body of work, offering the first quantitative empirical investigation of the role of memory in reconciliation. Our findings provide support for the hypothesis of David Blight (2009), who holds that the alienation of African Americans was instrumental to national reconciliation after the US Civil War.

We thus shed light on an unexplored side of the Civil War and Reconstruction eras, enriching the empirical economic history literature that explores the deployment (Dippel and Heblich, 2021; and Costa and Kahn, 2010) and consequences of the US Civil War (Costa, Yetter, and DeSomer, 2018; Ager, Boustan, and Eriksson, 2019; Feingenbaum, Lee, and Mezzanotti, 2018) and the history of discrimination against African Americans (Boustan, 2010; Fouka, Mazumder, and Tabellini, 2020;

Shertzer and Walsh, 2019; and Tabellini, 2010, among others).

On a broader scale, by exploring the consequences of the diffusion of the Lost Cause narrative, our research takes up the challenge posed by Shiller (2017) to provide "serious quantitative study of changing popular narratives," thus joining the work of Akerlof and Snower (2016), Cantoni et al. (2017), Mukand and Rodrik (2018), Nyman et al. (2018), and Benabou, Falk and Tirole (2018).

This paper also explores a new fundamental dimension of the reconciliation phase that follows a civil war, and in doing so builds on a nascent empirical literature exploring the effects of reconciliation, peace- and nation-building policies (Cilliers, Dube and Siddiqi, 2016; Fearon, Humphreys, and Weinstein, 2015; Depetris-Chauvin and Durante, 2020).[5] To our knowledge, ours represents the first effort to provide a quantitative empirical analysis of the role of memory for reconciliation relying on plausibly exogenous variation.

In a similar vein, our project also adds to studies exploring the legacies of war. One the one hand, influential research in economics and political science demonstrates that civil wars leave persistent legacies that go well beyond the short- and medium-term physical and economic disruption they cause. To this regard, Blattman and Miguel (2010) argue that the "social and institutional legacies of conflict are arguably the most important but least understood of all war impacts" and may, in fact, explain cycles of violence (Collier, 2003).[6] On the other hand, a rapidly growing body of work – extensively reviewed in Bauer et al. (2016) – shows that direct exposure to war violence can push people to behave in altruistic and pro-social ways in their communities. These findings resonate with a broad literature in evolutionary biology that suggests that external threats can magnify parochial altruism (see, among others, Henrich and Boyd, 2001; Richerson and Boyd, 2001; Bowles, 2006; Choi and Bowles, 2007).

We contribute to these two strands of literature by showing that the legacy of war might crucially depend on the types of *narratives* that prevail in the aftermath of a conflict. Narratives that magnify an internal or external threat might foster in-group cohesion, breaking the persistence of historical tensions and divisions. Our findings, at a more general level, shed light on an open question in the field of cultural persistence, and in doing so respond to Voth's (2020) call for research that elucidates "mechanisms and the conditions under which persistence breaks down." Specifically, we show how the spread of a new narrative about the past can break half a century of tensions and hostilities.

Lastly, by exploiting the diffusion of one of the most important movies in US history as a vector of a narrative, our project also relates to studies that investigate the role of propaganda and mass media in fostering changes in attitudes, beliefs and behaviors, such as political preferences (Adena

[5] Cilliers, Dube and Siddiqi (2016) study truth and reconciliation efforts in the context of post-civil war Sierra Leone, assessing the social and individual effect of reconciliation forums that brought together perpetrators and victims. Fearon, Humphreys, and Weinstein (2015) explore the role of development assistance in collective actions in post-conflict Liberia. Blattman et al. (2011) and Bauer et al. (2018), meanwhile, examine the reintegration of former soldiers into society. This body of research forms part of an extensive interdisciplinary effort involving historians, sociologists, anthropologists and psychologists. Depetris-Chauvin and Durante (2020), for example, use football victories to study the role of shared experience in shaping national identities.

[6] The literature exploring the human and economic legacies of internal warfare is vast. Blattman and Miguel (2010) and Justino (2010) provide extensive reviews. Among the more related contributions, exposure to civil violence has been shown to reduce social capital (Rohner et al., 2013), trust (Cassar et al., 2013), strengthen ethnic identity (Lupu and Peisakhin, 2017), and increase future criminal behaviors (Couttenier et al., 2016).

et al., 2013; Voigtlander and Voth, 2015; Stromberg, 2004; Gentzkow, 2006; Martin and Yurukoglu, 2017) and nationalism and nation-building (Della Vigna et al. 2014; Blouin and Mukand, 2019).[7]

The impact of *The Birth of a Nation* on US history was unprecedented and unparalleled. As Pierce (2013) observes: "motion pictures in the tens and twenties - before network radio, television, cell phones, and the Internet - had an influence that is hard to imagine today." The most compelling empirical evidence of the movie's impact comes from Ang (2020), who documents how exposure to *The Birth of a Nation* increased lynchings of Blacks, race riots and local support for the KKK, with effects on racial violence that persist to the present day. Our paper takes a broader view by contextualizing the movie's main message in the framework of the Lost Cause narrative, which the movie faithfully reflects. Indeed, as suggested by the title, *The Birth of a Nation* is not only a movie about violence and the racial alienation of African Americans, it is also about post-Civil War reconstruction. This perspective allows us to show that the Lost Cause narrative played a key role in national reconciliation, with consequences that went well beyond racial violence. Moreover, we are able to shed light on the deep causes underlying the perpetuation of racial hatred – a question left unexplored by Ang (2020) – as we document how the racist content of the movie was in fact instrumentally exploited to foster reconciliation, structured by a common-enemy argument.[8]

The paper is structured as follows. Section 2 provides a brief historical background and describes our collection and processing of the data related to the movie. Section 3 discusses the empirical strategy. Baseline estimation results for each of the three outcome variables related to reconciliation are then reported in Sections 4, 5 and 6. We investigate the interplay between racial alienation and reconciliation in Section 7, followed by a comparison of the results between the North and the South in Section 8. Section 9 concludes.

## 2 Historical Background and The Birth of a Nation

### 2.1 The US after American Civil War: A Slow Way to Reconciliation

The American Civil War (1861-1865) stands out to this today as a conflict of unparalleled violence and destructiveness. According to estimates, more than 2% of the country's population died in the line of duty.[9] In addition to the direct losses of human and physical capital, however, the war's lasting legacy was a deeply fractured society (Foster, 1988; Blight, 2009). In the South, the country had to deal "with a disaffected people who had aspired to independence and failed" (Buck, 1973),

[7] See also the studies exploring the effect of television on gender norms (Jensen and Oster, 2009; La Ferrara et al., 2012; Kearney and Levine, 2015), consumption behavior (Bursztyn and Cantoni, 2015), and racism and violence (Dahl and Della Vigna, 2009; Yanagizawa-Drott, 2014; Voigtlander and Voth, 2015)

[8] Our study was carried out in parallel to Ang (2020). While both papers use similar (though not identical) data collection approaches, they rely on distinct sources for identifying variations in the instrumentation strategy and look at radically different outcome variables. It is reassuring that both papers provide consistent results in their measurements of the alienation of African Americans. However, here we are interested in forms of racial segregation that are less violent and spectacular, though potentially more pervasive and disruptive.

[9] Recent research combining newly digitized census data from the 19th century calculates the number of military victims to be around 750,000 (Hacker and McPherson, 2011). According Goldin's (1975) estimates, the South lost a total of 683,939 soldiers and civilians while the North lost 954,922. Note that the exact number of military and civilian casualties remains controversial, given the challenges of making reliable quantitative assessments.

while in the victorious North, politicians and public opinion alike appeared determined "to be assertive in the realization of newly found power" (Buck, 1973). While the war allowed the Union to be preserved, the peace did not bring an end to wartime hatred and partisan animosities.

Following the end of hostilities, public memory became a new battlefield. Contending *narratives* regarding the "true" causes of the war, the behaviors of soldiers in battle, the rightness of reconstruction policies and similar issues started to emerge on both sides.[10] In this context, the narrative of the supposed Lost Cause gained particular fame, shaping regional identity and race relations for generations. The Lost Cause was a "rationalization, a cover-up to vindicate the name and fame" of those who lost the war (Gallagher and Nolan, 2000) . Its main tenets : i) played down the importance of slavery in the onset of the conflict, which was rather driven by a desire for liberty on the part of Southerners; ii) idealized the life and values of the antebellum South, depicting the enfranchisement of Black people as a real threat for the country; and iii) pushed for national reconciliation, since neither side had been truly wrong and defeated Southerners had displayed unrivaled heroism and honor.[11] According to many scholars, the Lost Cause was fundamental to the reunification of the country (Buck, 1973; and Blight, 2009). This movement, however, was exclusively restricted to white Americans, and the narrative it supported gave precedence to the need for reconciliation between the white North and white South, obscuring the legacy of emancipation and Black participation in the war (Blight, 2009).

### 2.2 *The Birth of a Nation*

Although the Lost Cause narrative was initially put forth by Southern elites, many of them former Confederate generals, over time it gained popularity throughout the country. A particularly successful vector for the nationwide popularization of an extreme version of this narrative was a Hollywood blockbuster of unparalleled fame: the 1915 movie *The Birth of a Nation* by D. W. Griffith.

The movie's objective was to tell the "true" story of the Civil War (see Appendix A.1 for more details about the plot). In the film, the war is not caused by quarrels over slavery, but escalates because of the ambition of a greedy Northern Republican mulatto politician attempting to enforce Black rule in the region. The emancipation and enfranchisement of African Americans is depicted as a common threat for the white populations of both the North and the South. In the end, thanks to a renewed push by white Northerners and Southerners to unite against this common enemy, national reconciliation and white supremacy are finally restored. In other words, the movie proposes reconciliation between the white North and South based on an exclusively racial basis, blatantly discriminating against African Americans and ignoring their emancipation.

In spite of its disturbing message, the movie was an unprecedented box office success. Its positive reception was largely a consequence of its innovative techniques, aesthetic beauty and storytelling power, along with the comforting narrative it offered to the defeated white South, and more generally to a country seeking reconciliation. At the same time, civil rights activists accused

[10] For insightful accounts on the role of memory and narratives in the post-bellum US see, among others, Buck (1973); Silber (1997); Blight (2009); and Foster (1988).

[11] On the Lost Cause see, among others, Foster (1988) ; Gallagher and Nolan (2000); and Blight (2009).

the movie of fueling dangerous anti-Black sentiment with its racist portrayal of African Americans.

### 2.3 Measuring Movie Diffusion from Historical Newspapers

By 1930, an estimated fifty million people had seen *The Birth of a Nation*.[12] Despite this striking number, no accurate figures exist with respect to either the movie's overall profit or the exact size of the audiences across locations.[13] In the absence of official sources, the most complete, systematic and comparable source of information on the movie's distribution over time and space are local newspapers.[14] We therefore performed an extensive data collection exercise to map the distribution of the movie's diffusion by retrieving information on film screenings from the online newspaper archive *newspapers.com*. This archive covers an extremely large variety of *local* papers. For example, for the state of Alabama, between 1910 and 1920, newspaper records are available for 93 different cities.[15] Appendix Figure A9 maps the counties that were home to the headquarters of at least one newspaper.

Local newspapers provide information on the locations and dates of screenings through advertisements (left panel of Figure 1), movie times in local theaters (right top panel of Figure 1), and reviews of movies being shown in the area. To exploit information from these three sources, we started by collecting all newspaper pages containing the keyword *The Birth of a Nation*. We retrieved a total of 55,044 pages from 1837 newspapers, located in 845 counties.[16] *The Birth of a Nation* was a phenomenal success that set the stage for an intense public debate over the narrative endorsed by the movie. The items retrieved through this keyword search are thus not only related to movie's screenings but are also general reviews of the film or articles discussing reactions to it across the country (right panel of Appendix Figure A10). It should be noted, however, that the searching algorithm excludes the definite and indefinite articles and some hits might be completely unrelated to the movie. Given that our objective is to map screenings of the movie, we asked external judges to read each item in full and assess whether it refers to a movie screening; if not, the item was discarded. This filtering yielded 14,941 "verified" screening records from 990 newspapers located in 635 counties. Random checking of the accuracy of the judges' work indicated that 92% of the verified records correctly identified screenings of *The Birth of a Nation* in the county.

We then associated each verified screening record to the month of screening and the county where the newspaper's headquarters are located. In order to correctly measure the movie's diffusion, it is of utmost importance to distinguish county-months that had no screening (i.e., a true zero)

[12] A number set forth by Carl E. Milliken, secretary of the Motion Picture Producers and Distributors of America (MPPDA), cited by Stokes (2007). As Stokes (2007) recalls, "Local newspapers made approximate estimates of the audiences who had seen the film toward the end of its run in particular cities: 185,000 in Boston, 100,000 in Kansas..."

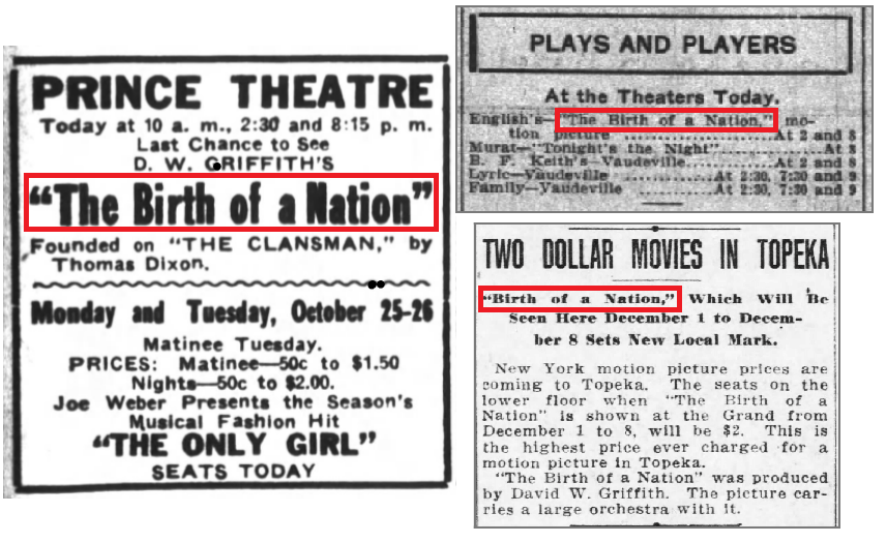
[13] This may also be due to the fact that the company that produced the movie (Epoch Producing Company) sold local distribution rights to other companies. For instance, Harry Sherman of Minneapolis acquired the rights to show the film in sixteen western states, while a syndicate headed by Louis B. Mayer acquired the rights to distribute the movie in New England (Stokes, 2007).

[14] This was also aided by the marketing strategy of the distributing companies to advertise the movie's release in theaters in "local newspapers across the country through press agencies and wire services" (Stokes, 2007).

[15] Note that other states, however, have much less fine-grained coverage. For the state of Wyoming, for instance, we only have access to newspapers from three cities.

[16] Note that the repository only allows us to retrieve the number of newspaper pages containing each keyword and not the total number of keywords.

Figure 1: Newspaper Page with *The Birth of a Nation* - Ads and Movie Table




NOTE: *Left Figure* An advertisement documenting the screening of the movie *The Birth of a Nation* at the Prince Theater in Houston, Texas. *The Houston Post*, 23 October 1915. *Right Top Figure* Movie Table documenting the screening of *The Birth of a Nation* in Indianapolis, Indiana. *The Indianapolis News*, 14 December 1915. *Right Bottom Figure* Newspaper article documenting the screening of *The Birth of a Nation* in Topeka, Kansas. *The Topeka Daily Capital*, 15 October 1915. Source: newspapers.com

from county-months for which no digitized newspapers exist in the archive (i.e., a missing record). To avoid the inclusion of missing records, we only consider county-months with at least one digitized newspaper page for the month.[17] In addition, we always control for newspaper coverage at the county-month level in our econometric specifications. Finally, since the set of newspapers covered by the newspapers.com archive varies over time, our final sample is unbalanced. We therefore also verify that our results hold with a balanced subsample of our dataset.

Our empirical analysis aims to assess changes in attitudes after exposure to the movie. As a baseline measure, we create a county-level treatment variable, which we label $\texttt{BoaN}_{ct}$, which takes a value of 1 (and 0 otherwise) for all months after at least two verified screening records were found in a given county.[18] Retaining a threshold of two records minimizes measurement error related to instances in which we erroneously consider the keyword *The Birth of a Nation* as proof of a screening (approx. 8% of items).[19] In our sensitivity analysis, we show that the results are

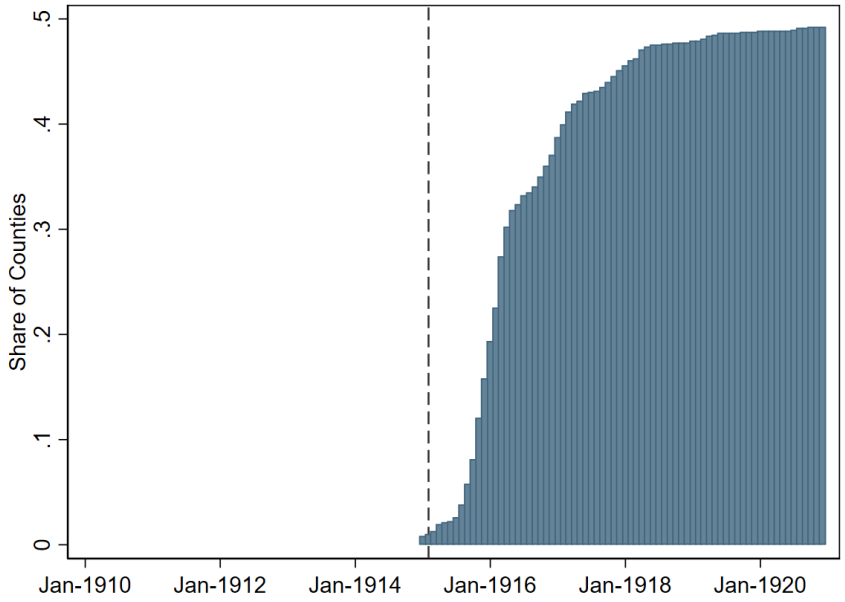
[17]We use two approaches to detect whether a county has at least one digitized newspaper page for a given month. First, we retrieve information on newspaper coverage provided by newspapers.com (coverage condition #1) available at https://www.newspapers.com/browse/us. Given that the archive contains an immense amount of digitized newspaper pages, coverage information may imprecisely capture month-to-month variation. To refine the precision of our coverage measure we add a second condition, searching all newspaper pages containing at least one of the keywords "he," "you" or "I" (coverage condition #2). If at least one of these two conditions is met, we include the county-month in the sample.

[18]Note that the verified screening records do not necessarily identify distinct screening events. For instance, in the case of an op-ed article announcing the arrival of the movie in the local theater and and a timetable of movie screenings in town, both records may refer to the same event.

[19]Some screening records may also refer to screenings taking place in the following month. The threshold of two records allows us to reduce the impact of information related to the following months, improving the precision of the

robust to alternative threshold values. Figure 2 displays the time-series evolution of the share of counties that had screened the movie, i.e., $\texttt{BoaN}_{ct} = 1$ for our baseline definition of the threshold (two records). A visual inspection reveals that the movie had been screened in about 16% of the counties in our sample by the end of 1915, 45% by the end of 1916 and 50% by the end of 1917. The left panel of Figure 3 displays the spatial dispersion of counties that had screened the movie (or not) by the end of our sample in 1920.

Figure 2: *The Birth of a Nation* over Time




NOTE: The figure depicts the share of counties that screened *The Birth of a Nation* over time, which corresponds to the monthly average of the variable $\texttt{BoaN}_{ct}$ as defined in the text.

# 3 Empirical Design

We are interested in assessing the impact of screenings of the movie *The Birth of a Nation* (BoaN) on an array of attitudes toward reconciliation and alienation between 1910 and 1920 at the county level.

## 3.1 Econometric Equation

The film was officially released in February 1915 and gradually aired across the country. Given its staggered distribution in space and time, a natural starting point for the empirical design is a a two-way fixed (FE) effects estimation, controlling for county and time fixed effects. Hence, our generic approach consists of estimating the treatment effect (i.e. the coefficient $\beta$) of the binary variable BoaN through OLS in the following econometric equation:

$$\texttt{outcome}_{ct} = \alpha + \beta \times \texttt{BoaN}_{ct} + \sum_{c} \texttt{FE}_{c} + \sum_{t} \texttt{FE}_{t} + \epsilon_{ct}. \tag{1}$$

measure.

The unit of observation is a county ($c$) $\times$ period ($t$) cell, where a period is a year-month (e.g., March 1915), the variable $\texttt{BoaN}_{ct}$ codes for the post-screening period and is equal to 1 in all periods following the screening and 0 otherwise, and $\texttt{FE}_c$ and $\texttt{FE}_t$ are county and time fixed effects. The variable `outcome` varies across specifications depending on the attitude we are scrutinizing: type of rhetoric in newspaper articles, enlistment in the Navy, cultural type of name, supremacism, or discrimination in the labor market. We explain below how each variable is constructed and provide the related estimation results. Note that in all cases, `outcome` is a county-level variable, except in the case of names, where it is defined at the individual level. In the baseline analysis, standard errors are clustered at the county level, which corresponds to the (relevant) aggregation level of the treatment $\texttt{BoaN}_{ct}$. In robustness checks, we allow for more complex clustering structures of the estimator of the VCV matrix that take into account the spatial nature of the error terms $\epsilon_{ct}$ (Collela et al. 2020).

In equation (1), the main estimation challenge relates to the potentially endogenous nature of the treatment $\texttt{BoaN}_{ct}$. Balancing tests (Appendix Table A1) quantitatively confirm that counties that aired the film at some point are different from those that did not. These tests are rejected for several observable factors: more populated areas and ones with a lower share of rural population or a higher share of literate inhabitants were more likely to receive the film. Importantly, these factors are also susceptible to correlating with $\texttt{outcome}_{ct}$. For example, cities tend to be more progressive than rural counties and to have more theaters, two features that impact their odds of airing the film (i.e., shifters of the movie demand and supply, respectively).[20]

The two-way fixed (FE) effects estimation strategy alleviates this endogeneity concern as long as the so-called parallel (or common) trend assumption holds. Here, identification assumes that factors affecting movie screening exert a time-invariant influence on the outcome variable that can be filtered out by the battery of fixed effects at the estimation stage. However, in the context of BoaN, the presence of time-varying confounders is a real concern. As documented above, the film fueled protests and social unrest and magnified existing cleavages in politically polarized places, with calls for censorship and attempts to ban the movie (Stokes 2007).[21] Areas in the midst of social conflict might be more or less likely to screen the movie depending on the local context. Another confounder could be economic distress, which drives both the odds of screening the movie and attitudes.[22] For example, in rural counties (73% of our sample), transient climate shocks could adversely impact the local economy and potentially reactivate frustration, angers and scapegoating.

[20]The presence of movie theaters is highly correlated with county-level population. Using 1910 census data on the share of projectionists in the population as a proxy for the presence of movie theaters, we find a correlation of 0.93 between population and the presence of theaters.

[21]Protests and organized political campaigns attempting to ban the movie started to emerge around the country with the very first public screenings, with mixed results. In February 1915, in Los Angeles, the National Association for the Advancement of Colored People (NAACP), together with other organizations, appealed to local authorities and courts for the movie to be banned, complaining that it encouraged "bitterness and strife between the races" (Stokes 2007). The movie matinée was canceled, but not the evening première. Similar mixed results were achieved in the rest of the country, where political campaigns to stop the movie clashed with a myriad of different local rules about film censorship (Stokes 2007). In some cities, such as Chicago, protesters managed to have the movie banned for brief periods of time, while it was prohibited for longer in Kansas and Ohio (Stokes 2007).

[22]There is a rich literature on economic distress and attitudes. Important contributions include Autor at al., 2020; Guiso, Herrera, Morella and Sonno, 2017; Dal Bo, Finan, Folke, Rikne, and Persson, 2017; Pastor and Veronesi, 2018; Piketty, 2018; Guriev, 2018; and, for a review, Guriev and Papaioannou, 2020.

Since we cannot rule out the existence of time-varying confounding factors, two-way fixed estimates might be biased. The previous examples make clear that the direction of the bias can go in both directions, upward or downward. One way of dealing with this issue is to include flexible controls and county-specific polynomial time trends in the OLS estimation of (1) (see Section 4.4). Both solutions are demanding of the data and somewhat limited in scope. Flexible controls deal only with observable factors – e.g., urban areas with fast-growing populations and quickly evolving opinions and attitudes – but our historical data restricts the set of observables at hand.[23] Meanwhile, county-specific trends may involve unobserved factors (e.g., local memories of the Civil War), but at the cost of imposing parametric restrictions on their time-series impact.[24] Given these limitations, our preferred solution is to instrument the treatment in a 2SLS version of equation (1). We explain the logic and construction of our instrument in the next section.

### 3.2 Instrumenting the Screening of *The Birth of a Nation*

***The Million Dollar Mystery.*** The idea behind our instrumental variable strategy is that the distribution of movies in the US of the early 20th century followed recurrent spatio-temporal patterns because of logistical constraints related to movie theaters, pre-existing agreements between theaters, and shipping limitations on the number of film reels across locations.

As our baseline instrument, we rely on the movie *The Million Dollar Mystery* (MDM), which was the highest-grossing movie before BoaN. It was released in June 1914, 231 days before BoaN. The movie tells the story of a secret society that attempts to gain control of a lost million dollars. Hence, its plot is *radically different* from that of BoaN.[25] We retrieved information on screenings on the *The Million Dollar Mystery* using a data collection strategy similar to that employed for *The Birth of a Nation*.[26] We then define our instrumental variable $\texttt{MDM}_{ct}$ as an indicator variable coding for the post-screening period of the movie, transposed 231 days later: Specifically, $\texttt{MDM}_{ct} = 1$ (0 otherwise) if 2 MDM screening records in county $c$ have been collected from local newspapers before date $t - 231$ days. Our premise is that $\texttt{MDM}_{ct}$ is a relevant and exogenous predictor of $\texttt{BoaN}_{ct}$ in both the cross-sectional and time-series dimensions.

**Identification Assumption.** The relevance of the instrument relies on the fact that the spatio-temporal distribution of MDM *transposed 231 days later* correlates quite closely with the spatio-temporal distribution of BoaN. The correlation is tied to logistical constraints specific to the film

[23] In practice, flexible controls imply that for all observables $X_i$, which appear to be unbalanced, we include $X_i \times FE_t$ on the RHS of equation (1).

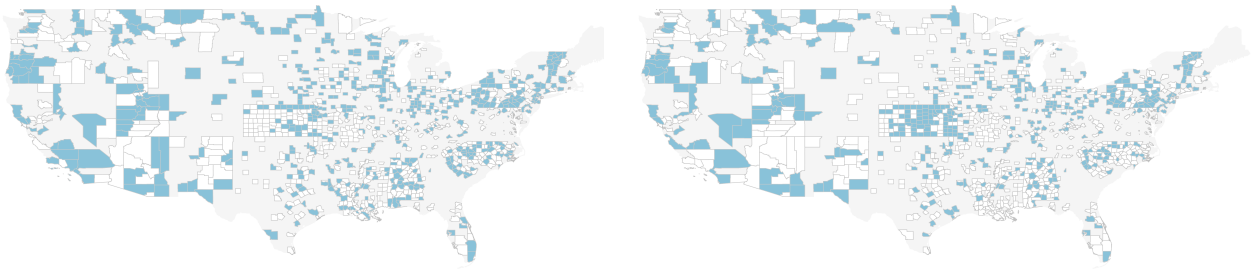
[24] Whenever the effect of interest changes over time, group-specific linear time trends are less than ideal, since the trend might also capture the difference in the evolution of the outcome between treated and control variables, biasing estimates (see, for instance, Wolfers, 2006; and Kahn-Lang and Lang, 2019).

[25] *The Million Dollar Mystery* (MDM) is an ideal choice because, while it deals with very different topics with respect to BoaN, it also created a true mania across the country. Its serial format provided a richer and more engaging experience, increasing audiences and revenues. Bean (2017) describes the mobs congregating at theaters on "Mystery Nights," waiting to find a seat. This format also increases the likelihood of the movie being screened in large storefront theaters with a regular movie programming season, which were also the theaters where *The Birth of a Nation*, distributed with a cohort of projectionists, electricians, other technicians and orchestra, was more likely to be screened.

[26] In a nutshell, we started with a search for the keyword "The Million Dollar Mystery," which resulted in the retrieval of 25,858 items. Out of these, judges verified that 18,168 (70%) were related to actual screenings of the movie in a county.

industry of this period. The first and main constraint pertains to the absence of a movie theater in many places. This feature makes MDM a powerful instrument for BoaN in the cross-sectional dimension (i.e., whether BoaN was screened or not). Secondly, the number of reels available for a given film was limited and they had to be physically shipped across the country. On top of this, established distribution practices targeted big cities first, where higher admission prices could be charged (Nowell-Smith, 1996). Consequently, the time elapsed between national release and local screenings tended to be similar across movies. These features make MDM an appropriate instrument for the time-series dimension (i.e., when BoaN was screened). Finally, our key empirical assumption is that the airing of MDM is exogenous to local opinions and attitudes, in particular those related to reconciliation and alienation. This makes sense, given that MDM was basically a comedy with a politically neutral plot. As robustness tests, we replicate our instrumentation strategy with other movies that were released before MDM: *Traffic in Souls*, released in 1913, and *What Happened to Mary*, released in 1912.[27]

Figure 3: BoaN and MDM Screenings - Maps



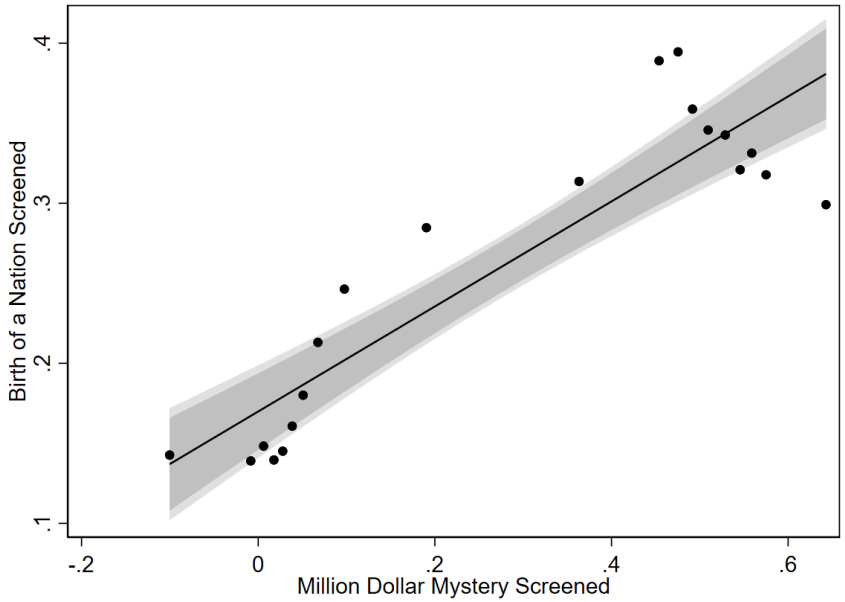
NOTE: *Left Figure* The figure maps the counties that had screened *The Birth of a Nation* by 1920. Blue counties were exposed to the movie, while white counties were not. *Right Figure* The figure maps the counties that had screened *The Million Dollar Mystery* by 1920. The correlation between counties that screened BoaN and counties that screened MDM is visible to the naked eye, with the notable exception of Kansas, where *Birth of a Nation* was banned starting from 1916.

**Some Evidence.** Figure 3 (right panel) maps the spatial distribution of counties that had screened *The Million Dollar Mystery* by 1920. They represent 43% of counties in our sample and around 70% of the counties that had also screened *The Birth of a Nation* (see Appendix Table A2). Figure 4 provides statistical evidence that $\texttt{MDM}_{ct}$ is a strong and relevant predictor of $\texttt{BoaN}_{ct}$. The binned scatter plot displays the conditional relationship between the two variables in the full panel, after filtering out county and month-year fixed effects (see equation 1).

**Instrumented specifications.** We estimate two instrumented versions of equation (1): a standard 2SLS estimator with $\texttt{MDM}_{ct}$ as the instrument and a reduced form specification (RF) with $\texttt{MDM}_{ct}$ as the explanatory variable. OLS and 2SLS estimated coefficients might differ for several reasons. As

[27] As further discussed in Appendix Section B.4, as expected, *The Million Dollar Mystery* reached a similar share of counties as *The Birth of a Nation*. Meanwhile, *Traffic in Souls* and *What Happened to Mary* were screened in a sizeably lower number of counties.

Figure 4: BoaN and MDM Screenings - Binned Scatterplots




NOTE: Binned scatter of the conditional correlation between $\texttt{BoaN}_{ct}$ and $\texttt{MDM}_{ct}$ after filtering out county and month-year fixed effects.

discussed above, OLS can be biased downward or upward depending on the confounding factors at play. Moreover, dealing with two dichotomous variables as endogenous variable and instrument can concur to magnify the difference between OLS and IV estimates (see, for instance, Pischke, 2007 and Black, Berger and Scott, 1995).[28]

# 4 Reconciliation in the Public Debate

In the United States of the 1910s, before the advent of radio and television, public debates took place entirely in newspapers, which were inexpensive and ubiquitous. Local newspapers are therefore the natural place to monitor the prevalent attitudes towards reconciliation throughout the country. To measure their evolution, we study how language evolved in newspapers in the online archive *newspapers.com* (more than 25 million pages from around 3760 newspapers with headquarters located in over 1000 US counties). The main limitation of this historical depository is that it does not permit access to the full text of articles. We therefore perform a text analysis exercise with a *bag of words* approach, in the spirit of Gentzkow and Shapiro (2010) and Enke (2020), among many others.

## 4.1 Measuring Rhetoric in Newspaper Articles

We selected a set of 12 keywords that relate to reconciliation, patriotism and national unity – which we label *Patriotic* – and a set of 12 keywords that recall the Civil War, Reconstruction and the sec-

[28] Such a magnification arises if the share of counties that screened BoaN and are erroneously classified as non-BoaN is higher that the share of counties that did not screen BoaN and are erroneously coded as BoaN counties. In this situation, OLS is biased downward while the IV coefficient is biased upward. OLS and IV estimates can then be used to bound the true coefficient. The issue is further discussed in Appendix Section B.3.3.

tionalism of the times – which we label *Divisive*.[29] Given that the keywords were selected on a discretionary basis, we devise an array of strategies to demonstrate that any subset and/or recombination of our baseline lists provide consistent and highly correlated measures of reconciliation and division (see Appendix B.2.2). For each keyword $i$ and county$\times$month cell, we compute the log frequency of the keyword's occurrence as

$$\texttt{logFreq}_{i,c,t} \equiv \log\left[\frac{1+\texttt{Page}_{i,c,t}}{1+\texttt{TotPage}_{c,t}}\right] \tag{2}$$

where $\texttt{Page}_{i,c,t}$ stands for the number of newspaper pages where $i$ appears (at least once) and $\texttt{TotPage}_{c,t}$ measures the total number of newspaper pages in county $c$ and month $t$.[30] Importantly, to avoid the possibility of keywords on our list referring directly to the plot of *The Birth of a Nation*, we exclude all pages containing the keyword *The Birth of a Nation* from the word counting exercise.

Considering frequency (rather than number of occurrences) controls for variations in the availability of digitized newspapers and limits measurement errors by reducing the influence of outliers.[31] Moreover, taking the log allows for comparing changes across keywords with different baseline average frequencies of occurrence (i.e., rare vs frequent keywords). When computing frequency, the +1 transformation is a standard and convenient way of dealing with zeros in keyword occurrence. However, this functional form can cause distortions in the distribution of the variable, especially when occurrence is low. Hence, in Appendix B.2.1, we propose alternative definitions and coding options for keyword occurrence (e.g., without log, without +1, etc.). Further, to make sure that our results are not driven by anomalous variations in newspaper coverage, we account for the total number of newspaper pages available in the county-month in a flexible way by including fixed effects for page number percentiles in the baseline analysis (and decile fixed effects as a robustness exercise).

### 4.2 Regionalism vs United Country

As a first pass at the data, our empirical analysis starts with log frequency as the dependent variable in Equation (1). This preliminary approach has the virtue of transparency but is not parsimonious because it must be conducted separately for each keyword. For the sake of illustration, we show the results for two keywords. Specifically, we look at how the screening of BoaN impacted references to a "United Country" compared to references to the regional identity of the former enemies, namely

[29] The *Patriotic* list includes the terms: American Revolution, Star Spangled Banner, Star and Stripes, US Flag, American people, American together, Americans, Fraternity, Patriotic, Reconciliation, United Country, United States. The *Divisive* list includes: Bloody Shirt, Carpetbagger, Civil War, Confederacy, Confederate, Confederate Flag, Northerner, Scalawag, Secession, Secession Flag, Sectionalism, Southerner.

[30] Since information on the total number of printed words (across all newspapers in a county) is unavailable, we proxy the total number of pages available for each county during each month by taking the maximum value between: i) the total number of pages as computed by newspapers.com and ii) total number of pages containing at least one of the terms "he", "you" or "I".

[31] Certain counties display some digitized pages for a few months and then disappear for most of our sample. For our baseline exercise, we thus focus on a subset of counties with a sufficient coverage of historical newspapers digitized on *newspapers.com* from between 1910 and 1920. We define the concept of sufficient coverage as all counties that have at least one newspaper page digitized for 25% of the 132 months in the sample. We present robustness exercises including all counties irrespective of coverage and focusing on a balanced sample of counties covered throughout the 132 months.

"Northerner(s)" or "Southerner(s)."

We begin by examining the *relative* frequency of these references. In the first three columns of Table 1, the dependent variable takes the form of $\texttt{logFreq}_{i,c,t} - \texttt{logFreq}_{j,c,t}$, namely the (log) relative frequency of the *patriotic* word $i$ (United Country) with respect to a *divisive* keyword $j$ (Northerner *or* Southerner). Indeed, we are primarily interested in documenting how the movie affected rhetoric and the salience of patriotic words relative to divisive words. We then explore the effect on the *absolute* frequency of the terms with $\texttt{logFreq}_{i,c,t}$ (col.4-6) and $\texttt{logFreq}_{j,c,t}$ (col.7-9) as dependent variables.

Columns (1), (4) and (7) display the estimation results of the OLS empirical model using Equation (1), where the main explanatory variable is the (non-instrumented) treatment $\texttt{BoaN}_{ct}$. In Columns (2), (5) and (8), we estimate the reduced form version by replacing the treatment with its instrument $\texttt{MDM}_{ct}$. Columns (3), (6) and (9) present the 2SLS results with $\texttt{MDM}_{ct}$ as instrument. The first stage estimates, reported in Appendix Section B.1, confirm that $\texttt{MDM}_{ct}$ is a strong and relevant predictor of $\texttt{BoaN}_{ct}$ with a Kleibergen-Paap statistic of around 166. Switching $\texttt{MDM}_{ct}$ from 0 to 1 increases the probability of BoaN being screened by about one third. For each outcome variable, we observe a substantial difference between the OLS and 2SLS estimates. While the various reasons explaining this pattern were discussed in Section 3, it is worth rementioning the importance of the fact that both the endogenous variable and the instrument are dichotomous variables. Reassuringly, the coefficient expansion in the rest of our analysis is always smaller than that we observe with these three specific keywords.

In the first three columns, we see that the treatment effect is positive and statistically significant at the 1% threshold. In terms of magnitude, using the 2SLS estimate as a benchmark, exposure to BoaN brings about a 40% increase in the relative frequency of the keyword "United Country" with respect to "Northerner/Southerner." We interpret this finding as evidence that the screening of BoaN brought about a shift in local papers' rhetoric away from regionalist language and toward a relative increase in references to a united country. When looking at the effect of the movie on unification and regionalism separately, we see that the treatment effect is statistically significant in both cases with a reversal of the coefficient from a positive to a negative sign. Hence, beyond its relative effect, the movie increased (decreased) the salience of patriotic (divisive) words in absolute terms: the effect amounts to a 6.8% increase in references to a United Country and a 31% decrease in references to the identities of the former enemies.

### 4.3 Reconciliation Rhetoric

In the baseline analysis, we opt for a more compact approach and reduce the dimensionality of the dataset using a principal component analysis (PCA). More precisely, we compress information on the log frequencies of the 24 keywords into two scalars, $\texttt{Patriotic}_{c,t}$ and $\texttt{Divisive}_{c,t}$, which correspond to the first principal component of the sets of patriotic and divisive keyword log frequencies, respectively.[32] The two scalars have comparable scales because the PCA is conducted after stan-

[32]For the bag of patriotic keywords, the first eigenvector explains 81% of the variance, with a corresponding eigenvalue of 9.68. In the first principal component, all words from the list have positive weights and the overall score for the Kaiser-

Table 1: North/South vs United Country - Newspaper Analysis

| | United Country - North/South | | | United Country | | | North/South | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.079*** | | 0.399*** | 0.012*** | | 0.068*** | -0.067*** | | -0.331*** |
| | (0.012) | | (0.046) | (0.003) | | (0.01) | (0.012) | | (0.043) |
| Million Dollar Mystery | | 0.127*** | | | 0.022*** | | | -0.105*** | |
| | | (0.013) | | | (0.003) | | | (0.013) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.561 | 0.562 | - | 0.185 | 0.185 | - | 0.652 | 0.652 | - |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Dep. Var. Mean | -.444 | -.444 | -.444 | .052 | .052 | .052 | .497 | .497 | .497 |
| Dep. Var. Std. Dev. | .722 | .722 | .722 | .209 | .209 | .209 | .758 | .758 | .758 |

NOTE: The table reports OLS (Columns 1, 4 and 7), reduced form (Columns 2, 5 and 8), and 2SLS (Columns 3, 6 and 9) estimates. The dependent variable is the log frequency of "United Country" minus the log frequency of "Northerner/Southerner" ($\texttt{logFreq}_{i,c,t} - \texttt{logFreq}_{j,c,t}$) in Columns (1) to (3), the log frequency of "United Country" ($\texttt{logFreq}_{i,c,t}$) in Columns (4) to (6), and the log frequency of "Northerner/Southerner" ($\texttt{logFreq}_{j,c,t}$), in Columns (7) to (9). See Section 4.2 for further details. The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

dardizing the data. Thus, we can compute our main dependent variable, $\texttt{Reconciliation}_{c,t}$, as the difference $\texttt{Patriotic}_{c,t} - \texttt{Divisive}_{c,t}$. This variable has a natural interpretation: It represents the relative log frequencies of patriotic and divisive keywords in local newspapers. Its log linear nature makes it additively separable, a convenient feature for analyzing the contribution of each of its components.

Equipped with these variables, we can now extend the previous analysis to the full set of 24 keywords in order to scrutinize the semantics of reconciliation in a multidimensional manner. To this end, we replicate the previous table with $\texttt{Reconciliation}_{c,t}$ as a dependent variable in Equation (1). Table 2 reports the results for OLS in Column (1), reduced form in Column (2) and 2SLS in Column (3). The coefficient of interest is consistently positive across specifications, showing that the screening of the movie generally tilted the content of local newspapers toward a more reconciliatory tone. The pattern found above with the three specific keywords thus still holds with this more comprehensive measure of reconciliation. However, given the PCA-induced rescaling of the dependent variable, assessing the magnitude of the effect must be performed in a different way than with plain log frequencies. The point estimate in Column (3) indicates that screenings led to an increase in $\texttt{Reconciliation}_{c,t}$ of 0.6 standard deviations.

In Columns (4) to (9), we again examine the impact of the movie on patriotic and divisive words separately. To this end, we break down the dependent variable into its two components $\texttt{Patriotic}_{c,t}$ and $\texttt{Divisive}_{c,t}$ and estimate the OLS, reduced form and 2SLS specifications for each of them. We continue to find evidence of antagonistic patterns, with a post-screening increase in the salience of patriotic words and a decrease in the salience of divisive words.

Meyer-Olkin measure of sampling adequacy is 0.96. For the bag of divisive keywords, the first eigenvector explains 96% of the variance, with a corresponding eigenvalue of 11.48. Again, all words in the first principal component receive positive weights and the overall score for the Kaiser-Meyer-Olkin measure of sampling adequacy is 0.98. As the share of variance explained by the first eigenvectors is high in both cases we decide to retain only the first principal components.

Table 2: Patriotic vs Divisive Discourse - Newspaper Analysis

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.158*** | | 0.599*** | 0.097*** | | 0.343*** | -0.061*** | | -0.256*** |
| | (0.0156) | | (0.058) | (0.019) | | (0.062) | (0.01) | | (0.035) |
| Million Dollar Mystery | | 0.190*** | | | 0.109*** | | | -0.081*** | |
| | | (0.016) | | | (0.019) | | | (0.01) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.801 | 0.802 | - | 0.933 | 0.933 | - | 0.981 | 0.981 | - |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). See Section 4.3 for further details. The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## 4.4 Sensitivity Analysis

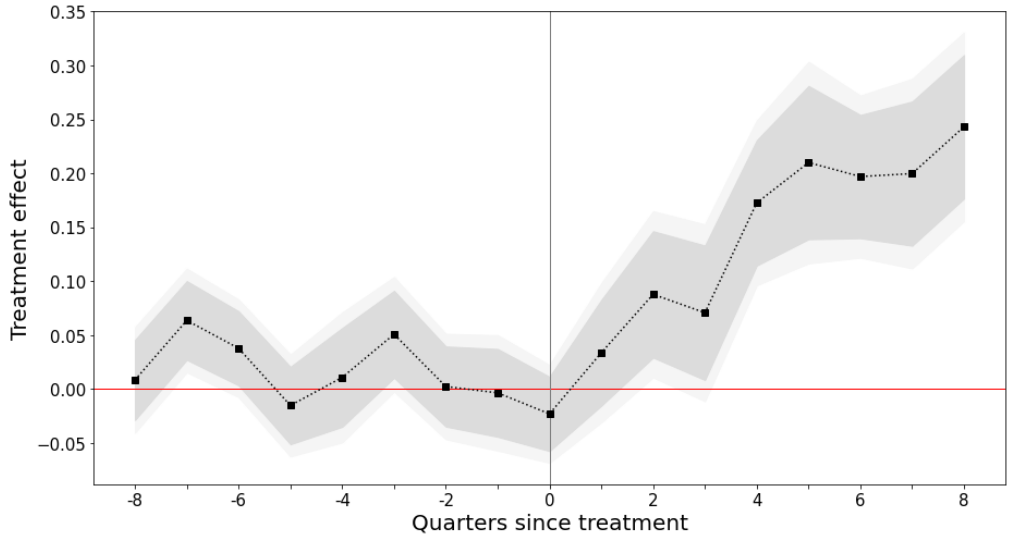
We perform a battery of sensitivity checks to test for the robustness of the baseline estimates from Table 2. In what follows, we report only a summary of the sensitivity analysis; all tables and a detailed discussion can be found in the Online Appendix.

**Measurement of Reconciliation Rhetoric.** A substantial part of the additional analysis reported in the Appendix is dedicated to alternative measurements of reconciliation in local newspapers. In Section B.2.1, we show that the results continue to hold when keyword occurrence is not measured as log frequency or when we propose different logarithmic transformations of the frequencies. More importantly, our starting list of *Patriotic* and *Divisive* terms was largely discretionary; keywords were selected based on their importance in the movie script and on the tenets of the Lost Cause narrative. Section B.2.2 demonstrates that our results are confirmed with any possible recombination of our baseline words, which actually provide very consistent and highly correlated measures of reconciliation and division. To prove this, we devise a flexible approach in the choice of *Patriotic* and *Divisive* words: For each list, we sample 100 different random subsets containing 50% of the words in the starting lists. Figure B1 shows that exposure to *The Birth of Nation* increases $\texttt{Reconciliation}_{c,t}$, no matter how the lists are constructed.

**Timing of the Effect.** Our premise is that the movie was a powerful vehicle of the Lost Cause narrative. Following the logic of Dawkins (1976) and Shiller (2019), we view narratives as memes spreading within a society in a viral manner. The Lost Cause narrative initially "infected" the county when the movie was aired. In the months following the screening, the narrative kept on spreading locally in the form of fiery debates and controversies within the echo chambers of the

time – churches, political party branches, charities, family gatherings, newspapers, etc. Hence, shifts in attitudes were likely gradual, cumulative and potentially not immediate. It takes time for a narrative to spread across segments of the local society and for echo chambers to deploy their full effect. To summarize, our interpretation of the post-screening changes in attitudes is that they capture not only the initial spark, but also the entire chain of reactions.

Figure 5: Patriotic vs Divisive Discourse - Lags and Leads




NOTE: The figure plots estimates of the effect on the variable $\texttt{Reconciliation}_{c,t}$ of 8 of lags and 8 leads (quarterly increments) of the variable $\texttt{BoaN}_{ct}$, estimated following De Chaisemartin and D'Haultfoeuille (2020).

A natural way to document the change of the effect over time is to complement our analysis by augmenting the two-way fixed effect estimator setup with richer dynamics of lags and leads. Therefore, we estimate a variant of equation (1) that embeds a set of quarterly leads and lags of the treatment effect, relying on the routine devised by Chaisemartin and D'Haultfoeuille (2020).[33] The estimated coefficients are reported in Figure 5. We find a gradual and cumulative impact of the screening of the movie on our reconciliation measure, in line with our expectations. Moreoever, data do not seem to reject the parallel trend assumption as pre-treatment effects appear to be absent.

**Measurement of *The Birth of a Nation* and Alternative Samples.** We perform various exercises to confirm that our results do not rely on any specific coding choice in the construction of the estimation sample and the explanatory variable. First, given that our variables are constructed based on information in newspapers, we need to verify that the findings are not driven by anomalous variations in newspaper coverage. Appendix Section B.3.1 accounts for variations in newspaper total coverage across counties and months in a variety of ways. In Appendix Section B.3.2, we consider alternative samples, including a balanced panel of counties with constant newspaper coverage throughout our sample period. The outbreak of the First World War might have fostered attitudes of reconciliation in the country. To rule out the confounding effect of the war in particu-

[33] We present results of specifications including 8 of lags and 8 of leads, though variations of the lags/leads structure would leave the results unaffected.

lar areas, we replicate our baseline results in Appendix Section B.3.2 but exclude all months after March 1917 and then all months December 1916. Finally, the definition of exposure to the movie was based on evidence of movie screenings and attempted to minimize misclassification errors by imposing a threshold of at least two movie screening records for a county to be "treated." In Section B.3.3, we show that the results do not rely on this choice and are robust to different definitions of the treatment variable.

**Alternative Instrumental Variables.** *The Million Dollar Mystery* (MDM) was released about eight months before *The Birth of a Nation* (BoaN). As demonstrated above, the transposed spatio-temporal diffusion of MDM is a good predictor for the diffusion of BoaN. In principle, it is possible that certain persistent shocks such as an income boom across certain dynamic cities, might drive both the screening of MDM and the screening of BoaN eight months later. To rule out this possibility, we replicate our baseline results using the spatio-temporal diffusion of two different movies as instruments. These films, *Traffic in Souls* and *What Happened to Mary*, are thematically unrelated to reconciliation or alienation and were released 2-3 years before *The Birth of a Nation*. In Section B.4 we present baseline results with these alternative instruments. Both movies do indeed follow recurrent spatio-temporal patterns, confirming our initial findings.

**Alternative Empirical Specifications.** Our next step is to investigate alternative empirical models (Appendix Section B.5). To control for trends that may be particular to treated counties, we account for county-specific linear time trends (Section B.5.1). Section B.5.2 presents results of specifications that include time-varying effects of variables that have been shown to differ between counties that were exposed to the movie and those that were not. In order to tackle time-varying confounders at the year level, we replicate our estimates accounting for state-specific year fixed effects (Section B.5.3) and show that the results remain unchanged.

Our two-way fixed effects (FE) regression baseline estimates can be conceived as a weighted sum of several difference-in-differences into a single one. To understand why, we can for instance take the example of counties that are exposed to BOAN first: these counties are compared with both counties that will never be treated and to counties that will be exposed to BOAN in the following years. As a way to understand whether the effect is only driven by the comparison with counties that are never exposed to BOAN, we replicate our baseline analysis focusing only on the subset of counties that screen the movie by the end of 1920. Counties that screen BoaN early are then compared, in the period immediately following the screening, with counties that would go on to screen BoaN later. Appendix Table B20 shows that the baseline patterns are confirmed.[34] Last, we allow for a richer clustering structure by taking into account the spatial nature of the error terms $\epsilon_{ct}$ in Equation (3.1) (Collela et al. 2020). These results are presented in Appendix Section B.5.5.

[34]The distance between OLS and IV estimates increases, however, because the first stage has much less predictive power. The instrument turns out to be better at predicting *where* the movie is going to be screened than *when* the movie is going to be screened.

**Word-based Regressions.** In Section B.5.6, we move to a word-based approach. Specifically, instead of looking at the frequency of vectors of words in a county-month, we shift to each single word $i$ in a county-month as the unit of observation. In these specifications, the dependent variable will be $\texttt{Hit}_{i,c,t}$, which measures the occurrence of each word in a county-month. The setting allows us to control for county-word fixed effects, on top of county-month fixed effects, which accounts for all events affecting counties at a specific point in time, such as local elections, income shocks, protests or strikes. Our results show that words belonging to the patriotic (divisive) list appears more (less) often after a county is exposed to the movie *The Birth of a Nation*.

# 5 Reconciliation & Patriotism: Evidence from War Casualties

In this section, we examine how the diffusion of the Lost Cause narrative spurred a new national sentiment and a renewed commitment to the national cause. We measure patriotism by looking at individuals' decision to volunteer for the United States Navy. This analysis complements the previous examination of changes in opinion by looking at how reconciliation translated into real behavioral changes.

## 5.1 Measurement of Patriotism through Navy Enlistment

Existing datasets on army volunteering (see, for instance, Fouka, 2019; and Caprettini, Schmidt-Fischbach and Voth, 2020) provide information on the geographical origin of volunteers but not on the date of enlistment. This limitation in the data is problematic given that our empirical strategy relies on within-county time-series variations in movie exposure. We must therefore construct a new dataset with information on the geographical and temporal dimensions of enlistment. We make use of data on casualties (deaths) suffered by the US Navy in the First World War – an *exhaustive*, and so far unexploited, dataset collected by the Bureau of Navigation of the Navy Department (1920).[35] Casualties involved individuals who enrolled in the Navy in the years surrounding the US's April 1917 entry into the war, between January 1913 and November 1918.[36]

The dataset contains 7,569 casualties, the vast majority of which occurred in 1917 and 1918 (851 and 5,847, respectively). For each deceased Navy sailor, we note his enlistment date $t$ and county of origin $c$ by assigning him the county of his next of kin's address (parents, spouse, etc.). Following the logic of the previous section (see the discussion below equation 2), we build our new outcome variable as the log share of enlistment at the county-month level:

$$\texttt{logList}_{c,t} \equiv \log\left[\frac{1+\texttt{sailors}_{c,t}}{1+\texttt{Pop}_c}\right] \tag{3}$$

where $\texttt{sailors}_{c,t}$ is the number of deceased sailors from $c$ who enlisted at time $t$ and $\texttt{Pop}_c$ is the pre-sample population of the county in 1910. Strictly speaking, this variable corresponds to the

[35] Data were digitized by G. Smith and are available at https://www.naval-history.net/WW1NavyUS-Casualties.htm.

[36] Data on Army enlistment could have been another valid potential source of information. Unfortunately, materials on the Army contain only the list of soldiers who lost their lives in WWI, lackingany information on enlistment dates (the original data was lost in a fire in 1973 (source)).

enlistment of sailors who died during the war. However, when estimating Equation (1), we interpret it as a proxy for enlistment in the Navy at large. This view is reasonable given the inclusion of time and county fixed effects in our empirical design and the log transformation of the dependent variable.[37] Given that effort in battle and the related mortality rate could be also related to national sentiment and patriotism, as a robustness analysis, we replicate our approach including only sailors who perished because of infectious diseases.

Figure A15 maps the spatial distribution of the enlistment measure $\texttt{logList}_{c,t}$ across the country in our estimation sample. We see that variations are substantial and not confined to the coastal areas, two desirable features for identifying the treatment effect.[38] The main limitation of our data is that there is no distinction between individuals who volunteered and those who were drafted. This is of potential concern only for the period following the introduction of the draft with the Selective Service Act of May 1917. However, draft evasion was a widespread phenomenon, amounting to more than 15 percent of all eligible men – a higher share than during the Vietnam War – and public authorities were never able to strictly enforce the draft (Kazin 2017). It is therefore likely that the local enrollment rate is a valid proxy of patriotism even after April 1917. Hence, for the sake of completeness, we consider two versions of the estimation sample: one covers 1913-1918 (the full period for which enrollment data are available), while the other excludes all months after April 1917.

### 5.2 Reconciliation and Patriotism: Empirical Results

Table 3 reports the estimation results of Equation (1) with log enrollment share ($\texttt{logList}_{c,t}$) as the dependent variable. Specifications replicate the structure of Columns (1) to (3) in Table 2, with OLS in Column (1), reduced form in Column (2) and 2SLS in Column (3). In Columns (4) to (6), we redo the analysis for the subsample of observations preceding April 1917, which only involve individuals that enlisted on a purely voluntary basis. Reassuringly, the coefficients of OLS, reduced form and 2SLS estimates are consistently positive and very precisely estimated. Using 2SLS estimates in Column (3) as a benchmark, the coefficient size indicates that exposure to *The Birth of a Nation* increased the number of volunteers by 7%. A more conservative estimation in column (6) yields a comparable but smaller effect (around 3%).

**Sensitivity Analysis.** We use the following checks to verify the robustness of our findings, available in the Online Appendix. In Appendix Section C.1.1 we show that results are robust to using

[37] Our approach basically rests on the empirical assumption that the likelihood of dying in service at a given point in time, conditional on an individual's military experience (i.e., date of enlistment), is not influenced by county-specific time-varying factors. Technically, let us assume that the mortality risk has two multiplicatively separable components, $\mu_c$ and $\nu_t$. Then, abstracting from the +1 transformation, we get $\texttt{logList}_{c,t} = \log \mu_c + \log \nu_t + \texttt{logList}^{\texttt{tot}}_{c,t}$ where $\texttt{logList}^{\texttt{tot}}$ is the log share of total enlistment. When estimating Equation (1), the terms $\log \mu_c$ and $\log \nu_t$ are absorbed by the fixed-effect structure. As a consequence, having logList as a dependent variable yields the same coefficient of interest $\beta$ as the ideal specification with the unobserved variable $\texttt{logList}^{\texttt{tot}}$. Finally, note that about 1.4% of all the individuals recruited died during the sample period. The total number of US Navy personnel during WWI, according to figures from the New York State Archives (see http://www.archives.nysed.gov/education/total-navy-personnel-state-world-war-i-c-1920), was 551,736.

[38] Appendix Figure A14 shows the variation in total enlistment over time.

the same sample of counties as that in Table 2. Appendix Section C.1.2 considers alternative definitions of the treatment variable, while Section C.2.1 focuses on alternative log transformations of the outcome variable. We also look at the extensive margin of enrollment (instead of the intensive margin), considering an alternative model in which we use an indicator variable for at least one enlistment in the county-month as an outcome variable. Additionally, Appendix Section C.2.2 looks at the subsample of enlisted men who died of infectious diseases to distinguish between combat and non-combat fatalities. For the Navy, the latter subsample is particularly relevant given of the impact of the Spanish flu.

Table 3: *The Birth of a Nation* and Enlistments

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.031*** | | 0.067*** | 0.012*** | | 0.026*** |
| | (0.004) | | (0.014) | (0.003) | | (0.008) |
| Million Dollar Mystery | | 0.024*** | | | 0.007*** | |
| | | (0.005) | | | (0.002) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.978 | 0.978 | - | 0.995 | 0.995 | - |
| 1st stage F-Stat | - | - | 299 | - | - | 251 |
| Dep. Var. Mean | -10.09 | -10.09 | -10.09 | -10.11 | -10.11 | -10.11 |
| Dep. Var. Std. Dev. | .96 | .96 | .96 | .98 | .98 | .98 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$. See Section 5.1 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## 6 Reconciliation & Cultural Convergence in Naming Patterns

A key factor in successful nation-building is a sense of a common identity with shared symbols and cultural elements (see Anderson, 1983, among others). In this section, we study this specific dimension of post-conflict reconciliation by showing how the Lost Cause narrative contributed to cultural convergence between former enemies. We focus on the decision by parents to give their child a first name associated with the regional culture of the former enemy. More precisely, we test whether the movie increased the popularity of "Unionist-sounding" names among babies born in former Confederate states and *vice-versa*.

We view naming patterns as an appropriate object of inquiry because first names are widely considered important markers of cultural identity. - Moreover, the choice of a first name is avail-

able to all parents, without material constraints (Lieberson, 1992). A large body of literature in economics has looked at naming patterns to measure parents' racial, social, cultural and even political attitudes (e.g., Fryer and Levitt, 2004; Fouka, 2019; Abramitzky et al., 2020; Bazzi, Fizsbein and Gebresilasse, 2020; and Algan, Malgouyres, Mayer and Thoenig, 2021).[39]

### 6.1 Empirical Design and the Enemy-Sounding Name Index

We use first names to proxy the prevailing sentiments of the former Civil War enemies towards each other over time. We identify the former enemies as two groups, $S$ and $N$, which correspond to the former Confederate and Unionist states.[40] Drawing on Fryer and Levitt (2004), and more recently Fouka (2019, 2020) and Abramitzky et al. (2020), we compute an index of name distinctiveness that measures how popular a given name is within the former enemy's population relative to the population at large. More precisely, for a given baby $i$, born in group $g(i) \in \{S, N\}$ between 1910 and 1920, we build an index of how "enemy-sounding" her name $n(i)$ is, as follows:

$$\texttt{ENI}_i = \frac{\texttt{Name}_{n(i),g_{-1}(i)}}{\texttt{Name}_{n(i),g_{-1}(i)} + \texttt{Name}_{n(i),g(i)}} \times 100 \tag{4}$$

where Name represents the name frequencies taken separately for the child's own group $g(i)$ and the former enemy's group $g_{-1}(i)$. These (pre-sample) frequencies are computed for US-born white individuals born before 1910 to US-born parents in territory $g$ or $g_{-1}$ (thus limiting the cultural influence of recent migrants).[41] We construct $\texttt{ENI}_i$ for the subsample of individuals born between 1910 and 1920 and covered in the 1% sample of the Integrated Public Use Microdata Series (IPUMS) of the 1920 and 1930 Censuses. Given that censuses report only the county of residence, we must assume that counties of birth and residence are identical.

The Enemy-Sounding Name Index, ENI, gauges how much a given name evokes the former enemy's regional identity. Names with a value of zero are only used by people born in the region of origin $g(i)$ of the individual $i$. A value of 100, in contrast, means that the name is encountered only in the territory of the former enemy $g_{-1}(i)$.[42] Below, we also consider a binary version of the

[39] Fryer and Levitt (2004) provide additional evidence on the cultural component of first names by showing that the surge in distinctively Black names in the US since the seventies can be associated to a rise in Black cultural identity. In their study of two major waves of immigration in the United States, Abramitzky et al. (2020) emphasize the attractiveness of first names as a measure of assimilation. They argue that first names are more likely to reflect preferences and less likely to reflect constraints imposed by the host society than alternative measures, such as intermarriage, which could reflect both the demand and supply determinants of assimilation opportunities. Mazumder (2019) finds that immigrants' military service in the US army during World War I increased their rate of cultural assimilation, with potentially positive economic returns. Particularly relevant to our study is the finding by Fouka (2019) that German immigrants and their descendants responded to discrimination in the US during WWI by increasing their assimilation efforts, including by changing the "Americanness" of their names.

[40] We choose to define Unionist states as all states and former US territories that did not belong to the Confederacy. Note that regressions with alternative definitions of this group (provided in the online appendix) show that the results do not rely on this choice. The list of former Confederate States comprises Alabama, Arizona, Arkansas, Florida, Georgia, Louisiana, Mississippi, New Mexico, North Carolina, South Carolina, Tennessee, Texas, Virginia and West Virginia.

[41] Frequencies are retrieved from the 1880 10% and 1900 5% Integrated Public Use Microdata Series (IPUMS) samples, publicly available datasets based on census data.

[42] Anecdotally, the relevance of this index is also illustrated by the choice of character names in the movie *The Birth of a Nation*. The two main characters are named *Austin* Stoneman (the abolitionist leader from the North) and *Ben* Cameron

ENI which takes a value of 1 when $\texttt{ENI}_i > 50$ and zero otherwise. In this case, the binary index has a straightforward interpretation as it indicates names that are *more* popular in the former enemy's territory than in the region of origin.

We then estimate the impact of the screening of *The Birth of a Nation* on the ENI of babies born over the 1910-1920 period. To this end, the baseline econometric equation (1) has to be slightly modified in order to accommodate individual-level data:

$$\texttt{ENI}_i \equiv \alpha + \beta \times \texttt{BoaN}_{c(i),y(i)} + \sum_c \texttt{FE}_c + \sum_g \texttt{FE}_g + \sum_{y,s} \texttt{FE}_{y,s} + \epsilon_i. \tag{5}$$

where the unit of observation is the baby $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The treatment variable has to be adjusted accordingly and now varies at the county-year level: $\texttt{BoaN}_{cy}$ codes for the post-screening period and is equal to 1 in all years following the screening and 0 otherwise. The variables $\texttt{FE}_c$, $\texttt{FE}_g$ and $\texttt{FE}_{ys}$ respectively stand for county, gender and (year of birth $\times$ census year) fixed effects. Our analysis departs from that presented in the previous section in two main respects. Firstly, the outcome of interest varies at the individual level rather than the county level. Secondly, the time dimension is now collapsed to the year level given that censuses do not report individuals' month of birth.

In Equation (5), a positive $\beta$ captures the extent to which a baby born in a Unionist county has a Confederate-sounding name or vice versa. We are tempted to interpret a convergence in naming patterns as an indication of fading stigmas attached to names that were historically distinctive of the former enemy's culture. However, rather than a change in cultural norms, convergence in naming patterns could also be driven by migration. This would be the case if the post-screening period is systematically associated with an increase in the inflow of migrants from the states of the former enemy (who bring with them a distinctive set of names for their babies). While interesting and also related to reconciliation, this alternative channel seems, in our view, less plausible. Below, we perform several empirical exercices to control for this migration-related channel. In particular, we exclude first- and second-generation migrants by restricting our estimation sample to babies whose state of birth and/or parental state of birth correspond to their state of residence at the time of the census.

### 6.2 Convergence in Naming Patterns: Results

The baseline estimation results of Equation (5) are displayed in Table 4. Following the logic of the previous tables, the first three specifications correspond to OLS (Column 1), reduced form (Column 2) and 2SLS (Column 3). The last three columns replicate the same set of regressions with the binarized version of the ENI. In these cases, the empirical model has to be interpreted as a linear probability model.

In all columns, the coefficient of interest is positive and statistically significant at conventional thresholds. These results show that the screening of the movie influenced naming decisions of

(savior of the South and "proud" founder of the Ku Klux Klan). The corresponding ENI values for the two names ( $\texttt{ENI}_{Austin}$=40.27 and $\texttt{ENI}_{Ben}$=25.95) seem to reflect a desire to further typify the two movie characters.

babies by increasing the prevalence of enemy-sounding names at the intensive margin (Columns 1 to 3) as well as the extensive margin (Columns 4 to 6). Point estimates in columns (4) and (6) of Table 4 imply that airing the movie increased the likelihood that a baby born in the county would receive a popular enemy-sounding name by 1.8 to 4.7 percentage points (sample mean: 36 pp). As discussed above, this evidence of a convergence in naming patterns is consistent with our hypothesis that the large-scale diffusion of the Lost Cause narrative fostered cultural reconciliation between the groups.

Table 4: *The Birth of a Nation* and Naming Patterns

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.687*** | | 1.635*** | 0.018*** | | 0.047*** |
| | (0.222) | | (0.461) | (0.006) | | (0.012) |
| Million Dollar Mystery | | 0.854*** | | | 0.024*** | |
| | | (0.240) | | | (0.007) | |
| Observations | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 |
| R-squared | 0.070 | 0.070 | - | 0.064 | 0.064 | - |
| 1st stage F-Stat | - | - | 144 | - | - | 144 |
| Dep. Var. Mean | 44.07 | 44.07 | 44.07 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.82 | 16.82 | 16.82 | 16.82 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

**Sensitivity Analysis.** In the Online Appendix, we explore the robustness of these findings to alternative measures of name connotations and group definitions (Section D.1.1), alternative county coverage (Section D.1.2) as well as alternative definitions of exposure to the movie (Section D.1.3), alternative empirical specifications (Section D.2), and alternative definitions of former enemies (Section D.3).

We also explore whether migration patterns or different family composition could explain our results (Section D.4). We do this by first replicating our baseline analysis using only individuals from the 1920 Census, whose year of birth is closer to the census year, which presumably decreases the chance that they resided in a county other than their county of birth (Table D10). We then replicate our main results using only individuals from former Confederate states for whom the head of the family was also born in the former Confederacy, and repeat this analysis for the former

Union (Table D9). Finally, we verify the robustness of our estimates when including a set of family characteristics FEs (Table D8). Overall, the results remain very similar and consistent with our baseline estimates presented in Table 4.

Finally, we present a set of falsification exercises to validate our empirical approach. First, we replicate our analysis using Black newborns, expecting to find no effect within this subsample of the population. Second, in order to verify that our explanatory variable of interest does not capture pre-existing trends in naming patterns that affect counties where the movies are more likely to be screened, we fictitiously anticipate the release of the movies *The Birth of a Nation* and *The Million Dollar Mystery*. The corresponding results suggest that the screening of *The Birth of a Nation* did not produce a significant change in naming patterns of Black individuals from counties in the former Confederacy or Union (Section D.5.1), nor did it affect naming patterns when we fictitiously anticipated the release of the movies (Section D.5.2).

# 7 From Alienation to Reconciliation

The message of *The Birth of a Nation* was a call for national reconciliation between former enemies. The racist portrayal of African Americans played an instrumental role in accomplishing this goal, with the emancipation of Blacks depicted as a threat for the entire country. Whites from the North and South were exhorted to reunite on the basis of racial solidarity. In this section, we explore whether this logic of a common enemy is at work in the data: Did the movie promote reconciliation by promoting the alienation of African Americans? We start by documenting how the movie increased racial alienation by looking at references to race in the public discourse and discrimination on the labor market. We then perform a mediation analysis to assess the extent to which movie-induced racial alienation contributed to reconciliation.

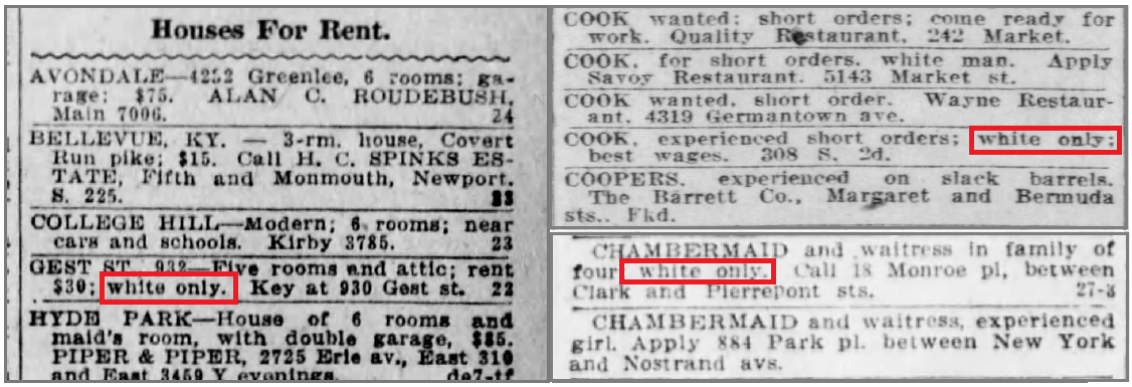
## 7.1 Movie-Induced Alienation

We use our data on local newspapers to build three measures of racial alienation at the county level. Our first outcome variable, $\texttt{Supremacism}_{c,t}$, is a proxy for racial nationalism that captures the salience of race and whiteness in the public discourse on national identity. In our baseline exercise, $\texttt{Supremacism}_{c,t}$ is computed using Equation (2) to find the log frequency of newspaper pages containing the keyword "White Americans." In robustness exercises, we look at references to "True Americans" as well as explore more extreme forms of supremacism with the keyword "Aryan(s)." Our second outcome variable, $\texttt{Discrimination}_{c,t}$, is a proxy for racial discrimination based on ads retrieved in local newspapers. Two examples of such discriminatory job ads are displayed in Figure 6 (both retrieved from newspapers located in a former Unionist state). As a baseline measure, we compute the log frequency of newspaper pages containing ads with the common formulation "White Only." Such wording generally pertains to housing, labor market and other forms of racial discrimination, though a small fraction of these pages are unrelated to such discrimination, leading to a certain amount of measurement error. In the Online Appendix, we validate our results with two other keywords that are less frequent but more unambiguously related to discrimination:

the log frequency of the keywords "White Men Only" and "White Only Wanted." Finally, our third variable, $\texttt{Alienation}_{c,t}$, corresponds to the first principal component of the previous two measures.

Table 6 reports OLS, reduced form, and 2SLS estimation results of Equation (1) with these three outcome variables as dependent variables. The coefficients are positive and statistically significant in all columns. Exposure to BoaN increases references to "White Americans" by 10% and racial discrimination in job ads by 21%. These results are confirmed using alternative measurements of supremacism (see Appendix Section E.1.2), alternative measures of racial discrimination (Section E.1.1) and alternative measures of exposure to *The Birth of a Nation* (Section E.1.3).

Figure 6: Racial Discrimination in the Housing Market and the Labor Market




NOTE: Three advertisements documenting racial discrimination in the housing and labor markets. *Left Figure The Cincinnati Enquirer*, Ohio. *Top Right Figure The Philadelphia Enquirer*, Pennsylvania. *Bottom Right Figure The Brooklyn Daily Eagle*, New York. Source: newspapers.com.

Table 5: *The Birth of a Nation* and Alienation of African Americans

| | Supremacism | | | Discrimination | | | Alienation | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.028*** | | 0.097*** | 0.056*** | | 0.212*** | 0.028*** | | 0.104*** |
| | (0.005) | | (0.018) | (0.012) | | (0.043) | (0.005) | | (0.017) |
| Million Dollar Mystery | | 0.031*** | | | 0.067*** | | | 0.033*** | |
| | | (0.006) | | | (0.013) | | | (0.005) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.951 | 0.951 | - | 0.808 | 0.808 | - | 0.934 | 0.934 | - |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Dep. Var. Mean | -4.576 | -4.576 | -4.576 | -4.247 | -4.247 | -4.247 | -.76 | -.76 | -.76 |
| Dep. Var. Std. Dev. | 1.178 | 1.178 | 1.178 | .971 | .971 | .971 | .671 | .671 | .671 |

NOTE: This table reports OLS (Columns 1, 4 and 7), reduced form (Columns 2, 5 and 8) and 2SLS (Columns 3, 6 and 9) estimates. The dependent variable is the salience of whiteness in the public discourse ($\texttt{Supremacism}_{c,t}$) in Columns (1) to (3), racial discrimination ($\texttt{Discrimination}_{c,t}$) in Columns (4) to (6) and the first principal component of $\texttt{Supremacism}_{c,t}$ and $\texttt{Discrimination}_{c,t}$, in Columns (7) to (9). See Section 7 for further details. The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### 7.2 Alienation as an Instrument of Reconciliation

With the aim of documenting the extent to which the racist narrative of the movie boosted reconciliation, we test whether the film's impact was stronger in areas where the fears triggered by its content struck a nerve. The analysis that follows thus focuses on our preferred outcome variable, $\texttt{Reconciliation}_{c,t}$, namely the measure of reconciliatory rhetoric in local newspapers. Specifically, we seek to decompose the total treatment effect in the baseline Equation (1) into a direct effect of the movie and an indirect effect mediated by alienation.

One approach could consist of performing a naïve mediation analysis where we estimate the baseline equation with $\texttt{Reconciliation}_{c,t}$ on the LHS, controlling for $\texttt{Alienation}_{c,t}$ on the RHS. An attenuation of the treatment effect with respect to its baseline counterpart together with a non-zero coefficient of the control would provide insight on the mediating effect of alienation. Though simple, such an empirical design suffers from well-known estimation biases (the so-called "bad control problem") and hence we refrain from using it.[43] Note also that we cannot apply the recent procedure developed by Dippel, Ferrara and Heblich (2020) on causal mediation analysis in instrumental-variables regressions. Indeed, in our setup, we cannot exclude the possibility of endogeneity arising from confounders that jointly influence the treatment (i.e., movie screening) and the outcome of interest (i.e., reconciliation) – a key requirement for the latter methodology to be valid.

Our empirical strategy therefore follows a different route and exploits exogenous shifters of the movie's effect on alienation that are not expected to modulate its direct effect on reconciliation. Specifically, we retrieve the pre-sample spatial distribution of African Americans within the US population from the 1910 Census and build $\texttt{Black}_c$, an indicator variable that codes for counties with a share of African Americans above the 33rd percentile. Our strategy relies on two assumptions. The first (testable) hypothesis $[H_1]$ is that the movie exerted a larger effect on alienation in counties with a large Black population ($\texttt{Black}_c = 1$) relative to other counties ($\texttt{Black}_c = 0$). The second identifying assumption, $[H_2]$, is that *for a given level of alienation, the direct effect of the movie on reconciliation is identical for both types of counties*. This statement is not testable and is akin to the exclusion restriction in a 2SLS setup. Both $[H_1]$ and $[H_2]$ seem empirically relevant. Indeed, the movie depicts African Americans as dangerous criminals and their political empowerment as a threat to white Americans. It is expected that this disturbingly racist and anxiety-provoking message resonated more among white Americans in areas with a large Black population. By contrast, there is no reason to expect that the movie's message of reconciliation had a larger impact on attitudes in such areas, with a possible exception related to slavery and war destruction (see below).

More precisely, we estimate Equation (1) augmented with the interaction term $\texttt{BoaN}_{ct} \times \texttt{Black}_c$ as an explanatory variable

$$\texttt{outcome}_{ct} = \alpha + \beta \times \texttt{BoaN}_{ct} + \gamma \times \texttt{BoaN}_{ct} \times \texttt{Black}_c + \sum_c \texttt{FE}_c + \sum_t \texttt{FE}_t + \epsilon_{ct}. \tag{6}$$

We denote $(\hat{\beta}_{\texttt{Ali}}, \hat{\gamma}_{\texttt{Ali}})$ and $(\hat{\beta}_{\texttt{Rec}}, \hat{\gamma}_{\texttt{Rec}})$ as the point estimates recovered from the estimation of this

[43] The estimation results, available upon request, are nonetheless consistent with a mediating effect of alienation.

equation with $\texttt{Alienation}_{c,t}$ and $\texttt{Reconciliation}_{c,t}$ as outcome variables, respectively. Equipped with this set of four coefficients, we first test $[H_1]$ by looking at the coefficient of the interaction term obtained with $\texttt{Alienation}_{c,t}$. The hypothesis is not rejected by the data if $\hat{\gamma}_{\texttt{Ali}} > 0$. We then test for the mediation effect using the interaction term attached to $\texttt{Reconciliation}_{c,t}$. A finding that $\hat{\gamma}_{\texttt{Rec}} > 0$ is evidence that the overall effect of the movie on reconciliation is larger in counties with large Black populations. Under $[H_2]$, this necessarily means that part of the effect is indirectly mediated by the increase in alienation. In Appendix Section E.2.1, we discuss the econometric reasoning in greater detail and explain how we can precisely quantify the contribution of mediation to the overall treatment effect using this set of coefficients.

Table 6 reports the 2SLS estimation results of Equation (6) with $\texttt{MDM}_{ct} \times \texttt{Black}_c$ as the instrumental variable of the interaction term (see Appendix E.2.2 for OLS and reduced form results). In Column (1), the outcome variable is $\texttt{Alienation}_{c,t}$. The coefficient of the interaction term is, as expected, positive and statistically significant at conventional thresholds, confirming that [H1] is empirically relevant and the effect of the movie on alienation is stronger in places with a larger share of African Americans. We then turn to $\texttt{Reconciliation}_{c,t}$ as the dependent variable in Column (2).[44] Here again, the interaction term is positive and statistically significant. As explained above, this is evidence that the movie's impact on reconciliation was partly mediated by the increase in racial alienation. Quantitatively, the coefficients of Columns (1) and (2) imply that 55% of the total treatment effect of the movie on reconciliation is mediated by racial alienation.

Our approach is valid as long as $[H_2]$ holds. The latter may not, however, apply in places that were heavily impacted by war destruction or those in which slavery was practiced before the Civil War. In such places, the share of African Americans in 1910 and the level of resentment and discontent related to the war are likely to be correlated. To address this concern, we repeat the analysis of the first two columns for counties in free states where slavery was already prohibited before the war (Columns 3 and 4) and for counties where per-capita war casualties were below the median (Columns 5 and 6).[45] Quite remarkably, in spite of the drastic sample size reductions, the coefficients are barely affected and the quantitative assessment of the contribution of alienation to the total treatment effect is very stable. Specifically, it is equal to 53% for the subsample of free states and 53% for counties with lower Civil War casualties.

[44] Note that we cannot apply our methodology to enlistment in the Navy because enlistment was also open to African Americans. Hence, the threat that hypothesis $[H_2]$ is violated cannot be ignored in the case of this variable. The Enemy Name Index, meanwhile, is an individual-level variable while Equation (6) is defined at the county level.

[45] Columns (3) and (4) display results obtained by removing states in which slavery was legal before the American Civil War. The list of free states includes: California, Colorado, Connecticut, Idaho, Illinois, Indiana, Iowa, Kansas, Massachusetts, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New York, North Dakota, Ohio, Oregon, Pennsylvania, Rhode Island, South Dakota, Utah, Vermont, Washington, Wisconsin and Wyoming. Columns (5) and (6) display results obtained using only states in which Civil War military casualties as a share of the 1860 population was below the median. The list of states with comparatively low Civil War casualties includes: Arizona, California, the District of Columbia, Florida, Georgia, Idaho, Louisiana, Maryland, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, Pennsylvania, Rhode Island, Tennessee, Texas, Utah, Washington, West Virginia and Wyoming.

Table 6: Alienation as a Means of Reconciliation

| | Full Sample | | Free States | | Low Int. Civil War | |
|---|---|---|---|---|---|---|
| | (1) Alien. | (2) Reconc. | (3) Alien. | (4) Reconc. | (5) Alien. | (6) Reconc. |
| Birth of a Nation | 0.078*** | 0.507*** | 0.092*** | 0.565*** | 0.040 | 0.329** |
| | (0.017) | (0.061) | (0.024) | (0.099) | (0.028) | (0.128) |
| Birth of a Nation x Black | 0.042*** | 0.129*** | 0.057*** | 0.156*** | 0.056** | 0.174*** |
| | (0.014) | (0.040) | (0.014) | (0.050) | (0.029) | (0.076) |
| Observations | 102,772 | 102,772 | 63,899 | 63,899 | 36,441 | 36,441 |
| Mediation Effect (% Total) | - | 55.2 | - | 53.2 | - | 53.8 |

NOTE: The table reports the 2SLS estimates. The dependent variable is $\texttt{Supremacism}_{c,t}$ in Columns (1), (5) and (7), and $\texttt{Reconciliation}_{c,t}$ in Columns (2), (6) and (8). See Section 7 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all baseline counties in Columns (1) and (2), only counties located within free states in Columns (3) and (4), and only counties with a lower than median share of Civil War casualties in Columns (5) and (6). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). The variable $\texttt{Black}_c$ is an indicator variable taking a value of 1 for counties with a population share of African Americans above the 33rd percentile. All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# 8 A Nation-wide Reconciliation

In this section, we show that the push for national reconciliation advocated by the movie, with its associated instrumental message of racial alienation, reached even areas of the country that had historically never practiced any outright forms of segregation. Importantly, we show that in these places, the message of racial alienation transmitted by the film took root in more subtle, but potentially equally disruptive, ways.

Specifically, we replicate our baseline analysis for each outcome variable, comparing the treatment effect of the movie in former Confederate and Unionist states.[46] Figure 7 reports the estimated effects of the movie recovered from 2SLS estimations (see online Appendix Section F.1 for OLS and reduced form results). To permit comparison, we standardize all dependent variables with a zero mean and unitary standard deviation. We do not report the results for $\texttt{Reconciliation}_{c,t}$ on the figure, since this would make the scale too compressed and visually unappealing. Coefficients in black correspond to the estimated effects of the movie for the full sample; those in dark and light gray report the effects when estimated separately for the former Unionist and Confederate states respectively (as per Equation (7)). For each coefficient, the scale reads as standard deviation of the

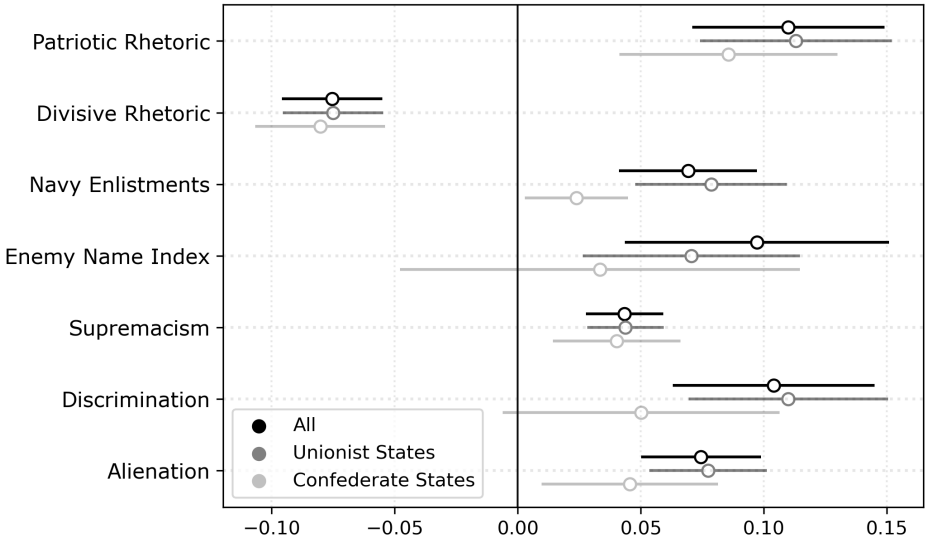
[46] We augment the baseline model (1) with group-specific treatment effects

$$\texttt{Outcome}_{ct} = \alpha + \beta \times \texttt{BoaN}_{ct} \times \texttt{Unionist}_c + \gamma \times \texttt{BoaN}_{ct} \times \texttt{Confederate}_c + \sum_c \texttt{FE}_c + \sum_t \texttt{FE}_t + \epsilon_{ct} \quad (7)$$

where the variables $\texttt{Unionist}_{ct}$ and $\texttt{Confederate}_{ct}$ indicate whether the county belongs to a former Unionist or Confederate state, respectively.

variable of interest.

Figure 7: A Nationwide Phenomenon




NOTE: The figure summarizes regression estimates of the effects of exposure to BoaN on our outcome variables, standardized to have a zero mean and unitary standard deviation. See the text for further details.

Results indicate that *The Birth of a Nation* fostered reconciliation in both former Confederate and Unionist states. Public debate moved toward more patriotic (approx. 0.12 SD) and less divisive rhetoric (approx. 0.08 SD) throughout the entire country. Changes in behaviors, meanwhile, measured in terms of Navy enlistment and naming choices, were slightly larger in the former Union states. Importantly, a vision of "White America" and racial segregation permeated with equal force former Confederate and Unionist states.

Our findings on naming choices resonate with the work of David Blight (2009) and other scholars, suggesting that while the North won the war on the battlefield, the South managed to win the war on memory (see, among others, Goldfield, 2013). The estimation results show, admittedly with some statistical imprecision, that the Enemy Name Index changed more dramatically in the former Unionist states. As names are a close proxy of cultural shifts, these findings are consistent with the view that the movie advanced the cause of the South in the "culture war," renewing an idealized imaginary of the plantation South.

## 9 Conclusion

This paper provides evidence on the powerful role of reconciliation narratives in the aftermath of a civil conflict by looking at the dissemination of the Lost Cause tenets through the movie *The Birth of a Nation* (1915). Our results suggest that reconciliation was fostered by substituting the North/South cleavage with a Black/White cleavage. Specifically, the Lost Cause narrative forged the myth of a common threat to all white Americans from the North and the South: Blacks and their fight for enfranchisement.

Several caveats apply to our endeavor to study the Lost Cause narrative. First, our empirical approach allows us to assess only effects that relate to the influence of the movie *The Birth of a Nation*. Yet the Lost Cause narrative was, in fact, popularized and disseminated by a host of other cultural channels, such as literary books and political campaigns. Its overall impact may therefore be larger than that captured herewith our empirical framework. Second, by focusing on exogenous variation in exposure to the narrative, our analysis cannot shed light on the fundamental drivers of the emergence of the narrative, which we leave to further research. Third, our analysis does not allow us to discern whether the observed changes in opinions and attitudes were driven by genuine persuasion or the reactivation of a silent majority.

Our findings offer a new interpretation of the role of the Lost Cause in molding internal cleavages that remain at the core of political debate today. More broadly, our paper questions the ability to reconcile as one nation when that reconciliation is based on the exclusion of some of its members. We leave to future research the study of factors leading to truly inclusive reconciliation.

## References


Abramitzky, R., L. Boustan, and K. Eriksson (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights 2*(1), 125–41.

Adena, M., R. Enokopolov, M. Petrova, and V. Santarosa (2015). Radio and the rise of the nazis in pre-war germany. *The Quarterly Journal of Economics 130*(4), 1885–1939.

Ager, P., L. P. Boustan, and K. Eriksson (2019). The intergenerational effects of a large wealth shock: White southerners after the civil war. Technical report, National Bureau of Economic Research.

Akerlof, G. A. and D. J. Snower (2016). Bread and bullets. *Journal of Economic Behavior & Organization 126*, 58–71.

Algan, Y., T. Mayer, and M. Thoenig (2013). The economic incentives of cultural transmission: Spatial evidence from naming patterns across france.

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso books.

Ang, D. (2020). The birth of a nation: Media and racial hate.

Autor, D., D. Dorn, G. Hanson, and K. Majlesi (2020). Importing political polarization? the electoral consequences of rising trade exposure. *American Economic Review 110*(10), 3139–83.

Bakshy, E., S. Messing, and L. A. Adamic (2015). Exposure to ideologically diverse news and opinion on facebook. *Science 348*(6239), 1130–1132.

Bauer, M., C. Blattman, J. Chytilová, J. Henrich, E. Miguel, and T. Mitts (2016). Can war foster cooperation? *Journal of Economic Perspectives 30*(3), 249–74.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2020). Frontier culture: The roots and persistence of "rugged individualism" in the united states. *Econometrica 88*(6), 2329–2368.

Bean, J. M. (2017). Early mystery-crime films, scientific seriality, and the imagination of wonder. *Velvet Light Trap* (79), 90.

Bénabou, R., A. Falk, and J. Tirole (2018). Narratives, imperatives, and moral reasoning. Technical report, National Bureau of Economic Research.

Black, D. A., M. C. Berger, and F. A. Scott (2000). Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association 95*(451), 739–748.

Blight, D. (2009). *Race and reunion*. Harvard University Press.

Blouin, A. and S. W. Mukand (2019). Erasing ethnicity? propaganda, nation building, and identity in rwanda. *Journal of Political Economy 127*(3), 1008–1062.

Boustan, L. P. (2010). Was postwar suburbanization "white flight"? evidence from the black migration. *The Quarterly Journal of Economics 125*(1), 417–443.

Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. *Science 314*(5805), 1569–1572.

Boyd, R. and P. J. Richerson (2005). *The origin and evolution of cultures*. Oxford University Press.

Buck, P. H. (1937). *The Road to Reunion, 1865-1900*, Volume 72. Little Brown & Company.

Bursztyn, L. and D. Cantoni (2016). A tear in the iron curtain: The impact of western television on consumption behavior. *Review of Economics and Statistics 98*(1), 25–41.

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

Caprettini, B., F. Schmidt-Fischbach, and H.-J. Voth (2018). From welfare to warfare: New deal spending and patriotism during world war ii.

Cassar, A., P. Grosjean, and S. Whitt (2013). Legacies of violence: trust and market development. *Journal of Economic Growth 18*(3), 285–318.

Choi, J.-K. and S. Bowles (2007). The coevolution of parochial altruism and war. *science 318*(5850), 636–640.

Cilliers, J., O. Dube, and B. Siddiqi (2016). Reconciling after civil conflict increases social capital but decreases individual well-being. *Science 352*(6287), 787–794.

Collier, P. et al. (2003). *Breaking the conflict trap: Civil war and development policy*. World Bank Publications.

Collier, P. and A. Hoeffler (2005). Resource rents, governance, and conflict. *Journal of conflict resolution 49*(4), 625–633.

Costa, D. L. and M. E. Kahn (2010). *Heroes and cowards: The social face of war*. Princeton University Press.

Costa, D. L., N. Yetter, and H. DeSomer (2018). Intergenerational transmission of paternal trauma among us civil war ex-pows. *Proceedings of the National Academy of Sciences 115*(44), 11215–11220.

Couttenier, M., V. Preotu, D. Rohner, and M. Thoenig (2016). The violent legacy of victimization: post-conflict evidence on asylum seekers, crimes and public policy in switzerland.

Dahl, G. and S. DellaVigna (2009). Does movie violence increase violent crime? *The Quarterly Journal of Economics 124*(2), 677–734.

Dal Bo, E., F. Finan, O. Folke, J. Rickne, and T. Persson (2017). Economic losers and political winners: The rise of the radical right in sweden. Technical report, mimeo Uppsala University.

De Chaisemartin, C. and X. d'Haultfoeuille (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–96.

DellaVigna, S., R. Enikolopov, V. Mironova, M. Petrova, and E. Zhuravskaya (2014). Cross-border media and nationalism: Evidence from serbian radio in croatia. *American Economic Journal: Applied Economics 6*(3), 103–32.

Depetris-Chauvin, E., R. Durante, and F. Campante (2020). Building nations through shared experiences: Evidence from african football. *American Economic Review 110*(5), 1572–1602.

Derouen, K. R. J. and J. Bercovitch (2008). Enduring Internal Rivalries: A New Framework for the Study of Civil War. *Journal of Peace Research 45*(1), 55–74.

Dippel, C., A. Ferrara, and S. Heblich (2020). Causal mediation analysis in instrumental-variables regressions. *The Stata Journal 20*(3), 613–626.

Dippel, C. and S. Heblich (2021). Leadership in social movements: Evidence from the" forty-eighters" in the civil war. *American Economic Review 111*(2), 472–505.

Enke, B. (2020). Moral values and voting. *Journal of Political Economy 128*(10), 3679–3729.

Fearon, J. D., M. Humphreys, and J. M. Weinstein (2015). How does development assistance affect collective action capacity? results from a field experiment in post-conflict liberia. *American Political Science Review 109*(3), 450–469.

Feigenbaum, J., J. Lee, and F. Mezzanotti (2018). Capital destruction and economic growth: The effects of sherman's march, 1850-1920. *NBER working paper* (w25392).

Foster, G. M. (1988). *Ghosts of the Confederacy: Defeat, the Lost cause, and the Emergence of the New South, 1865-1913*. Oxford University Press.

Fouka, V. (2019). Backlash: The Unintended Effects of Language Prohibition in U.S. Schools after World War I. *The Review of Economic Studies 87*(1), 204–239.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in us schools after world war i. *The Review of Economic Studies 87*(1), 204–239.

Fouka, V., S. Mazumder, and M. Tabellini (2020). From immigrants to americans: Race and assimilation during the great migration.

Fryer Jr, R. G. and S. D. Levitt (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics 119*(3), 767–805.

Gallagher, G. W. and A. T. Nolan (2000). *The myth of the lost cause and Civil War history*. Indiana University Press.

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

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

Goldin, C. D. and F. D. Lewis (1975). The economic cost of the american civil war: Estimates and implications. *The Journal of Economic History 35*(2), 299–326.

Guiso, L., H. Herrera, M. Morelli, and T. Sonno (2017). Populism: demand and supply. Technical report.

Guriev, S. (2018). Economic drivers of populism. In *AEA Papers and Proceedings*, Volume 108, pp. 200–203.

Guriev, S. and E. Papaioannou (2020). The political economy of populism. *Available at SSRN 3542052*.

Hacker, J. D. and J. M. McPherson (2011). A census-based count of the civil war dead: With introductory remarks by james m. mcpherson. *Civil War History 57*(4), 307–348.

Halberstam, Y. and B. Knight (2016). Homophily, group size, and the diffusion of politican information in social networks: Evidence from twitter. *Journal of Public Economics 143*, 73–88.

Henrich, J. and R. Boyd (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of theoretical biology 208*(1), 79–89.

Jensen, R. and E. Oster (2009). The power of tv: Cable television and women's status in india. *The Quarterly Journal of Economics 124*(3), 1057–1094.

Kazin, M. (2017). *War Against War: The American Fight for Peace, 1914–1918*. Simon and Schuster.

Kearney, M. S. and P. B. Levine (2015). Media influences on social outcomes: The impact of mtv's 16 and pregnant on teen childbearing. *American Economic Review 105*(12), 3597–3632.

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

Lieberson, S. and E. O. Bell (1992). Children's first names: An empirical study of social taste. *American Journal of sociology 98*(3), 511–554.

Lupu, N. and L. Peisakhin (2017). The legacy of political violence across generations. *American Journal of Political Science 61*(4), 836–851.

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

Mazumder, S. (2019). Becoming white: How military service turned immigrants into americans. Technical report, Center for Open Science.

Mukand, S. and D. Rodrik (2018). The political economy of ideas: on ideas versus interests in policymaking. Technical report, National Bureau of Economic Research.

Nowell-Smith, G. (1996). *The Oxford history of world cinema*. Oxford University Press.

Nyman, R., S. Kapadia, D. Tuckett, D. Gregory, P. Ormerod, and R. Smith (2018). News and narratives in financial systems: exploiting big data for systemic risk assessment.

Pastor, L. and P. Veronesi (2018). Inequality aversion, populism, and the backlash against globalization. Technical report, National Bureau of Economic Research.

Pierce, D. (2013). The survival of american silent feature films, 1912-1929. Council on Library and Information Resources and the Library of Congress.

Piketty, T. (2018). Brahmin left vs merchant right: Rising inequality and the changing structure of political conflict. *WID. World Working Paper 7*.

Pischke, S. (2007). Lecture notes on measurement error. *London School of Economics, London*.

Rohner, D., M. Thoenig, and F. Zilibotti (2013). Seeds of distrust: Conflict in uganda. *Journal of Economic Growth 18*(3), 217–252.

Shiller, R. J. (2017). Narrative economics. *American Economic Review 107*(4), 967–1004.

Silber, N. (1997). *The romance of reunion: Northerners and the South, 1865-1900*. Univ of North Carolina Press.

Stokes, M. (2007). *DW Griffith's the Birth of a Nation: a history of the most controversial motion picture of all time*. Oxford University Press on Demand.

Strömberg, D. (2004). Radio's impact on public spending. *The Quarterly Journal of Economics 119*(1), 189–221.

Tabellini, M. (2020). Racial heterogeneity and local government finances: Evidence from the great migration.

Voigtländer, N. and H.-J. Voth (2012). Persecution Perpetuated: The Medieval Origins of Anti-Semitic Violence in Nazi Germany. *The Quarterly Journal of Economics 127*(3), 1339–1392.

Voigtlander, N. and H.-J. Voth (2015). Nazi indoctrination and anti-semitic beliefs in germany. *Proceedings of the National Academy of Sciences 112*(26), 7931–7936.

Voth, H.-J. (2020). Persistence–myth and mystery. *CEPR Working Paper*.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: Evidence from the rwandan genocide. *The Quarterly Journal of Economics 129*(4), 1947–1994.

ONLINE APPENDIX - For Online Publication Only

# RECONCILIATION NARRATIVES: –"The Birth of a Nation" after the US Civil War–

Elena Esposito Tiziano Rotesi Alessandro Saia Mathias Thoenig

This Online Appendix accompanies the paper –Reconciliation Narratives: "The Birth of a Nation" after the US Civil War–. Section A presents additional material complementing the Context and Data section in the main text. Extended Results and robustness analysis of the effect of exposure to Birth of a Nation on newspaper rhetoric, Navy enlistment and naming patterns are reported in Section B, C, and D. Section E further develops the analysis on the mediating effect of the common enemy argument on reconciliation, while Section F complements the analysis on the geographical heterogeneity of the effect.

## Contents

**A Context and Data** **III**
A.1 The Narrative of the Birth of a Nation . . . III
A.2 Measuring Movie Diffusion From Historical Newspapers . . . VI
A.3 Instrumenting the Screening of The Birth of Nation: The Million Dollars Mystery . . X
A.4 Reconciliation and Patriotism: Measurement of Patriotism through Navy Enlistment XI

**B Exposure to The Birth of a Nation and Public Debate: Extended Results** **XII**
B.1 First Stage Estimates . . . XII
B.2 Alternative Measures of Rhetoric . . . XIV
B.2.1 Robustness Exercises . . . XIV
B.2.2 Measurement of Patriotic and Divisive Rhetoric . . . XVII
B.3 Measurement of Screenings, Sample and Coverage . . . XVIII
B.3.1 Alternative Ways to Control for Newspaper Coverage . . . XVIII
B.3.2 Alternative Samples . . . XX
B.3.3 Alternative Measures of Screenings . . . XXIII
B.4 Instrumental Variable Strategy: Other Movies . . . XXV
B.5 Alternative Specifications . . . XXVII
B.5.1 County-Specific Linear Time-Trends . . . XXVII
B.5.2 Accounting for Time-Varying Confounders . . . XXVIII
B.5.3 Accounting for State Year Fixed Effects . . . XXIX
B.5.4 Only Counties that Screened BoaN by 1920 . . . XXX
B.5.5 Accounting for Spatial Correlation in the Errors . . . XXXI
B.5.6 Word-Based Analysis . . . XXXII

**C Reconciliation and Patriotism: Extended Results** **XXXIII**
C.1 Robustness . . . XXXIII
C.1.1 Alternative Sample . . . XXXIII
C.1.2 Alternative Measures of BoaN Exposure . . . XXXIV

C.2 Alternative Measures of Enlistment . . . XXXVI
C.2.1 Transformation of the Dependent Variable . . . XXXVI
C.2.2 Enlistment Measured with Deaths of Infectious Disease . . . XXXIX

**D Reconciliation Culture and Names of the Former Enemies: Extended Results XL**
D.1 Robustness Analysis . . . XL
D.1.1 Alternative Measures of Names' Connotation . . . XL
D.1.2 Alternative Coverage Thresholds . . . XLI
D.1.3 Alternative Definitions of BoaN Exposure . . . XLII
D.2 Collapsed Sample: County-Year Level Analysis . . . XLV
D.3 Alternative Definitions of Former Enemies . . . XLVI
D.4 Accounting for Migration Flows . . . XLVII
D.5 Falsification Exercises . . . L
D.5.1 Falsification Exercise - Black Individuals . . . L
D.5.2 Falsification Exercise - Older Cohorts . . . LIII
D.5.3 Accounting for Spatial Clustering in the Errors . . . LVII

**E White Reconciliation and Racial Discrimination LVIII**
E.1 BoaN Exposure, Alienation and Discrimination . . . LVIII
E.1.1 Alternative Measures of Supremacism . . . LVIII
E.1.2 Alternative Measures of Racial Discrimination . . . LIX
E.1.3 Alternative Measures of BoaN Screenings . . . LX
E.2 Alienation and Reconciliation . . . LXI
E.2.1 Alienation and Reconciliation: Mediation Analysis Approach . . . LXI
E.2.2 Mediation Analysis Approach: OLS and Reduced Form . . . LXIII

**F A Pervasive Reconciliation LXV**
F.1 Heterogeneous Effects in Confederate and Unionist States . . . LXV

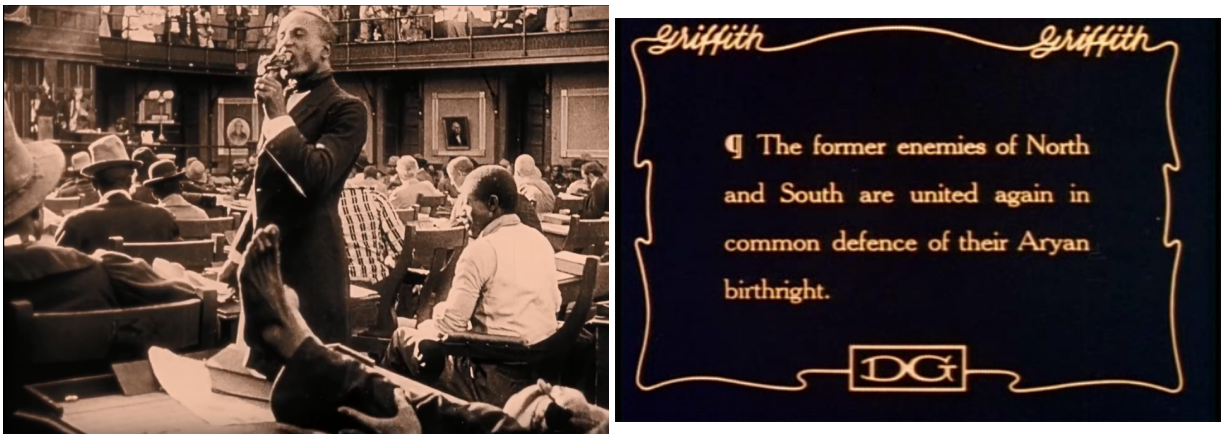
# A Context and Data

## A.1 The Narrative of the Birth of a Nation

Figure A1: The "true" Story of the Civil War and the Reconstruction

The movie "*The Birth of a Nation*" wants to tell the "true" Story of the Civil War and the Reconstruction. According to its plot, the Civil War was caused by the ambitions of a greedy Northern Republican mulatto politician (portrayed in the *Left Figure*). The ante-bellum South is told as a place of peace and cohesion, where black slaves are loyal and devoted servants of well-meaning masters (*Right Figure*).

Figure A3: The Common Threat: The Emancipation of African Americans

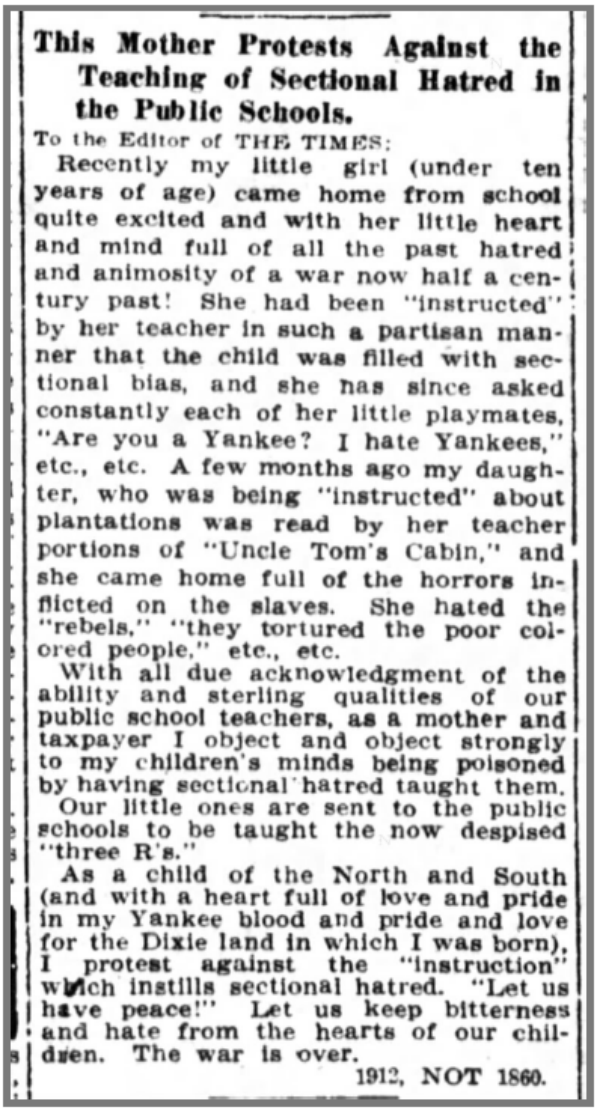



The emancipation and enfranchisement of African Americans is presented as a common threat for the white population in the North and the South. In the *Left Figure* a scene of the movie depicting a state Parliament dominated by Black representatives. In the *Right Figure*, a script from the movie inviting for reconciliation against the common threat.

Figure A5: The Happy Ending: Reconciliation of White Americans


The movie ends with the reconciliation of the North and the South, message reinforced by the marriage of members from the two families (from the North and the South) at the center of the movie.

Figure A7: Civil War Memories and North-South Resentment



**This Mother Protests Against the Teaching of Sectional Hatred in the Public Schools.**

To the Editor of THE TIMES:

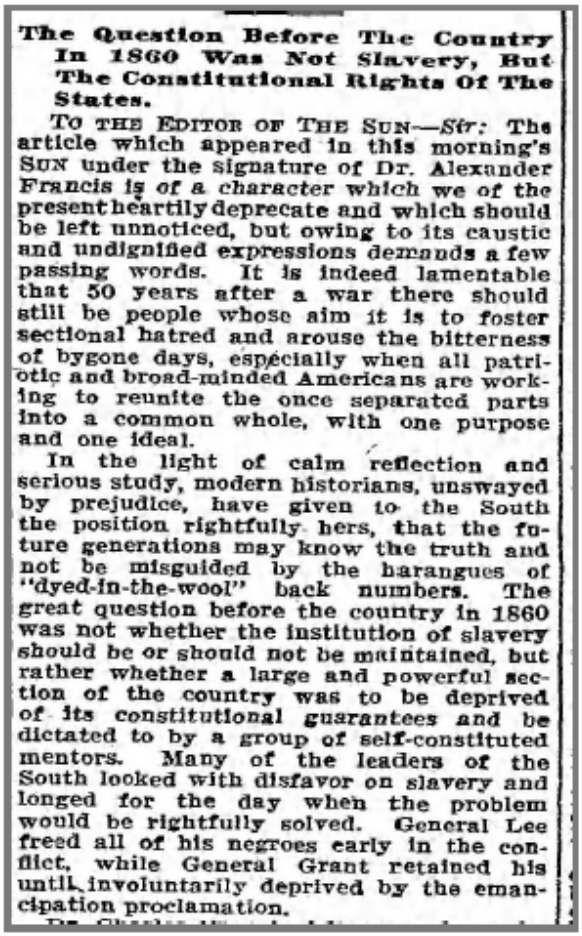
Recently my little girl (under ten years of age) came home from school quite excited and with her little heart and mind full of all the past hatred and animosity of a war now half a century past! She had been "instructed" by her teacher in such a partisan manner that the child was filled with sectional bias, and she has since asked constantly each of her little playmates, "Are you a Yankee? I hate Yankees," etc., etc. A few months ago my daughter, who was being "instructed" about plantations was read by her teacher portions of "Uncle Tom's Cabin," and she came home full of the horrors inflicted on the slaves. She hated the "rebels," "they tortured the poor colored people," etc., etc.

With all due acknowledgment of the ability and sterling qualities of our public school teachers, as a mother and taxpayer I object and object strongly to my children's minds being poisoned by having sectional hatred taught them.

Our little ones are sent to the public schools to be taught the now despised "three R's."

As a child of the North and South (and with a heart full of love and pride in my Yankee blood and pride and love for the Dixie land in which I was born), I protest against the "instruction" which instills sectional hatred. "Let us have peace!" Let us keep bitterness and hate from the hearts of our children. The war is over.

1912, NOT 1860.



**The Question Before The Country In 1860 Was Not Slavery, But The Constitutional Rights Of The States.**

TO THE EDITOR OF THE SUN—*Sir:* The article which appeared in this morning's SUN under the signature of Dr. Alexander Francis is of a character which we of the present heartily deprecate and which should be left unnoticed, but owing to its caustic and undignified expressions demands a few passing words. It is indeed lamentable that 50 years after a war there should still be people whose aim it is to foster sectional hatred and arouse the bitterness of bygone days, especially when all patriotic and broad-minded Americans are working to reunite the once separated parts into a common whole, with one purpose and one ideal.

In the light of calm reflection and serious study, modern historians, unswayed by prejudice, have given to the South the position rightfully hers, that the future generations may know the truth and not be misguided by the harangues of "dyed-in-the-wool" back numbers. The great question before the country in 1860 was not whether the institution of slavery should be or should not be maintained, but rather whether a large and powerful section of the country was to be deprived of its constitutional guarantees and be dictated to by a group of self-constituted mentors. Many of the leaders of the South looked with disfavor on slavery and longed for the day when the problem would be rightfully solved. General Lee freed all of his negroes early in the conflict, while General Grant retained his until involuntarily deprived by the emancipation proclamation.

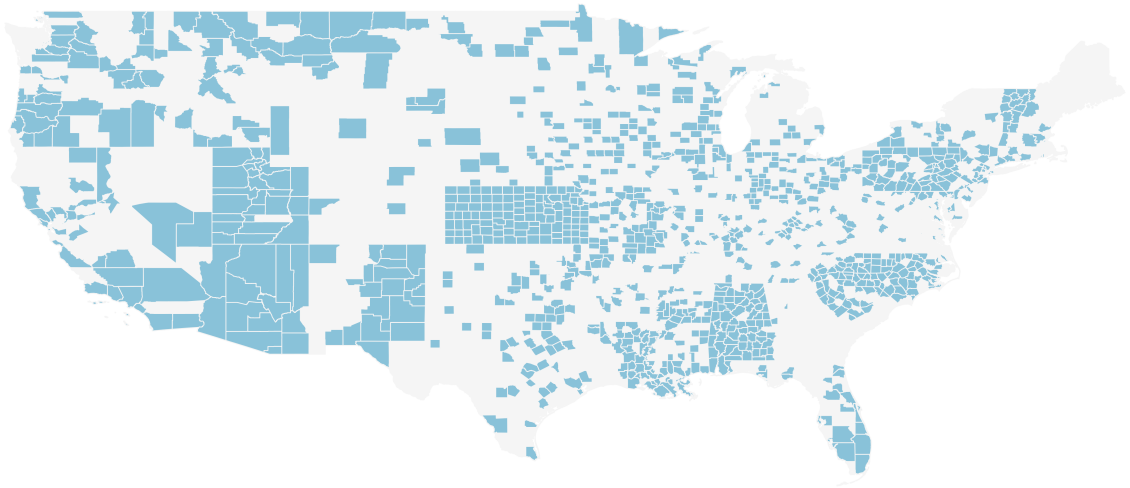
NOTE: Examples of persistent and divisive underlying tensions between North and South emerging on newspapers in the first decades 20th century. *Left Figure* The Washington Times, 21 May 1912. *Right Figure* The Baltimore Sun, 6 September 1914. Source: newspapers.com

## A.2 Measuring Movie Diffusion From Historical Newspapers

We map the distribution of the movie across counties and over time retrieving information on movie screenings from the online archive *newspapers.com*, the largest online newspaper archive. Our baseline analysis is conducted at monthly level, and includes all months from 1910 to 1920 (132 months).

**Newspaper Coverage and Sample Definition** Between 1910 and 1920, we rely on more than 3,700 newspapers with headquarters located across 1,911 cities. The first issue is to associate newspapers to counties. We attribute each newspaper to the county where the newspaper's headquarter is located. For the state of Alabama, for instance, between 1910 and 1920, we can rely on newspapers with headquarters in 99 cities, located across 63 counties.[47]

Figure A9: Newspapers Headquarters



NOTE: Blue polygons represent counties with a newspaper's headquarter as per our database. We consider all newspapers for which we could retrieve at least one newspaper page digitized on newspapers.com between 1910 and 1920. See the text for further details.

Many counties do not have newspaper coverage for some/all the months we consider. The second issue, therefore, is to define our sample based on newspaper coverage. This step is necessary to disentangle counties/months that do not feature screenings of "*The Birth of a Nation*" from counties/months that do not have historical newspapers digitized on *newspapers.com* on that month, and exclude the latter from our sample of interest.

To do so, we only keep counties/months that have at least one newspaper page digitized in the month. We use two criteria to detect whether a county has at least one newspaper page digitized in

[47] These cities are: Centre Lovan, Butler, Cullman, Fayette, Geneva, Jackson, Marion, Albertville, Andalusia, Anniston, Atmore Birmingham, Brewton, Camden, Carrollton, Centreville, Claiborne, Clanton, Columbia, Decatur, Demopolis, Dothan, Enterprise, Fairhope, Falkville, Florence, Foley, Gadsden, Gainesville, Georgiana, Goodwater, Hayneville, Heflin, Huntsville, Jasper, LaFayette, Leighton, Linden, Luverne, Mobile, Monroeville, Montgomery, New Decatur, Newton, Ozark, Piedmont, Rockford, Russellville, Scottsboro, Seale, Talladega, Thomaston, Troy, Tuscaloosa, Tuscumbia, Tuskegee, Uniontown, Wetumpka, Abbeville, Albany, Alexander City, Ashland, Ashville, Bay Minette, Boaz, Elba, Epes, Eutaw, Evergreen, Florala, Fort Payne, Girard, Greensboro, Greenville, Grove Hill, Guntersville, Haleyville, Hamilton, Hartselle, Headland, Lineville, Livingston, Moulton, Muscle Shoals, Oneonta, Opp, Phil Campbell, Prattville, Roanoke, Selma, St. Stephens, Stevenson, Union Springs, Vernon, West Blocton, Lanett, Clayton, Grand Bay.

the month. First, we retrieve information on newspaper coverage as provided by newspapers.com (https://www.newspapers.com/browse/us). To refine this first criteria, we search all newspaper pages containing one of the keyword "he", "you" or "I". If at least one of these conditions is met, we include the county/month in the sample.

We end up with an unbalanced sample including 1,068 counties. The average county is covered for around 100 months, and we observe on average 800 counties per month.

**Collecting Newspaper Pages referring to The Birth of a Nation** We then proceed collecting all newspapers pages containing the keywords "The Birth of a Nation" in our sample of interest. We retrieve a total of 56,140 items, from 1,837 newspapers, located across 838 counties. These hits include three main types of items: i) movie ads (see an example on the left panel of Figure 1); ii) movie time table (see an example on the right panel of Figure 1); iii) general articles referring to the movie (see an example on the left panel of Figure A10).

A large share of items retrieved are not related to movie screenings but are general reviews of the movie or articles discussing the reactions to the movie across the country (see an example on the left panel of Figure A10). We ask external judges to read each item collected containing the keywords "*The Birth of a Nation*". The judges are asked to identify movie ads and movie time table. Judges are also asked to read long articles and disentangle cases such as the left and right panel of Figure A10. This process allows us to obtain 14,933 "true" screening records, from 990 newspapers located across 632 counties. Random checking of the work of judges suggests that 92% of items related to a screening are correctly identified, with only an average of 8% of our items remaining in the sample as false screening records. Table A12 depicts the total number of newspaper pages containing the keywords "*The Birth of a Nation*" by month (# BoaN Keywords) and the "true" screening records verified judges (# Verified Screening Records). Finally, when considering treated counties as those with at least two verified screening records, we obtain a sample of 526 counties treated counties and 542 untreated counties.

Table A1: Counties with and without BoaN

| Variable | (1) Untreated Counties | (2) Treated Counties | (3) Difference |
|---|---|---|---|
| Share of Illiterate (10+), 1910 | 0.112 | 0.082 | -0.030*** |
| | (0.113) | (0.097) | (0.006) |
| Share of Blacks, 1910 | 0.152 | 0.120 | -0.032** |
| | (0.219) | (0.202) | (0.013) |
| Share of Males, 1910 | 0.525 | 0.520 | -0.005** |
| | (0.033) | (0.030) | (0.002) |
| Share of Rural Population, 1910 | 0.847 | 0.642 | -0.205*** |
| | (0.216) | (0.275) | (0.015) |
| Share of Population in Cities (25k), 1910 | 0.036 | 0.129 | 0.093*** |
| | (0.153) | (0.275) | (0.014) |
| Share of Urban Population, 1910 | 0.152 | 0.351 | 0.199*** |
| | (0.214) | (0.272) | (0.015) |
| Share of Foreign Population, 1910 | 0.063 | 0.106 | 0.042*** |
| | (0.081) | (0.101) | (0.006) |
| Observations | 542 | 526 | 1,068 |

NOTE: The table displays demographic characteristics of counties that screened *The Birth of a Nation* by the end on 1920 (i.e. Treated Counties) and counties that did not screen the movie by the end of 1920 (i.e. Untreated Counties). The last column presents the difference in the mean value of each variable between the two groups of counties.

Figure A10: Newspaper Page with *The Birth of a Nation* - A True Positive and a False Positive

# Birth of Nation Must Not Show In Denver.

Klu Klux to be Born in Colorado. South Pushes Jim Crowism and Klu Klux Plays North and West. Will Colorado be like Carolina, Home of Ben Tilllman?

**NEGROES PACK SHORTER TUESDAY NIGHT AT MASS MEETING**

The Birth of a Nation really means the new nation that was born under the immortal Emancipation Proclamation of Abraham Lincoln, in the beginning of 1865. As written and staged by Tom Dixon, it means the nation born of the loins of the political, secret, oath-bound bands of assassins known throughout the Southern States as the Ku Klux Klan. The Klan began began in 1865, at Pulaski, Tenn., and expanded itself into a great federation of regulators, having 50,000 members in the state of Tennessee and 550,000 throughout the states then lately in the Rebellion. The entire South was name by them the "Invisible Empire", placed under the domination of Gen. M. B. Forrest, who was called the "Grand Wizard." This was the same [illegible] who, as general of the Confederates at Fort Pillow [illegible] cred ihe Negro soldiers after their surrender at the battle. Each state was under a Dragon, each county under a Titan, and similar divisions under Grand Giants. Grand Cyclops, etc. These weird titles, taken from demonolgy, [illegible] the ghoulish and diabolical character of these oath-bound organizations of assassins.

# VETERANS AGAIN LIVE OVER WAR IN GREAT MOVIE

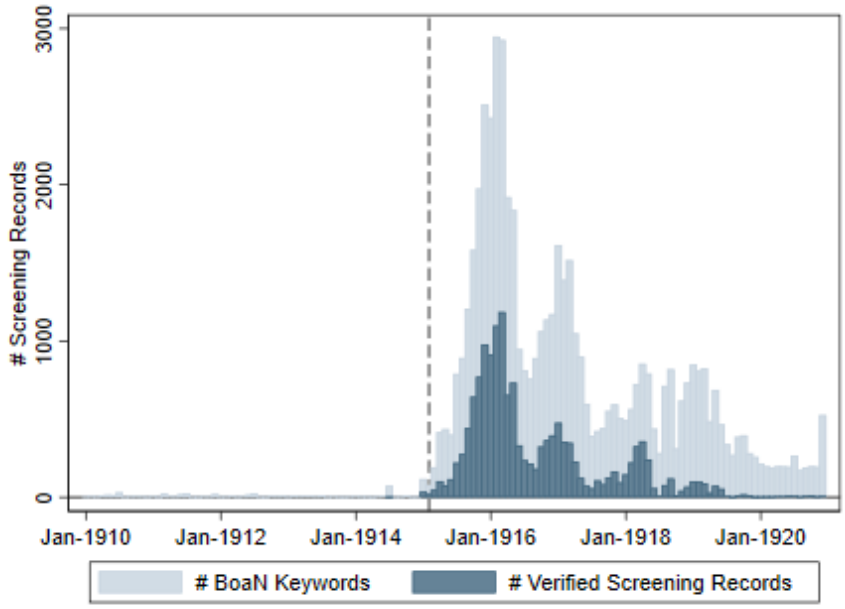
Old Soldiers Declare That "Birth of a Nation" Battle Scenes Are Faithful to Life

Some rulers seem to fear that there is no burning and suffering hereafter to which their faithful and loving subjects may be consigned after death, so they hurry into war to supply them with that necessity beforehand.

Any one who imagines there is a halo of beautiful romance about the soldiers' lot—that battle is grand and suffering and falling on the firing line the proud goal to which man's ambition should always aim, will be most beautifully disillusionized when they experience the "roar, the fury, the Hell of war," as only D. W. Griffith, in a thousand decisive battles of the Civil war, has shown it up to the white light of day, in his overwhelming spectacle, "The Birth of a Nation," which appears at the Majestic theater for a week beginning Sunday afternoon.

NOTE: *Left Figure* An article about the movie The Birth of a Nation, with no information about screenings in the local theathers. Source: The Indianapolis News, 14 December 1915 [newspapers.com]. *Right Figure* An article about the movie The Birth of a Nation with information regarding a forthcoming screening of the movie at cinema Majestic. Source: La Crosse Tribune, 12 November 1915 [newspapers.com].

Figure A12: Total Number of Screening Records Over Time

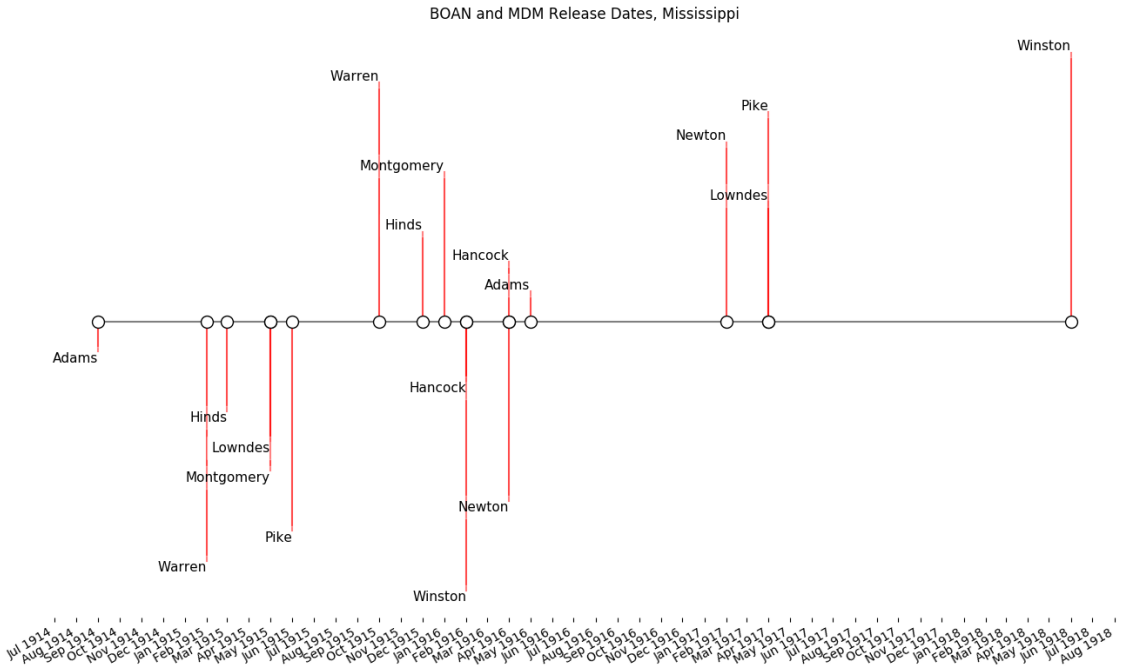



NOTE: Light blue bars depict the total number of newspaper pages containing the keyword "*The Birth of a Nation*" at each point in time. The blue bars depict the total number of screening records verified by external judges.

## A.3 Instrumenting the Screening of The Birth of Nation: The Million Dollars Mystery

Our instrumental variable strategy exploits the time and geographical patterns in the distribution of "*The Million Dollar Mystery*", a very successful movie produced and distributed around 8 months before "*The Birth of a Nation*", pretending that MDM was released on the same date of *The Birth of a Nation*. Figure A13 provides a graphical illustration of the logic of our instrument by looking at Mississippi. On the top panel we report for each month the names of counties where a screening of *The Million Dollar Mystery* was advertised, the bottom panel does the same for *The Birth of a Nation*. First, the two set of counties where the two movies are screened largely overlap. Second, while the difference in movie release is not always 8 months, it does tend to be around 8 months for several counties. Figure A2 presents, in a 2 $x$ 2 tables, the total number of counties that screened (and did not screen) *The Birth of a Nation* and *Million Dollar Mystery* by the end of 1920.

Figure A13: BoaN and MDM Dates of Release: Mississippi Timeline




NOTE: On the top panel, the date of release of MDM in Mississippi counties, on the bottom panel, the date of release of BoaN.

Table A2: Counties that Screened *The Birth of a Nation* and *Million Dollar Mystery*

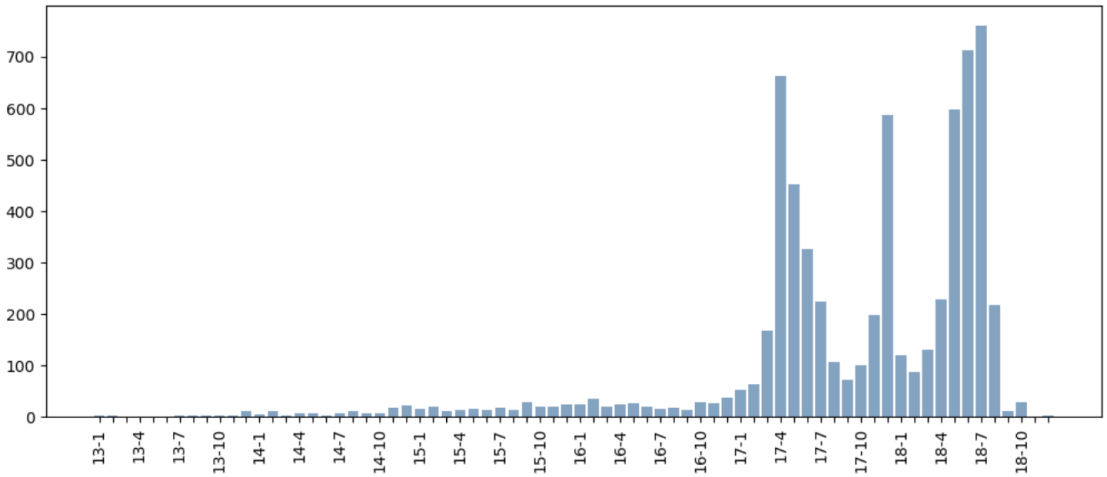
| | MDM | | |
|---|---|---|---|
| **BoaN** | Untreated | Treated | Total |
| Untreated | 449 | 93 | 542 |
| Treated | 164 | 362 | 526 |
| Total | 613 | 455 | 1,068 |

NOTE: The table depicts the number of counties that screened *The Birth of a Nation* and *Million Dollar Mystery* by the end of 1920 within a 2x2 tables.

## A.4 Reconciliation and Patriotism: Measurement of Patriotism through Navy Enlistment

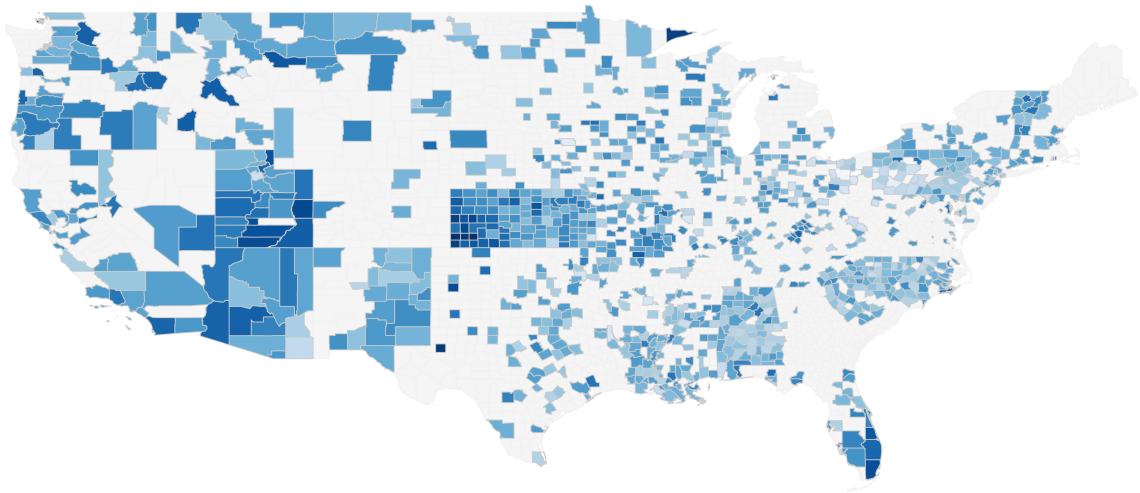
To measure patriotism, we construct a new database mapping Navy enlistment across counties over time, relying on information from the the pool of individuals that died at War. Figure A14 presents the time variation of our data, depicting the total number of marines enlisted by month. Figure A15 documents the total cross-county variation in enlistment by the end of our period of interest.

Figure A14: Enlistment in the Navy




NOTE: The histogram depicts the total number of marines enlisted by month, among the pool of individuals that died at War. See Section 5.1 for further details.

Figure A15: The Geographical Distribution of Navy Enlistments



NOTE: The figure maps the total enlistment across counties in our sample, measured as log share of total enlistment $\texttt{logList}^{\texttt{tot}}$ by the end of 1920. Darker shades of blue are associated with higher values.

# B Exposure to The Birth of a Nation and Public Debate: Extended Results

## B.1 First Stage Estimates

This section report first stage estimates of the relationship between $\texttt{BoaN}_{ct}$ and $\texttt{MDM}_{ct}$. Column (1) of Table B1 is the baseline first stage that is relevant for tables in the main text. We then replicate the baseline first stage estimate of the baseline table (Table 2) for all the alternative specifications presented in the paper.

Table B1 reports the first stage estimates with our different definitions of $\texttt{BoaN}_{ct}$, i.e taking value 1 after the threshold of 2, 3 and 5 screening ads in the county respectively. Table B2 verifies the robustness of the first stage relationship with alternative specifications, accounting for coverage in different ways and focusing on a balanced sample of counties. In particular, Column (1) reproduces our baseline first stage estimate, Column (2) does not include coverage percentile fixed effects, Column (3) include coverage decile fixed effects, Column (4) does not restrict the sample to counties with at least 25% of newspaper coverage, and Column (5) reproduces the first stage on the balanced sample of counties that have full coverage between 1910 and 1920. Table B3 replicates first stage estimates on shorter samples, including all months until 1916 and 1917 (Columns 2 and 3 respectively), proposes a specification including county-specific linear time trends (Column 4), and considers only counties exposed to *The Birth of a Nation* by the end of the sample period (Column 5).

Table B1: BoaN and MDM: First Stage Estimates

| | Birth of a Nation | | |
|---|---|---|---|
| | (1) Threshold=2 | (2) Threshold=3 | (3) Threshold=5 |
| Million Dollar Mystery (Threshold=2) | 0.318*** | | |
| | (0.0247) | | |
| Million Dollar Mystery (Threshold=3) | | 0.338*** | |
| | | (0.0247) | |
| Million Dollar Mystery (Threshold=5) | | | 0.360*** |
| | | | (0.0249) |
| Observations | 102,772 | 102,772 | 102,772 |
| R-squared | 0.690 | 0.668 | 0.641 |
| F-stat | 166.244 | 186.923 | 209.305 |

NOTE: The table reports first stage estimates of the relationship between $\texttt{BoaN}_{ct}$ and $\texttt{MDM}_{ct}$. As instrument, we use Million Dollar Mystery, which is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise. To assign a screening, we use three different thresholds. Column (1) reports our baseline specification, where we consider a movie as screened in the county if we detect at least 2 screening records. In Column (2) the threshold is 3 screening records, while in Column (3) the threshold is 5 screening records. The unit of observation is the county ($c$) in a particular month-year ($t$). All regressions control for county, month-year, and coverage percentile fixed effects, and restrict the sample to counties with at least 25% of newspaper coverage. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B2: BoaN and MDM: First Stage Estimates

| | Birth of a Nation | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Million Dollar Mystery | 0.318*** | 0.322*** | 0.319*** | 0.320*** | 0.279*** |
| | (0.0247) | (0.0248) | (0.0247) | (0.0246) | (0.0341) |
| Observations | 102,772 | 102,772 | 102,772 | 104,584 | 61,116 |
| R-squared | 0.690 | 0.688 | 0.689 | 0.691 | 0.706 |
| F-stat | 166.244 | 168.95 | 167.113 | 169.811 | 66.616 |

NOTE: The table reports first stage estimates of the relationship between $\texttt{BoaN}_{ct}$ and $\texttt{MDM}_{ct}$. As instrument, we use Million Dollar Mystery, which is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise. The unit of observation is the county ($c$) in a particular month-year ($t$). Column (1) reports our baseline specification, in which we control for county, month-year, and coverage percentile fixed effects, while restricting the sample to counties with at least 25% of newspaper coverage. Column (2) does not include coverage percentile fixed effects. Column (3) includes coverage decile fixed effects. Column (4) does not restrict the sample to counties with at least 25% of newspaper coverage. Column (5) restricts the analysis to the balanced sample of counties that have full coverage between 1910 and 1920. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: BoaN and MDM: First Stage Estimates

| | Birth of a Nation | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Million Dollar Mystery | 0.318*** | 0.223*** | 0.239*** | 0.153*** | 0.0637*** |
| | (0.0247) | (0.0185) | (0.0194) | (0.0190) | (0.0157) |
| Observations | 102,772 | 65,706 | 68,066 | 102,772 | 64,521 |
| R-squared | 0.690 | 0.494 | 0.525 | 0.854 | 0.849 |
| F-stat | 166.244 | 144.946 | 151.693 | 64.731 | 16.361 |

NOTE: The table reports first stage estimates of the relationship between $\texttt{BoaN}_{ct}$ and $\texttt{MDM}_{ct}$. As instrument, we use Million Dollar Mystery, which is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise. The unit of observation is the county ($c$) in a particular month-year ($t$). Column (1) reports our baseline specification, in which we control for county, month-year, and coverage percentile fixed effects, while restricting the sample to counties with at least 25% of newspaper coverage. Column (2) restricts the sample to observations up to December 1916. Column (3) restricts the sample to observations up to March 1917. Column (4) reports results for a specification which includes county-specific linear time trends. Column (5) restricts the analysis to counties exposed to *The Birth of a Nation* by the end of the sample period. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## B.2 Alternative Measures of Rhetoric

### B.2.1 Robustness Exercises

We present an array of exercises to verify robustness of results to variation in the construction of the dependent variable. We first replicate our measure of log frequencies by modifying $log(1 + \texttt{Freq}_{i,c,t})$ into $log(0.01 + \texttt{Freq}_{i,c,t})$, Table B4, and $log(0.1 + \texttt{Freq}_{i,c,t})$, Table B5. Table B6 presents the analysis on the sine hyperbolic transformation of $\texttt{Freq}_{i,c,t}$. Table B7 uses, as outcome, the total sum of frequencies, without aggregating them into a scalar through principal component analysis.

Table B4: Robustness of the Logarithmic Transformation - $log(0.01 + \texttt{Freq}_{i,c,t})$

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.088*** | | 0.399*** | 0.026 | | 0.059 | -0.063*** | | -0.340*** |
| | (0.019) | | (0.063) | (0.026) | | (0.081) | (0.015) | | (0.053) |
| Million Dollar Mystery | | 0.127*** | | | 0.019 | | | -0.108*** | |
| | | (0.0190) | | | (0.026) | | | (0.016) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.650 | 0.651 | -0.014 | 0.678 | 0.678 | -0.000 | 0.702 | 0.703 | -0.024 |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Dep. Var. Mean | .249 | .249 | .249 | -1.582 | -1.582 | -1.582 | -1.831 | -1.831 | -1.831 |
| Dep. Var. Std. Dev. | 1.094 | 1.094 | 1.094 | 1.059 | 1.059 | 1.059 | .825 | .825 | .825 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county ($c$) in a particular month-year ($t$). Log frequencies for word $i$ are calculated as $log(0.01 + \texttt{Freq}_{i,c,t})$. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Robustness of the Logarithmic Transformation - $log(0.1 + \texttt{Freq}_{i,c,t})$

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.121*** | | 0.499*** | 0.057** | | 0.18** | -0.064*** | | -0.319*** |
| | (0.0178) | | (0.061) | (0.023) | | (0.074) | (0.013) | | (0.047) |
| Million Dollar Mystery | | 0.159*** | | | 0.057** | | | -0.101*** | |
| | | (0.018) | | | (0.024) | | | (0.014) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.715 | 0.716 | -0.031 | 0.762 | 0.762 | -0.003 | 0.886 | 0.886 | -0.030 |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Dep. Var. Mean | .19 | .19 | .19 | -1.68 | -1.68 | -1.68 | -1.87 | -1.87 | -1.87 |
| Dep. Var. Std. Dev. | .98 | .98 | .98 | 1.039 | 1.039 | 1.039 | 1.077 | 1.077 | 1.077 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county ($c$) in a particular month-year ($t$). Log frequencies for word $i$ are calculated as $log(0.1 + \texttt{Freq}_{i,c,t})$. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B6: Robustness of the Logarithmic Transformation - Sine Hyperbolic Transformation

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 1.084*** | | 4.246*** | 0.589*** | | 2.093*** | -0.496*** | | -2.153*** |
| | (0.105) | | (0.396) | (0.127) | | (0.412) | (0.0812) | | (0.292) |
| Million Dollar Mystery | | 1.350*** | | | 0.666*** | | | -0.685*** | |
| | | (0.107) | | | (0.127) | | | (0.086) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.846 | 0.847 | -0.086 | 0.916 | 0.916 | -0.020 | 0.974 | 0.974 | -0.042 |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Sample Mean | 11.901 | 11.901 | 11.901 | -46.909 | -46.909 | -46.909 | -58.81 | -58.81 | -58.81 |
| Sample SD | 6.578 | 6.578 | 6.578 | 8.752 | 8.752 | 8.752 | 12.266 | 12.266 | 12.266 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of the sine hyperbolic transformation of patriotic words' frequencies minus the first principal component of the sine hyperbolic transformation of divisive words' frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of the sine hyperbolic transformation of patriotic words' frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of the sine hyperbolic transformation of divisive words' frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B7: Total Sum of Frequencies

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.985*** | | 3.798*** | 0.564*** | | 2.019*** | -0.421*** | | -1.780*** |
| | (0.09) | | (0.345) | (0.107) | | (0.351) | (0.067) | | (0.239) |
| Million Dollar Mystery | | 1.208*** | | | 0.642*** | | | -0.566*** | |
| | | (0.093) | | | (0.107) | | | (0.071) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.866 | 0.867 | - | 0.946 | 0.946 | - | 0.984 | 0.984 | - |
| 1st stage F-stat | - | - | 166.24 | - | - | 166.24 | - | - | 166.24 |
| Dep. Var. Mean | 10.13 | 10.13 | 10.13 | -41.7 | -41.7 | -41.7 | -51.83 | -51.83 | -51.83 |
| Dep. Var. Std. Dev. | 5.756 | 5.756 | 5.756 | 9.071 | 9.071 | 9.071 | 12.571 | 12.571 | 12.571 |

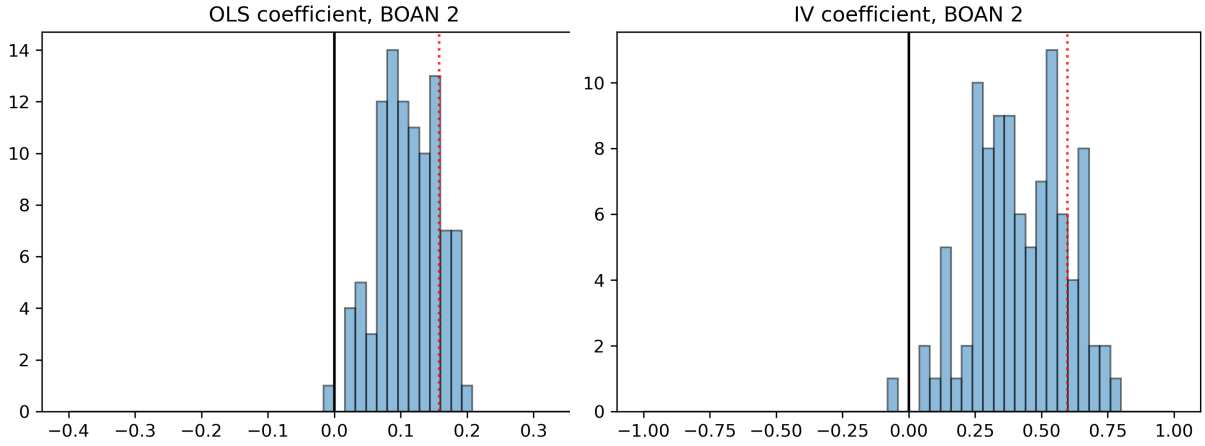
NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the sum of patriotic words' frequencies minus the sum of divisive words' frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the sum of patriotic words' frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the sum of divisive words' frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.2.2 Measurement of Patriotic and Divisive Rhetoric

This section presents results verifying our hypothesis with an array of alternative definitions of Patriotic and Divisive rhetoric. We start from our *Patriotic* and *Divisive* lists of words used for our baseline anaylsis. We then reconstruct 100 alternative *Patriotic* and *Divisive* lists by randomly composing new lists of words, where each word from the starting list has a 0.5 probability of being chosen for the new list. Armed with these new lists, we reconstruct 100 different versions of the variable $\texttt{Reconciliation}_{c,t}$.

Figure B1 summarizes results of 100 regressions that have, as outcome, $\texttt{Reconciliation}_{c,t}$ measured in our 100 different ways, as described above. On the left panel, we report OLS estimates, on the right panel, 2SLS estimates. The figure shows that all our measures of $\texttt{Reconciliation}_{c,t}$ provide results fully in line with our baseline one: *The Birth of a Nation* increases reference to a reconciliation rhetoric.

Figure B1: Conciliatory and Divisive List- Permutation Exercise




NOTE: The figures above each display 100 coefficients of the variable $\texttt{BoaN}_{ct}$ regressed on 100 alternative measures of $\texttt{Reconciliation}_{c,t}$ (see the text for further details). On the left panel, we report OLS estimates. On the right panel, 2SLS estimates.

## B.3 Measurement of Screenings, Sample and Coverage

### B.3.1 Alternative Ways to Control for Newspaper Coverage

This section replicates our baseline results using alternative ways to control for the newspaper coverage of the county, measured as the total newspaper pages available in a county in a month. In our baseline table - Table 2 in the main text - we flexibly control for newspaper coverage by accounting for the fixed effects of coverage percentiles. Table B8 replicates Table 2 in the main text without accounting for the fixed effects of coverage percentile. In Table B9, instead, we account for decile fixed effects of coverage.

Table B8: Not Accounting for Coverage FE

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.184*** | | 0.647*** | -0.003 | | 0.137 | -0.187*** | | -0.511*** |
| | (0.0158) | | (0.06) | (0.03) | | (0.103) | (0.029) | | (0.103) |
| Million Dollar Mystery | | 0.208*** | | | 0.044 | | | -0.164*** | |
| | | (0.017) | | | (0.033) | | | (0.033) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.791 | 0.792 | - | 0.840 | 0.840 | - | 0.895 | 0.895 | - |
| 1st stage F-stat | - | - | 168.95 | - | - | 168.95 | - | - | 168.95 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and month-year. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B9: Accounting for FE of Coverage Deciles

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.164*** | | 0.612*** | 0.083*** | | 0.303*** | -0.081*** | | -0.309*** |
| | (0.0156) | | (0.058) | (0.02) | | (0.065) | (0.012) | | (0.043) |
| Million Dollar Mystery | | 0.195*** | | | 0.097*** | | | -0.099*** | |
| | | (0.016) | | | (0.02) | | | (0.013) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.799 | 0.800 | - | 0.917 | 0.917 | - | 0.967 | 0.967 | - |
| 1st stage F-stat | - | - | 167.113 | - | - | 167.113 | - | - | 167.113 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage deciles fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.3.2 Alternative Samples

Our baseline results focus on counties with at least 25% of monthly coverage. Table B10 replicate our baseline estimates restricting the sample to counties with newspaper coverage for all the months in the 1910-1920 period (132 months). Table B11 replicates baseline results imposing no restriction based on coverage.

Our results might be amplified by the outbreak of the first World War. To show that results are not driven by nationalism and patriotism related to the start of the War, we replicate baseline results restricting the sample to months before April 1917, Table B12. Given that even the months strictly preceding the US entry into the War might be characterized by a patriotic rhetoric, we further restrict the samples to months before January 1917, Table B13. Importantly, there is no relevant change in the size of the coefficients of interest.

Table B10: Balanced Sample

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.144*** | | 0.719*** | 0.073*** | | 0.391*** | -0.071*** | | -0.328*** |
| | (0.017) | | (0.092) | (0.022) | | (0.09) | (0.013) | | (0.058) |
| Million Dollar Mystery | | 0.200*** | | | 0.109*** | | | -0.091*** | |
| | | (0.018) | | | (0.022) | | | (0.014) | |
| Observations | 61,116 | 61,116 | 61,116 | 61,116 | 61,116 | 61,116 | 61,116 | 61,116 | 61,116 |
| R-squared | 0.810 | 0.812 | - | 0.930 | 0.930 | - | 0.979 | 0.979 | - |
| 1st stage F-stat | - | - | 66.616 | - | - | 66.616 | - | - | 66.616 |
| Dep. Var. Mean | .318 | .318 | .318 | -1.971 | -1.971 | -1.971 | -2.289 | -2.289 | -2.289 |
| Dep. Var. Std. Dev. | .783 | .783 | .783 | 1.31 | 1.31 | 1.31 | 1.624 | 1.624 | 1.624 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). The sample is restricted to include only counties that have full coverage between 1910 and 1920. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## Table B11: No Restriction on Coverage

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.158*** | | 0.597*** | 0.097*** | | 0.342*** | -0.061*** | | -0.256*** |
| | (0.016) | | (0.057) | (0.019) | | (0.061) | (0.01) | | (0.035) |
| Million Dollar Mystery | | 0.191*** | | | 0.109*** | | | -0.082*** | |
| | | (0.016) | | | (0.019) | | | (0.01) | |
| Observations | 104,584 | 104,584 | 104,584 | 104,584 | 104,584 | 104,584 | 104,584 | 104,584 | 104,584 |
| R-squared | 0.801 | 0.802 | - | 0.934 | 0.934 | - | 0.981 | 0.981 | - |
| 1st stage F-stat | - | - | 169.811 | - | - | 169.811 | - | - | 169.811 |
| Dep. Var. Mean | .116 | .116 | .116 | -1.683 | -1.683 | -1.683 | -1.799 | -1.799 | -1.799 |
| Dep. Var. Std. Dev. | .785 | .785 | .785 | 1.431 | 1.431 | 1.431 | 1.693 | 1.693 | 1.693 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). With respect to the baseline, the sample here is not restricted to counties with at least 25% coverage. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## Table B12: Sample Restriction: Before 1917

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.128*** | | 0.638*** | 0.081*** | | 0.421*** | -0.046*** | | -0.216*** |
| | (0.013) | | (0.062) | (0.017) | | (0.07) | (0.01) | | (0.041) |
| Million Dollar Mystery | | 0.152*** | | | 0.101*** | | | -0.051*** | |
| | | (0.013) | | | (0.016) | | | (0.009) | |
| Observations | 68,066 | 68,066 | 68,066 | 68,066 | 68,066 | 68,066 | 68,066 | 68,066 | 68,066 |
| R-squared | 0.776 | 0.777 | - | 0.948 | 0.948 | - | 0.981 | 0.981 | - |
| 1st stage F-stat | - | - | 151.693 | - | - | 151.693 | - | - | 151.693 |
| Dep. Var. Mean | -0.065 | -0.065 | -0.065 | -1.837 | -1.837 | -1.837 | -1.772 | -1.772 | -1.772 |
| Dep. Var. Std. Dev. | .685 | .685 | .685 | 1.422 | 1.422 | 1.422 | 1.656 | 1.656 | 1.656 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). The sample is restricted to include only observations up to March 1917. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B13: Sample Restriction: Before 1916

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.1*** | | 0.608*** | 0.052*** | | 0.369*** | -0.049*** | | -0.239*** |
| | (0.013) | | (0.064) | (0.017) | | (0.073) | (0.011) | | (0.044) |
| Million Dollar Mystery | | 0.135*** | | | 0.082*** | | | -0.053*** | |
| | | (0.012) | | | (0.016) | | | (0.009) | |
| Observations | 65,706 | 65,706 | 65,706 | 65,706 | 65,706 | 65,706 | 65,706 | 65,706 | 65,706 |
| R-squared | 0.775 | 0.776 | -0.084 | 0.950 | 0.950 | -0.033 | 0.981 | 0.981 | -0.023 |
| 1st stage F-stat | - | - | 144.946 | - | - | 144.946 | - | - | 144.946 |
| Dep. Var. Mean | -.076 | -.076 | -.076 | -1.847 | -1.847 | -1.847 | -1.771 | -1.771 | -1.771 |
| Dep. Var. Std. Dev. | .675 | .675 | .675 | 1.425 | 1.425 | 1.425 | 1.655 | 1.655 | 1.655 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). The sample is restricted to include only observations up to December 1916. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.3.3 Alternative Measures of Screenings

This section presents results using alternative proxies of movie screenings. In our baseline analysis, we define $\texttt{BoaN}_{ct}$ as an indicator variable taking on value of 1 when a county reaches the threshold of 2 verified movie screening ads. Table B14 replicates baseline results with a threshold of 3, and Table B15 with a threshold of 5. Increasing the threshold tend to decrease the size of IV estimates. To understand why, we can define $p_1$ as the probability that we observe a county to have screened BoaN when this truly happened, and $p_0$ as the probability that we observe a county to have screened BoaN when, in fact, BoaN was not screened. It can be shown (see, for instance, Pischke, 2007; and Black, Berger and Scott, 1995) that the instrumental variable estimate simplifies to:

$$\texttt{plim}\,\hat{\beta_{IV}} \equiv \left[\frac{\beta}{p_1 - p_0}\right] \tag{8}$$

In the presence of misclassication of the variable $\texttt{BoaN}_{ct}$, therefore, the instrumental variable estimates can be biased. The direction of the bias depends on the relative size of $p_1$ and $p_0$. By increasing the threshold of ads that define exposure to the movie, we are implicitly increasing $p_1$ and decreasing $p_0$. Thus, increasing the threshold is likely reducing the upward bias in $\hat{\beta_{IV}}$, and this is consistent with results presented in this section.

Table B14: Conciliatory vs Divisive - Screening of the Movie: At Least 3 Ads

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.162*** | | 0.592*** | 0.096*** | | 0.328*** | -0.066*** | | -0.265*** |
| | (0.017) | | (0.055) | (0.02) | | (0.058) | (0.011) | | (0.033) |
| Million Dollar Mystery | | 0.200*** | | | 0.111*** | | | -0.089*** | |
| | | (0.016) | | | (0.019) | | | (0.01) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.801 | 0.802 | - | 0.933 | 0.933 | - | 0.981 | 0.981 | - |
| 1st stage F-stat | - | - | 186.923 | - | - | 186.923 | - | - | 186.923 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise. Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. We consider a movie as screened in the county if we detect at least 3 screening records (the threshold is 2 in the baseline specification). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B15: Conciliatory vs Divisive - Screening of the Movie: At Least 5 Ads

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.190*** | | 0.588*** | 0.111*** | | 0.335*** | -0.08*** | | -0.253*** |
| | (0.018) | | (0.051) | (0.022) | | (0.055) | (0.012) | | (0.031) |
| Million Dollar Mystery | | 0.212*** | | | 0.120*** | | | -0.091*** | |
| | | (0.016) | | | (0.019) | | | (0.011) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.802 | 0.803 | - | 0.933 | 0.933 | - | 0.981 | 0.981 | - |
| 1st stage F-stat | - | - | 209.305 | - | - | 209.305 | - | - | 209.305 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise. Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. We consider a movie as screened in the county if we detect at least 5 screening records (the threshold is 2 in the baseline specification). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.4 Instrumental Variable Strategy: Other Movies

As per baseline instrument, we use the spread of *The Million Dollar Mystery*, a movie released 231 days before *The Birth of a Nation*, to predict the diffusion of *The Birth of a Nation*. *The Million Dollar Mystery* is a natural choice as it is the second movie of the period in terms of revenue and in terms of geographic spread across the country. In fact, while *The Birth of a Nation* was screened by the end of 1920 in around 50% of counties in our sample, *The Million Dollar Mystery* reaches 43% of counties. While other movies reached a more limited diffusion, given the recurrent spatio-temporal diffusion of movies across the country, they could also be exploited as instrument, as long as their plot is unrelated with the topics of the Birth of a Nation.

We verify the consistency of the logic of our instrument with 2 other movies: "*Traffic in Souls*" (1913), a movie that was screened in around 25% of the counties in our sample; and "*What Happened to Mary*" (1912), screened in 32% of counties. We construct two additional instruments, $\texttt{Traffic}_{ct}$ and $\texttt{Mary}_{ct}$, mimicking the construction of our baseline instrument $\texttt{MDM}_{ct}$. The only difference is that while for *Birth of a Nation* and *Million Dollar Mystery* the screening proofs were verified by external judges, for the two other movies for simplicity we use the total number of keyword hits found on newspapers.

Table B16 reports 2SLS and reduced form estimates using as an instrument: our baseline instrument $\texttt{MDM}_{ct}$ constructed with screening proof verified by judges, $\texttt{MDM}_{ct}$ constructed using the total number of keyword hits found on newspapers (with no check of judges), $\texttt{Mary}_{ct}$ constructed based on the diffusion of the movie "*What Happened to Mary*" (with no check of judges), $\texttt{Traffic}_{ct}$ constructed on the movie "*Traffic in Souls*" (with no check of judges). Reduced form and instrumental variable estimates are remarkably close to our baseline ones, suggesting that the spatio-temporal diffusion of movies followed very similar trajectories.

Table B16: Alternative Movies as Instrument

| | Million Dollar Mystery | | Million Dollar Mystery (no check) | | What Happened to Mary | | Traffic in Souls | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Dep Var: Reconciliation | 2SLS | RF | 2SLS | RF | 2SLS | RF | 2SLS | RF |
| Birth of a Nation | 0.599*** | | 0.560*** | | 0.572*** | | 0.593*** | |
| | (0.058) | | (0.055) | | (0.056) | | (0.061) | |
| Million Dollar Mystery | | 0.190*** | | | | | | |
| | | (0.0159) | | | | | | |
| Million Dollar Mystery (no check) | | | | 0.178*** | | | | |
| | | | | (0.016) | | | | |
| What Happened to Mary | | | | | | 0.194*** | | |
| | | | | | | (0.016) | | |
| Traffic in Souls | | | | | | | | 0.192*** |
| | | | | | | | | (0.017) |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | - | 0.802 | - | 0.802 | - | 0.802 | - | 0.802 |
| 1st stage F-stat | 166.244 | - | 173.21 | - | 204.753 | - | 199.874 | - |
| Dep. Var. Mean | .122 | .122 | .122 | .122 | .122 | .122 | .122 | .122 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | .788 | .788 | .788 | .788 | .788 |

NOTE: The table reports 2SLS (Columns 1, 3, 5, and 7), and reduced form (Columns 2, 4, 6, and 8) estimates. The dependent variables is the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise. Columns (1) and (2) report results using Million Dollar Mystery, which is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. Columns (3) and (4) consider the same movie and includes screening records that were filtered out by judges. Columns (5) and (6) consider the movie *What Happened to Mary*, while Columns (7) and (8) are based on the movie *Traffic in Souls*. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## B.5 Alternative Specifications

### B.5.1 County-Specific Linear Time-Trends

Table B17 present baseline specification when including county-specific linear time trends. The overall effect on $\texttt{Reconciliation}_{c,t}$ is confirmed, while estimates of $\texttt{Divisive}_{c,t}$ rhetoric become noisy. Note, however, that an important caveat applies to these estimates. Figure 5 in the main text documents that the effect is increasing over time. Whenever the effect of interest changes over time, including group-specific linear time trends can bias estimates. The intuition is that the trend might also capture the difference in the evolution of the outcome between treated and control observations that relates to the treatment. For a complete discussion of this issue see, for instance, Wolfers (2006), and Kahn-Lang and Lang (2019).

Table B17: Accounting for County Specific Time Trends

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.069*** | | 0.708*** | 0.085*** | | 0.626*** | 0.016* | | -0.082 |
| | (0.014) | | (0.11) | (0.02) | | (0.121) | (0.009) | | (0.06) |
| Million Dollar Mystery | | 0.108*** | | | 0.096*** | | | -0.013 | |
| | | (0.014) | | | (0.018) | | | (0.009) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.828 | 0.828 | - | 0.945 | 0.945 | - | 0.984 | 0.984 | - |
| 1st stage F-stat | - | - | 64.731 | - | - | 64.731 | - | - | 64.731 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, coverage percentile fixed effects, and county specific linear time trends. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.5.2 Accounting for Time-Varying Confounders

Appendix Table A1 documents that counties that screened *The Birth of a Nation* are significantly different from counties that did not screen the movie along several dimensions, $\texttt{X}_c$ : a lower share of male inhabitants, a lower share of illiterate population, a slightly lower share of blacks, a lower (higher) share of rural (urban) population, a higher share of people living in cities of more than 25 thousands inhabitants, and a higher share of foreign population. In order to exclude that $\texttt{BoaN}_{ct}$ is capturing the specific trajectories of more inhabited and more culturally/economically dynamic counties, we include in our baseline specifications an interaction term between each dimension, $\texttt{X}_c$, and a time fixed effects. This approach allows to control very flexibly for potential time-varying effects related to each dimension. In the result we present below, we account for all dimensions that are found to differ significantly between counties that screen BoaN (at some point in time) and other counties. Table B18 contains regression results for specifications allowing the effect of $\texttt{X}_c$ to vary at month-year level.

Table B18: Accounting for Confounders Varying at Month-Year Level

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.089*** | | 0.680*** | 0.078*** | | 0.439*** | -0.012 | | -0.241*** |
| | (0.017) | | (0.111) | (0.021) | | (0.122) | (0.01) | | (0.067) |
| Million Dollar Mystery | | 0.135*** | | | 0.087*** | | | -0.048*** | |
| | | (0.017) | | | (0.023) | | | (0.012) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.811 | 0.812 | - | 0.935 | 0.935 | - | 0.982 | 0.982 | - |
| 1st stage F-stat | - | - | 61.317 | - | - | 61.317 | - | - | 61.317 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Additionally, regressions include the interaction between confounders and month-year fixed effects. The list of confounders includes the share of male inhabitants, share of illiterate population, share of blacks, share of rural population, share of urban population, share of population living in cities of more than 25K inhabitants, and share of foreign population. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.5.3 Accounting for State Year Fixed Effects

Another way to account for time-varying confounders that operate at the state-level, it is to account for state-specific year fixed effects. By doing so, all confounders that affect the outcome of each state in specific years is controlled for. Table B19 presents the baseline specification including state-year fixed effects.

Table B19: Accounting for State-Year Fixed Effects

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.168*** | | 0.604*** | 0.117*** | | 0.352*** | -0.051*** | | -0.251*** |
| | (0.015) | | (0.064) | (0.019) | | (0.069) | (0.011) | | (0.038) |
| Million Dollar Mystery | | 0.167*** | | | 0.097*** | | | -0.069*** | |
| | | (0.016) | | | (0.019) | | | (0.01) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.812 | 0.812 | - | 0.938 | 0.938 | - | 0.982 | 0.982 | - |
| 1st stage F-stat | - | - | 140.65 | - | - | 140.65 | - | - | 140.65 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, coverage percentile, and state-year fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.5.4 Only Counties that Screened BoaN by 1920

Table B20 replicates baseline results only in the sub-sample of counties that are screened *The Birth of a Nation* by the end of 1920. While results are more noisy, and the difference between OLS and IV increases, the baseline pattern is confirmed.

Table B20: Only Counties that Screened *The Birth of a Nation* by 1920

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.058*** | | 2.151*** | 0.06*** | | 0.99** | 0.002 | | -1.161*** |
| | (0.014) | | (0.583) | (0.017) | | (0.495) | (0.008) | | (0.342) |
| Million Dollar Mystery | | 0.137*** | | | 0.0630** | | | -0.074*** | |
| | | (0.02) | | | (0.027) | | | (0.016) | |
| Observations | 64,521 | 64,521 | 64,521 | 64,521 | 64,521 | 64,521 | 64,521 | 64,521 | 64,521 |
| R-squared | 0.797 | 0.799 | - | 0.910 | 0.910 | - | 0.976 | 0.976 | - |
| 1st stage F-stat | - | - | 16.361 | - | - | 16.361 | - | - | 16.361 |
| Dep. Var. Mean | .321 | .321 | .321 | -2.015 | -2.015 | -2.015 | -2.336 | -2.336 | -2.336 |
| Dep. Var. Std. Dev. | .804 | .804 | .804 | 1.255 | 1.255 | 1.255 | 1.582 | 1.582 | 1.582 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). See Section 4.3 for further details. The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). The sample restricts the analysis to counties exposed to *The Birth of a Nation* by the end of 1920. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.5.5 Accounting for Spatial Correlation in the Errors

Turning to statistical inference, this section replicates our baseline estimates accounting for spatial correlation in the errors. Table B21 presents results with standard errors adjusted for two-dimensional spatial dependence, using cutoff thresholds of 50, 100, and 200 km.

Table B21: Conley Standard Errors

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.158*** | | 0.599*** | 0.097*** | | 0.343*** | -0.061*** | | -0.256*** |
| *Spatial Cluster 50km* | (0.005) | | (0.015) | (0.005) | | (0.015) | (0.003) | | (0.01) |
| *Spatial Cluster 100km* | (0.005) | | (0.016) | (0.005) | | (0.016) | (0.003) | | (0.01) |
| *Spatial Cluster 200km* | (0.006) | | (0.018) | (0.006) | | (0.018) | (0.004) | | (0.011) |
| Million Dollar Mystery | | 0.190*** | | | 0.109*** | | | -0.081*** | |
| *Spatial Cluster 50km* | | (0.005) | | | (0.005) | | | (0.003) | |
| *Spatial Cluster 100km* | | (0.005) | | | (0.005) | | | (0.003) | |
| *Spatial Cluster 200km* | | (0.006) | | | (0.006) | | | (0.003) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.063 | 0.068 | - | 0.582 | 0.582 | - | 0.819 | 0.819 | - |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9) (see Section 4.3 for further details). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year, and coverage percentile fixed effects. The table reports standard errors adjusted for two-dimensional spatial dependence using cutoff thresholds of 50, 100, and 200km. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### B.5.6 Word-Based Analysis

This section presents an empirical exercise focusing on single words across counties over time. We construct a word-county level dataset using all divisive and patriotic keywords in our lists. We add to this list, 80 neutral words, selected to be unrelated to patriotic and divisive words. To investigate the heterogeneous effects of $\texttt{BoaN}_{ct}$ on Patriotic and Divisive keywords, we estimate the following specification:

$$\texttt{\# of word}_{ict} \equiv \beta_0 + \beta_1 \texttt{BoaN}_{ct} * \texttt{Divisive}_i + \beta_2 \texttt{BoaN}_{ct} * \texttt{Concil.}_i + \texttt{FE}_{ct} + \texttt{FE}_{ic} + \epsilon_{ict}$$

where $i$ stands for word, $c$ for county and $t$ for month. $\texttt{\# of word}_{ict}$ measures the number of pages containing the word $i$ in county $c$ at month $t$. $\texttt{FE}_{ct}$ include a full set of county-month fixed effects. The inclusion of this set of fixed effects allows us to more flexibly account for all possible unknown confounders operating at the county-month level. $\texttt{FE}_{ic}$ stand for county-words fixed effects, and account for average occurrence of each word in our list in a county. All in all, within this empirical setting all events affecting counties at a specific point in time, such as local elections, protests or strikes will be accounted for (as long as their effect on the occurence of words is the same for all the words in our lists). The coefficient $\beta_1$ measures the effect of exposure to Birth of Nation for the set of words in the Divisive list, when compared to the neutral words lists. The coefficient $\beta_2$ measures the effect of exposure to Birth of Nation for the set of words in the Patriotic list, when compared to the neutral words in the sample.

Table B22 presents OLS and ITT results for the word-based analysis. The exposure of Birth of Nation increases the relative frequency of Patriotic words, while decreasing the relative frequency of Divisive words.

Table B22: Word-Based Regression - Conciliatory and Divisive versus Neutral Words

| | (1) OLS | (2) RF | (3) OLS | (4) RF | (5) OLS | (6) RF |
|---|---|---|---|---|---|---|
| Birth of a Nation * Patriotic Word | 6.892*** | | | | 6.892*** | |
| | (0.431) | | | | (0.431) | |
| Birth of a Nation * Divisive Word | | | -1.511*** | | -1.511*** | |
| | | | (0.512) | | (0.512) | |
| Million Dollar Mystery * Patriotic Word | | 7.021*** | | | | 7.021*** |
| | | (0.456) | | | | (0.456) |
| Million Dollar Mystery * Divisive Word | | | | -1.644*** | | -1.644*** |
| | | | | (0.509) | | (0.509) |
| Observations | 9,557,796 | 9,557,796 | 9,557,796 | 9,557,796 | 10,791,060 | 10,791,060 |
| R-squared | 0.943 | 0.943 | 0.937 | 0.937 | 0.942 | 0.942 |

NOTE: The table reports OLS (Columns 1, 3, and 5) and reduced form (Columns 2, 4, and 6) estimates. The dependent variables are the occurrence of each word. The unit of observation is the word ($i$) in a particular county-month-year ($ct$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). Patriotic Word is an indicator variable that takes value 1 if word $i$ is a Patriotic keywords. Divisive Word is an indicator variable that takes value 1 if word $i$ belongs to the list of Divisive keywords (see Section 4 for further details). Estimates presented in Columns (1) and (2) restrict the analysis to Patriotic and Neutral words, estimates presented in Columns (3) and (4) restrict the analysis to Divisive and Neutral words, estimates presented in Columns (5) and (6) includes Patriotic, Divisive and Neutral words. All regressions control for county-month, word-county fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# C Reconciliation and Patriotism: Extended Results

## C.1 Robustness

### C.1.1 Alternative Sample

Table 3 in the main text includes all counties for which we measure exposure to *The Birth of a Nation*. As a robustness exercise, to verify consistency with results on the rhetoric of local newspapers, we replicate our analysis on the same samples of counties of the newspapers sample (i.e. counties that have newspaper coverage for at least 25% of months in our sample). Results are confirmed.

Table C1: BoaN and Enlistments - Alternative Newspaper Coverage

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.027*** | | 0.077*** | 0.012*** | | 0.031*** |
| | (0.004) | | (0.02) | (0.003) | | (0.012) |
| Million Dollar Mystery | | 0.02*** | | | 0.007*** | |
| | | (0.006) | | | (0.003) | |
| Observations | 55,694 | 55,694 | 55,694 | 40,018 | 40,018 | 40,018 |
| R-squared | 0.976 | 0.976 | - | 0.994 | 0.994 | - |
| 1st stage F-Stat | - | - | 136 | - | - | 128 |
| Dep. Var. Mean | -10.15 | -10.15 | -10.15 | -10.18 | -10.18 | -10.18 |
| Dep. Var. Std. Dev. | .98 | .98 | .98 | 1 | 1 | 1 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$ (see Section 5.1 for further details). The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). The sample here is restricted to counties with at least 25% coverage. All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### C.1.2 Alternative Measures of BoaN Exposure

Our baseline measure of exposure to *The Birth of a Nation*, $\texttt{BoaN}_{ct}$, relies on a threshold of 2. Table C2 and C3 replicate baseline results with the $\texttt{BoaN}_{ct}$ indicator taking on value 1 after respectively 3 and 5 movie screening proofs in the county.

Table C2: BoaN and Enlistments - 3 Ads as Threshold

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.039*** | | 0.07*** | 0.015*** | | 0.023*** |
| | (0.005) | | (0.014) | (0.004) | | (0.009) |
| Million Dollar Mystery | | 0.024*** | | | 0.006*** | |
| | | (0.005) | | | (0.002) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.979 | 0.978 | - | 0.995 | 0.995 | - |
| 1st stage F-Stat | - | - | 301 | - | - | 245 |
| Dep. Var. Mean | -10.09 | -10.09 | -10.09 | -10.11 | -10.11 | -10.11 |
| Dep. Var. Std. Dev. | .96 | .96 | .96 | .98 | .98 | .98 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$ (see Section 5.1 for further details). The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect at least 3 screening records (the threshold is 2 in the baseline specification). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: BoaN and Enlistments - 5 Ads as Threshold

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.049*** | | 0.076*** | 0.017*** | | 0.023** |
| | (0.006) | | (0.016) | (0.004) | | (0.01) |
| Million Dollar Mystery | | 0.026*** | | | 0.006** | |
| | | (0.006) | | | (0.003) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.979 | 0.978 | - | 0.995 | 0.995 | - |
| 1st stage F-Stat | - | - | 281 | - | - | 222 |
| Dep. Var. Mean | -10.09 | -10.09 | -10.09 | -10.11 | -10.11 | -10.11 |
| Dep. Var. Std. Dev. | .96 | .96 | .96 | .98 | .98 | .98 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$. See Section 5.1 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect at least 5 screening records (the threshold is 2 in the baseline specification). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## C.2 Alternative Measures of Enlistment

### C.2.1 Transformation of the Dependent Variable

Results presented in this Section replicate the baseline analysis with different transformations of the variable enlistment. In Table C4, we present results with the dependent variable calculated as:

$$\texttt{logList}_{c,t} \equiv \log\left[\frac{0.1 + \texttt{sailors}_{c,t}}{0.1 + \texttt{Pop}_c}\right] \tag{9}$$

While, in Table C5:

$$\texttt{logList}_{c,t} \equiv \log\left[\frac{0.01 + \texttt{sailors}_{c,t}}{0.01 + \texttt{Pop}_c}\right] \tag{10}$$

Table C6 presents results with the inverse hyperbolic sine transformation of the variables $\texttt{sailors}_{c,t}$ and $\texttt{Pop}_c$. Table C7 looks, instead, at the extensive margin of the effect. The dependent variable takes on value 1 with all county-month with at least one volunteer.

Table C4: BoaN and Enlistments - Robustness of the Log. Transformation - Log +0.1

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.09*** | | 0.184*** | 0.038*** | | 0.08*** |
| | (0.012) | | (0.035) | (0.01) | | (0.025) |
| Million Dollar Mystery | | 0.064*** | | | 0.022*** | |
| | | (0.013) | | | (0.007) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.828 | 0.828 | - | 0.948 | 0.948 | - |
| 1st stage F-Stat | - | - | 299 | - | - | 251 |
| Dep. Var. Mean | -12.34 | -12.34 | -12.34 | -12.39 | -12.39 | -12.39 |
| Dep. Var. Std. Dev. | .98 | .98 | .98 | .97 | .97 | .97 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$ calculated as in equation (9). The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: BoaN and Enlistments - Robustness of the Log. Transformation - Log +0.01

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.161*** | | 0.326*** | 0.07*** | | 0.147*** |
| | (0.021) | | (0.059) | (0.019) | | (0.047) |
| Million Dollar Mystery | | 0.114*** | | | 0.04*** | |
| | | (0.022) | | | (0.013) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.566 | 0.565 | - | 0.826 | 0.826 | - |
| 1st stage F-Stat | - | - | 299 | - | - | 251 |
| Dep. Var. Mean | -14.57 | -14.57 | -14.57 | -14.68 | -14.68 | -14.68 |
| Dep. Var. Std. Dev. | 1.14 | 1.14 | 1.14 | 1 | 1 | 1 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$ calculated as in equation (10). The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C6: BoaN and Enlistments - Robustness of the Log. Transformation - Inverse Hyperbolic Sine Transfomation

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.04*** | | 0.086*** | 0.016*** | | 0.033*** |
| | (0.006) | | (0.018) | (0.004) | | (0.011) |
| Million Dollar Mystery | | 0.03*** | | | 0.009*** | |
| | | (0.006) | | | (0.003) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.965 | 0.964 | - | 0.992 | 0.992 | - |
| 1st stage F-Stat | - | - | 299 | - | - | 251 |
| Dep. Var. Mean | -10.77 | -10.77 | -10.77 | -10.8 | -10.8 | -10.8 |
| Dep. Var. Std. Dev. | .96 | .96 | .96 | .97 | .97 | .97 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the inverse hyperbolic sine transformation of the variables enlistment in county $c$ at month-year $t$ and population in county $c$. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C7: BoaN and Enlistments - Extensive Margin

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.032*** | | 0.063*** | 0.014*** | | 0.03*** |
| | (0.004) | | (0.011) | (0.004) | | (0.01) |
| Million Dollar Mystery | | 0.022*** | | | 0.008*** | |
| | | (0.004) | | | (0.003) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.192 | 0.191 | - | 0.148 | 0.147 | - |
| 1st stage F-Stat | - | - | 299 | - | - | 251 |
| Dep. Var. Mean | .03 | .03 | .03 | .01 | .01 | .01 |
| Dep. Var. Std. Dev. | .17 | .17 | .17 | .1 | .1 | .1 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is an indicator function taking value 1 if at least one person was enlisted in county $c$ and month-year $t$. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### C.2.2 Enlistment Measured with Deaths of Infectious Disease

Behavior while serving for the Navy during the War and the related mortality rate could be, in principle, also affected by national sentiment and patriotism. It is therefore insightful to replicate baseline results with enlistment data constructed using a sub-sample of individuals who perished for reasons that are less likely to be related to heroism or dedication to the cause.

Table C8 replicates our results on the sub-sample of individuals that died of infectious diseases. The coefficient size and the precision of the estimate do not change.

Table C8: BoaN and Enlistments - Only Individuals who Died of Infectious Diseases

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.026*** | | 0.057*** | 0.009*** | | 0.018*** |
| | (0.004) | | (0.012) | (0.002) | | (0.005) |
| Million Dollar Mystery | | 0.02*** | | | 0.005*** | |
| | | (0.004) | | | (0.002) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.983 | 0.983 | - | 0.997 | 0.997 | - |
| 1st stage F-Stat | - | - | 299 | - | - | 251 |
| Dep. Var. Mean | -10.09 | -10.09 | -10.09 | -10.11 | -10.11 | -10.11 |
| Dep. Var. Std. Dev. | .97 | .97 | .97 | .98 | .98 | .98 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$. See Section 5.1 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We restrict the analysis to the sub-sample of individuals who died of infectious diseases. All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# D Reconciliation Culture and Names of the Former Enemies: Extended Results

## D.1 Robustness Analysis

### D.1.1 Alternative Measures of Names' Connotation

Names from different regions might differ only slightly, because of local dialects or traditions. To verify that our results are not driven by changes in the distribution of names with similar roots, in Table D1, we replicate Table 4 in the main text using the Soundex phonetic equivalent of the first name as dependent variable. The Soundex transformation, in fact, assigns the same value to similar-sounding names.

Table D1: *The Birth of a Nation* and Naming Patterns - Soundex Transformation of First Names

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.337** | | 1.132*** | 0.0103* | | 0.0314*** |
| | (0.152) | | (0.326) | (0.00551) | | (0.0112) |
| Million Dollar Mystery | | 0.585*** | | | 0.0162*** | |
| | | (0.166) | | | (0.00574) | |
| Observations | 129,509 | 129,509 | 129,509 | 129,509 | 129,509 | 129,509 |
| R-squared | 0.044 | 0.044 | - | 0.032 | 0.032 | - |
| 1st stage F-Stat | - | - | 149 | - | - | 149 |
| Dep. Var. Mean | 46.49 | 46.49 | 46.49 | .4 | .4 | .4 |
| Dep. Var. Std. Dev. | 13.53 | 13.53 | 13.53 | 13.53 | .49 | .49 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The Enemy-Sounding Name Index is computed using the Soundex phonetic equivalent of the first name. The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### D.1.2 Alternative Coverage Thresholds

Results presented in the main text are obtained using all counties for which at least one newspaper is observed over the period 1910-1920. Table D2 displays results obtained using 25% coverage threshold, which correspond to the threshold adopted for the baseline analysis on newspaper rhetoric. Results are similar when alternative coverage thresholds are adopted.

Table D2: *The Birth of a Nation* and Naming Patterns - Alternative Newspaper Coverage

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.511** | | 1.756*** | 0.015** | | 0.056*** |
| | (0.240) | | (0.630) | (0.007) | | (0.017) |
| Million Dollar Mystery | | 0.730*** | | | 0.023*** | |
| | | (0.259) | | | (0.007) | |
| Observations | 84,876 | 84,876 | 84,876 | 84,876 | 84,876 | 84,876 |
| R-squared | 0.068 | 0.068 | - | 0.062 | 0.062 | - |
| 1st stage F-Stat | - | - | 74 | - | - | 74 |
| Dep. Var. Mean | 43.92 | 43.92 | 43.92 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.74 | 16.74 | 16.74 | 16.74 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). With respect to the baseline, the sample here is restricted to counties with at least 25% coverage. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### D.1.3 Alternative Definitions of BoaN Exposure

This section documents that results are robust to different definitions of exposure to the movie. Tables D3 and D4 present results with a threshold of 3 and 5 screening proofs, respectively, to define screenings of the movie in a county-year. The analysis on names differ from the analysis of newspaper rhetoric as it relies on yearly variation (instead of year-month variation). In merging month-level data on *The Birth of a Nation* exposure with year-level data on naming patterns, we opted for the standard approach: If *The Birth of a Nation* reaches a county in the Fall of 1915, the county will be considered exposed to the movie for all the year 1915. This is a conservative approach, because even counties mildly treated (for instance counties that screened *The Birth of a Nation* in December 1915) are considered "fully" treated in 1915. To verify that results do not rest on this choice, we create an alternative measure of treatment which considers a county treated only if the movie is screened in the first six months of the year. Table D5 verifies the robustness of our findings when a county is considered to be treated only if the movie is screened in the first semester of the year. As expected, the size of the coefficients tend to be larger.

Table D3: *The Birth of a Nation* and Naming Patterns - Screening of the Movie: At Least 3 Ads

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.792*** | | 1.789*** | 0.020*** | | 0.051*** |
| | (0.216) | | (0.460) | (0.006) | | (0.013) |
| Million Dollar Mystery | | 0.894*** | | | 0.026*** | |
| | | (0.231) | | | (0.006) | |
| Observations | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 |
| R-squared | 0.070 | 0.070 | - | 0.064 | 0.064 | - |
| 1st stage F-Stat | - | - | 96 | - | - | 96 |
| Dep. Var. Mean | 44.07 | 44.07 | 44.07 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.82 | 16.82 | 16.82 | 16.82 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect at least 3 screening records (the threshold is 2 in the baseline specification). All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D4: *The Birth of a Nation* and Naming Patterns - Screening of the Movie: At Least 5 Ads

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.629*** | | 2.045*** | 0.018*** | | 0.051*** |
| | (0.219) | | (0.499) | (0.006) | | (0.013) |
| Million Dollar Mystery | | 1.026*** | | | 0.026*** | |
| | | (0.227) | | | (0.006) | |
| Observations | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 |
| R-squared | 0.070 | 0.070 | - | 0.064 | 0.064 | - |
| 1st stage F-Stat | - | - | 91 | - | - | 91 |
| Dep. Var. Mean | 44.07 | 44.07 | 44.07 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.82 | 16.82 | 16.82 | 16.82 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect at least 5 screening records (the threshold is 2 in the baseline specification). All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D5: *The Birth of a Nation* and Naming Patterns - Screening in the First 6 Months of the Year

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.681*** | | 1.725*** | 0.019*** | | 0.055*** |
| | (0.215) | | (0.494) | (0.006) | | (0.011) |
| Million Dollar Mystery | | 0.784*** | | | 0.0248*** | |
| | | (0.225) | | | (0.006) | |
| Observations | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 |
| R-squared | 0.070 | 0.070 | | 0.064 | 0.064 | |
| 1st stage F-Stat | - | - | 139 | - | - | 139 |
| Sample Mean | 44.07 | 44.07 | 44.07 | .36 | .36 | .36 |
| Sample SD | 16.82 | 16.82 | 16.82 | 16.82 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect a screening record in the first six months of the year. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## D.2 Collapsed Sample: County-Year Level Analysis

For our baseline analysis, we perform the analysis at the individual level, to preserve the underlying variation of our dependent variable. As a robustness exercise, Table D6 replicates baseline results aggregating our measure at the county-year level. In the first three columns, the outcome variables is the average of the index in the county-year, in Columns (4) to (6) the median of the index in the county-year, and in the last three Columns we use as dependent variable the average of the dummy indicator.

Table D6: *The Birth of a Nation* and Naming Patterns - County-Year Sample

| | Enemy Name Index | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Average | | | Mode | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 1.107*** | | 2.232*** | 1.246*** | | 2.461*** | 0.027*** | | 0.051** |
| | (0.334) | | (0.782) | (0.370) | | (0.865) | (0.009) | | (0.022) |
| Million Dollar Mystery | | 0.992*** | | | 1.094*** | | | 0.023** | |
| | | (0.349) | | | (0.387) | | | (0.010) | |
| Observations | 10,128 | 10,128 | 10,128 | 10,128 | 10,128 | 10,128 | 10,128 | 10,128 | 10,128 |
| R-squared | 0.264 | 0.263 | - | 0.247 | 0.247 | - | 0.230 | 0.230 | - |
| 1st stage F-Stat | - | - | 356 | - | - | 356 | - | - | 356 |
| Dep. Var. Mean | 44.87 | 44.87 | 44.87 | 44.88 | 44.88 | 44.88 | .39 | .39 | .39 |
| Dep. Var. Std. Dev. | 10.03 | 10.03 | 10.03 | 10.91 | 10.91 | 10.91 | 10.91 | .28 | .28 |

NOTE: The table reports OLS (Columns 1, 4 and 7), reduced form (Columns 2, 5 and 8) and 2SLS (Columns 3, 6 and 9) estimates. The dependent variable is the average Enemy-Sounding Name Index (ENI) observed in the country-year in Columns (1) to (3), the median of the index in Columns (4) to (6) and the dummy indicator in the last three columns. The unit of observation is computed using all indivuals from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county and year of birth fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## D.3 Alternative Definitions of Former Enemies

For for the construction of the index in our baseline results, the definition of the two groups of former enemies use former Confederate states on one side and non-Confederate states on the other side. Another option to construct the index would have been to use Confederate states on one side and former Unionist states on the other. Table D7 replicates our baseline analysis using this alternative definition of former enemies. Results are, again, unchanged.

Table D7: Former Confederate States versus Former Unionist States

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.694*** | | 1.695*** | 0.019*** | | 0.050*** |
| | (0.261) | | (0.560) | (0.007) | | (0.014) |
| Million Dollar Mystery | | 0.860*** | | | 0.025*** | |
| | | (0.283) | | | (0.007) | |
| Observations | 79,754 | 79,754 | 79,754 | 79,754 | 79,754 | 79,754 |
| R-squared | 0.057 | 0.057 | - | 0.068 | 0.068 | - |
| 1st stage F-Stat | - | - | 103 | - | - | 103 |
| Dep. Var. Mean | 43.85 | 43.85 | 43.85 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 17.98 | 17.98 | 17.98 | 17.98 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). With respect to the baseline (where all states and former US territories that did not belong to the Confederacy were defined as Unionist states), the sample here is restricted to individuals from former Confederate and former Unionist states. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## D.4 Accounting for Migration Flows

This section performs additional exercises to verify our results not to be driven by migration flows and population movements that affected the country in the first decades of the 20th century. Name prevalence at a specific point in time might change as a consequence of a large inflow of individuals from other areas. To flexibly account for this possibility, in Table D8, we replicate baseline results accounting for the fixed effects of the state of birth of the individual and of the head of her/his family.

Changes in naming patterns might also follow from a higher frequency of marriages with individuals coming from other areas of the country. Results presented in Table D9 are obtained with a sample that excludes families where the head of the family was born in the opposite group (Confederate states versus all other states).

We also replicate our analysis using individuals observed in the 1920 Census with the idea that the shorter is the distance between birthyear and Census year, the lower is the probability that the reported county of residence is different from the county of birth (Table D10).

Table D8: Accounting for Birth States FE of the Respondent and the Head of the Family

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.665*** | | 1.531*** | 0.017*** | | 0.043*** |
| | (0.221) | | (0.455) | (0.007) | | (0.012) |
| Million Dollar Mystery | | 0.802*** | | | 0.023*** | |
| | | (0.238) | | | (0.006) | |
| Observations | 93,187 | 93,187 | 93,187 | 93,187 | 93,187 | 93,187 |
| R-squared | 0.084 | 0.084 | - | 0.078 | 0.078 | - |
| 1st stage F-Stat | - | - | 145 | - | - | 145 |
| Dep. Var. Mean | 44.06 | 44.06 | 44.06 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.82 | 16.82 | 16.82 | 16.82 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender, (year of birth $\times$ census year) and (state of birth of the individual $\times$ state of birth of head of her/his family) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D9: Excluding Individuals with Head of the Family from the Opposite Group

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.672*** | | 1.685*** | 0.019*** | | 0.051*** |
| | (0.227) | | (0.476) | (0.006) | | (0.013) |
| Million Dollar Mystery | | 0.875*** | | | 0.026*** | |
| | | (0.243) | | | (0.007) | |
| Observations | 90,445 | 90,445 | 90,445 | 90,445 | 90,445 | 90,445 |
| R-squared | 0.068 | 0.068 | -0.000 | 0.063 | 0.063 | -0.000 |
| 1st stage F-Stat | - | - | 140 | - | - | 140 |
| Dep. Var. Mean | 43.84 | 43.84 | 43.84 | .35 | .35 | .35 |
| Dep. Var. Std. Dev. | 16.72 | 16.72 | 16.72 | 16.72 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. With respect to the baseline, we exclude from the sample individuals where the head of the family was born in the opposite group. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D10: Using Individuals from Census 1920 Only

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.591** | | 2.031*** | 0.018** | | 0.043** |
| | (0.297) | | (0.599) | (0.009) | | (0.0172) |
| Million Dollar Mystery | | 1.060*** | | | 0.022** | |
| | | (0.308) | | | (0.009) | |
| Observations | 48,610 | 48,610 | 48,610 | 48,610 | 48,610 | 48,610 |
| R-squared | 0.081 | 0.081 | - | 0.076 | 0.076 | - |
| 1st stage F-Stat | - | - | 149 | - | - | 149 |
| Dep. Var. Mean | 44.12 | 44.12 | 44.12 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.57 | 16.57 | 16.57 | 16.57 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 Census. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and year of birth fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## D.5 Falsification Exercises

### D.5.1 Falsification Exercise - Black Individuals

According to our working hypothesis, *The Birth of a Nation* fostered reconciliation between the white North and the white South, creating a common threat: the emancipation of African Americans. The movie is expected, therefore, to push for reconciliation of the white North and the white South, but this cultural reconciliation fostered by the movie should not affect black communities of the two regions, which were on the same side of the conflict during the Civil War. Following up on this intuition, this section presents a falsification exercise where we replicate our main empirical specification looking at black individuals (born in the years 1910-1919). Tables D11 and D12 display results obtained using as dependent variable the $ENI$, Enemy Name Index, indicator computed using black individuals.

In Table D11, the index is constructed using the frequency of names of blacks in former Confederate and non-Confederate counties. In other terms, we will be looking at whether exposure to *The Birth of a Nation* increases the likelihood of a baby born in a former Confederate counties to be named with a first name popular among blacks in former non-Confederate counties (and *vice-versa*). In Table D12 we replicate the same exercise constructing the index using the frequency of names of whites in former Confederate and non-Confederate counties. In other terms, we will be looking at whether newborn black babies from former Confederate counties are more likely to be named with first names popular among whites of former non-Confederate counties (and *vice-versa*).

As expected, estimates show no effect of cultural changes affecting black communities of the two regions as a consequence of exposure to *The Birth of a Nation*. These results provide further support to the validity of our empirical approach, as they indicate that our findings on white individuals are not mechanically capturing a broader cultural convergence between the two regions.

Table D11: Falsification Exercise on Black Newborns - ENI based on Black Individuals

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.0413 | | -0.873 | 0.010 | | -0.029 |
| | (0.483) | | (0.950) | (0.015) | | (0.030) |
| Million Dollar Mystery | | -0.449 | | | -0.015 | |
| | | (0.483) | | | (0.015) | |
| Observations | 16,633 | 16,633 | 16,633 | 16,633 | 16,633 | 16,633 |
| R-squared | 0.052 | 0.052 | - | 0.042 | 0.042 | - |
| 1st stage F-Stat | - | - | 104 | - | - | 104 |
| Dep. Var. Mean | 45.54 | 45.54 | 45.54 | .42 | .42 | .42 |
| Dep. Var. Std. Dev. | 16.45 | 16.45 | 16.45 | 16.45 | .49 | .49 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (ENI) of individual $i$ in Columns (1) to (3), and the binarized version of the ENI in Columns (4) to (6) (see Section 6.1 for further details). With respect to the baseline table, the ENI here is computed using black individuals only. The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all black native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and (year of birth $\times$ census year) effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D12: Falsification Exercise on Black Newborns - ENI based on White Individuals

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | -0.483 | | -0.378 | -0.022 | | -0.014 |
| | (0.476) | | (0.967) | (0.016) | | (0.031) |
| Million Dollar Mystery | | -0.194 | | | -0.007 | |
| | | (0.496) | | | (0.016) | |
| Observations | 14,105 | 14,105 | 14,105 | 14,105 | 14,105 | 14,105 |
| R-squared | 0.091 | 0.091 | - | 0.093 | 0.093 | - |
| 1st stage F-Stat | - | - | 98 | - | - | 98 |
| Dep. Var. Mean | 49.24 | 49.24 | 49.24 | .53 | .53 | .53 |
| Dep. Var. Std. Dev. | 15.47 | 15.47 | 15.47 | 15.47 | .5 | .5 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (ENI) of individual $i$ in Columns (1) to (3), and the binarized version of the ENI in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all black native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and (year of birth $\times$ census year) effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### D.5.2 Falsification Exercise - Older Cohorts

This section presents evidence providing support to our empirical strategy by means of a falsification exercise constructed fictitiously modifying the treatment time. More precisely, in order to see if our explanatory variable of interest is capturing a pre-existing trend in naming patterns affecting counties where the movies are more likely to be screened, we propose a falsification exercise where we fictitiously anticipate the release of the movies, *The Birth of a Nation* and *Million Dollar Mystery* (transposed), by five, ten, twenty and thirty years. Results, summarized in Tables D13, D14, D15, and D16 show that no pre-existing trend in naming patters is confounding our results, as we detect no positive effect of exposure to Birth of a Nation on the enemy name index as a consequence of the "fake" treatment.

Table D13: Falsification Exercise - Cohorts 1905-1915

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | -0.397* | | -0.551 | -0.009 | | -0.012 |
| | (0.225) | | (0.431) | (0.006) | | (0.012) |
| Million Dollar Mystery | | -0.295 | | | -0.007 | |
| | | (0.232) | | | (0.007) | |
| Observations | 86,125 | 86,125 | 86,125 | 86,125 | 86,125 | 86,125 |
| R-squared | 0.078 | 0.078 | - | 0.073 | 0.073 | - |
| 1st stage F-Stat | - | - | 164 | - | - | 164 |
| Dep. Var. Mean | 43.55 | 43.55 | 43.55 | .34 | .34 | .34 |
| Dep. Var. Std. Dev. | 16.46 | 16.46 | 16.46 | 16.46 | .47 | .47 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1905-1915] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We fictitiously anticipate the release of the movies, The Birth of a Nation and Million Dollar Mystery (transposed), by five years. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D14: Falsification Exercise - Cohorts 1900-1910

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | -0.205 | | -0.955** | -0.001 | | -0.019 |
| | (0.214) | | (0.402) | (0.007) | | (0.016) |
| Million Dollar Mystery | | -0.510** | | | -0.0101 | |
| | | (0.212) | | | (0.007) | |
| Observations | 83,495 | 83,495 | 83,495 | 83,495 | 83,495 | 83,495 |
| R-squared | 0.073 | 0.073 | - | 0.068 | 0.068 | - |
| 1st stage F-Stat | - | - | 163 | - | - | 163 |
| Dep. Var. Mean | 43.23 | 43.23 | 43.23 | .33 | .33 | .33 |
| Dep. Var. Std. Dev. | 16.05 | 16.05 | 16.05 | 16.05 | .47 | .47 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1900-1910] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We fictitiously anticipate the release of the movies, The Birth of a Nation and Million Dollar Mystery (transposed), by ten years. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D15: Falsification Exercise - Cohorts 1890-1900

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | -0.163 | | -0.411 | -0.006 | | -0.012 |
| | (0.135) | | (0.264) | (0.004) | | (0.008) |
| Million Dollar Mystery | | -0.216 | | | -0.006 | |
| | | (0.138) | | | (0.004) | |
| Observations | 237,493 | 237,493 | 237,493 | 237,493 | 237,493 | 237,493 |
| R-squared | 0.049 | 0.049 | - | 0.041 | 0.041 | - |
| 1st stage F-Stat | - | - | 205 | - | - | 205 |
| Dep. Var. Mean | 43.44 | 43.44 | 43.44 | .34 | .34 | .34 |
| Dep. Var. Std. Dev. | 15.47 | 15.47 | 15.47 | 15.47 | .47 | .47 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1890-1900] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We fictitiously anticipate the release of the movies, The Birth of a Nation and Million Dollar Mystery (transposed), by twenty years. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table D16: Falsification Exercise - Cohorts 1880-1900

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | -0.085 | | -0.414 | -0.005 | | -0.004 |
| | (0.133) | | (0.267) | (0.004) | | (0.008) |
| Million Dollar Mystery | | -0.219 | | | -0.002 | |
| | | (0.141) | | | (0.004) | |
| Observations | 206,166 | 206,166 | 206,166 | 206,166 | 206,166 | 206,166 |
| R-squared | 0.044 | 0.044 | - | 0.032 | 0.032 | - |
| 1st stage F-Stat | - | - | 221 | - | - | 221 |
| Dep. Var. Mean | 44.18 | 44.18 | 44.18 | .35 | .35 | .35 |
| Dep. Var. Std. Dev. | 14.97 | 14.97 | 14.97 | 14.97 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1880-1890] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We fictitiously anticipate the release of the movies, The Birth of a Nation and Million Dollar Mystery (transposed), by thirty years. All regressions control for county, gender and (year of birth $\times$ census year) fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### D.5.3 Accounting for Spatial Clustering in the Errors

We replicate our baseline table accounting for potential spatial correlation in the errors. Table D17 presents results with standard errors adjusted for two-dimensional spatial dependence, using cut-off thresholds of 50, 100, and 200 km.

Table D17: Birth of a Nation and Naming Patterns

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.687*** | | 1.635*** | 0.018** | | 0.047*** |
| *Spatial Cluster 50 Km* | (0.202) | | (0.409) | (0.006) | | (0.012) |
| *Spatial Cluster 100 Km* | (0.201) | | (0.409) | (0.006) | | (0.012) |
| *Spatial Cluster 200 Km* | (0.198) | | (0.411) | (0.006) | | (0.012) |
| Million Dollar Mystery | | 0.854*** | | | 0.024*** | |
| *Spatial Cluster 50 Km* | | (0.213) | | | (0.006) | |
| *Spatial Cluster 100 Km* | | (0.214) | | | (0.006) | |
| *Spatial Cluster 200 Km* | | (0.214) | | | (0.006) | |
| Observations | 95,040 | 85,645 | 85,645 | 85,645 | 85,645 | 85,645 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, gender and (year of birth $\times$ census year) effects. The table reports standard errors adjusted for two-dimensional spatial dependence using cutoff thresholds of 50, 100, and 200km. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# E White Reconciliation and Racial Discrimination

## E.1 BoaN Exposure, Alienation and Discrimination

### E.1.1 Alternative Measures of Supremacism

Our baseline measure of $\texttt{Supremacism}_{c,t}$ attempts to proxy racial nationalism and, more broadly, the salience of race and whiteness in the public discourse on national identity. To do so, it targets newspaper pages referring to "White Americans". As a way to capture other dimensions of supremacism, we further look at the frequency of references to "True Americans" and "Aryan(s)". These keyword are likely to proxy related but slightly different shades of supremacism. "True Americans" might proxy for racial nationalism but also relate to patriotic rhetoric more generally. The frequency of references to "Aryan(s)" , is interesting because it is also explicitly mentioned in the script of the movie The Birth of a Nation. Results indicate that exposure to the movie affect all these different dimensions of supremacism.

Table E1: BoaN and Racial Nationalism - Alternative Measures

| | Aryans | | | True Americans | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.018*** | | 0.046*** | 0.195*** | | 0.723*** |
| | (0.005) | | (0.015) | (0.015) | | (0.06) |
| Million Dollar Mystery | | 0.015*** | | | 0.23*** | |
| | | (0.005) | | | (0.015) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.955 | 0.955 | - | 0.855 | 0.856 | - |
| 1st stage F-Stat | - | - | 166.244 | - | - | 166.244 |
| Sample Mean | -4.584 | -4.584 | -4.584 | -4.316 | -4.316 | -4.316 |
| Sample SD | 1.197 | 1.197 | 1.197 | 1.072 | 1.072 | 1.072 |

NOTE: This table reports OLS (Columns 1, and 4), reduced form (Columns 2, and 5) and 2SLS (Columns 3, and 6) estimates. The dependent variable is the log frequency of *Aryans* in Columns (1) to (3), and the log frequency of *True Americans* in Columns (4) to (6). The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### E.1.2 Alternative Measures of Racial Discrimination

In our baseline analysis, we measure racial discrimination by collecting newspaper ads addressed to "White(s) Only". In practice, newspaper pages containing "White Only" refer to job ads, housing ads and more generally to public events reserved to whites. As a consequence, our baseline measure captures racial discrimination defined in a broad sense. The keywords "White Only" appearing on newspapers, however, might also refer to ads that have nothing to do with racial discrimination. The implicit assumption we are making is that the presence of ads unrelated to racial discrimination containing "White Only", which we erroneously categorize as racial discrimination, are not affected by exposure to *The Birth of a Nation*. As a consequence, this type of measurement error should increase the noise in the estimation but leave the coefficient unaffected. Another equally valid approach is to choose a set of keywords that are more targeted to identify specific forms of racial discrimination, such as labor market discrimination. We verify our results with this alternative approach and replicate our analysis searching for newspaper pages containing the keywords "White Men Only" or "White Only Wanted". Note that the average number of pages that we retrieve is lower, by construction. Coefficient sizes also differ depending on the keywords selected, the direction of the effect is however always the same and consistently in line with our baseline results.

Table E2: BoaN and Racial Discrimination - Alternative Measures

| | White Men Only | | | White Only Wanted | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation | 0.006** | | 0.018** | 0.046*** | | 0.162*** |
| | (0.003) | | (0.009) | (0.009) | | (0.035) |
| Million Dollar Mystery | | 0.006** | | | 0.052*** | |
| | | (0.003) | | | (0.011) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.992 | 0.992 | - | 0.901 | 0.901 | - |
| 1st stage F-Stat | - | - | 166.244 | - | - | 166.244 |
| Dep. Var. Mean | -4.668 | -4.668 | -4.668 | -4.454 | -4.454 | -4.454 |
| Dep. Var. Std. Dev. | 1.271 | 1.271 | 1.271 | 1.079 | 1.079 | 1.079 |

NOTE: This table reports OLS (Columns 1, and 4), reduced form (Columns 2, and 5) and 2SLS (Columns 3, and 6) estimates. The dependent variable is the log frequency of *White men only* in Columns (1) to (3), and the log frequency of *White only wanted* in Columns (4) to (6). The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

### E.1.3 Alternative Measures of BoaN Screenings

As usual, we replicate our findings with alternative measures of exposure to *The Birth of a Nation*. Table E3 replicates baseline results with a threshold of 3, and Table E4 with a threshold of 5.

Table E3: BoaN and Alienation of African Americans - 3 Ads as Thresholds

| | Supremacism | | | Discrimination | | | Alienation | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.035*** | | 0.097*** | 0.074*** | | 0.212*** | 0.037*** | | 0.104*** |
| | (0.006) | | (0.018) | (0.013) | | (0.041) | (0.005) | | (0.017) |
| Million Dollar Mystery | | 0.033*** | | | 0.072*** | | | 0.035*** | |
| | | (0.006) | | | (0.014) | | | (0.006) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.951 | 0.951 | - | 0.808 | 0.808 | - | 0.934 | 0.934 | - |
| Dep. Var. Mean | -4.576 | -4.576 | -4.576 | -4.247 | -4.247 | -4.247 | -.76 | -.76 | -.76 |
| Dep. Var. Std. Dev. | 1.178 | 1.178 | 1.178 | .971 | .971 | .971 | .671 | .671 | .671 |

NOTE: This table reports OLS (Columns 1, 4 and 7), reduced form (Columns 2, 5 and 8) and 2SLS (Columns 3, 6 and 9) estimates. The dependent variable is the salience of whiteness in the public discourse ($\texttt{Supremacism}_{c,t}$) in Columns (1) to (3), racial discrimination ($\texttt{Discrimination}_{c,t}$) in Columns (4) to (6) and the first principal component of $\texttt{Supremacism}_{c,t}$ and $\texttt{Discrimination}_{c,t}$, in Columns (7) to (9) (see Section 7 for further details). The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect at least 3 screening records (the threshold is 2 in the baseline specification) All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E4: BoaN and Alienation of African Americans - 5 Ads as Thresholds

| | Supremacism | | | Discrimination | | | Alienation | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation | 0.039*** | | 0.096*** | 0.093*** | | 0.22*** | 0.045*** | | 0.107*** |
| | (0.007) | | (0.017) | (0.015) | | (0.039) | (0.006) | | (0.016) |
| Million Dollar Mystery | | 0.035*** | | | 0.079*** | | | 0.039*** | |
| | | (0.006) | | | (0.014) | | | (0.006) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.951 | 0.951 | - | 0.808 | 0.808 | - | 0.934 | 0.934 | - |
| Dep. Var. Mean | -4.576 | -4.576 | -4.576 | -4.247 | -4.247 | -4.247 | -.76 | -.76 | -.76 |
| Dep. Var. Std. Dev. | 1.178 | 1.178 | 1.178 | .971 | .971 | .971 | .671 | .671 | .671 |

NOTE: This table reports OLS (Columns 1, 4 and 7), reduced form (Columns 2, 5 and 8) and 2SLS (Columns 3, 6 and 9) estimates. The dependent variable is the salience of whiteness in the public discourse ($\texttt{Supremacism}_{c,t}$) in Columns (1) to (3), racial discrimination ($\texttt{Discrimination}_{c,t}$) in Columns (4) to (6) and the first principal component of $\texttt{Supremacism}_{c,t}$ and $\texttt{Discrimination}_{c,t}$, in Columns (7) to (9) (see Section 7 for further details). The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). We consider a movie as screened in the county if we detect at least 5 screening records (the threshold is 2 in the baseline specification) All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

## E.2 Alienation and Reconciliation

Section E.2.1 present the econometric approaches used to quantify the mediated effect of alienation on reconciliation. Section E.2.2 presents OLS and reduced form estimates that complement results reported in the text.

### E.2.1 Alienation and Reconciliation: Mediation Analysis Approach

We assume that the true model of reconciliation is:

$$\texttt{REC}_{ct} = \beta_d \times \texttt{BoaN}_{ct} + \beta_m \times \texttt{ALI}_{ct} + \sum_c \texttt{FE}_c + \sum_t \texttt{FE}_t + \epsilon_{ct}. \tag{11}$$

where the mediating effect of alienation is captured by $\beta_m$.

As for alienation, the true model is:

$$\texttt{ALI}_{ct} = \beta_{\texttt{ALI}} \times \texttt{BoaN}_{ct} + \gamma_{\texttt{ALI}} \times \texttt{BoaN}_{ct} \times \texttt{Black}_c + \sum_c \texttt{FE}_c + \sum_t \texttt{FE}_t + \epsilon_{ct}. \tag{12}$$

where the impact on the treatment of the exogenous shifter $\texttt{Black}_c$ is captured by the coefficient $\gamma_{\texttt{ALI}}$. In virtue of the exclusion restriction $[H_2]$, the shifter has no such effect on $\texttt{REC}_{ct}$. Note that, in this setup, $\texttt{Black}_c$ can have a direct (linear) effect on both outcome variables and this influence is absorbed by the county-fixed effect.

Combining the two equations leads to the empirical model (6):

$$\texttt{REC}_{ct} = (\beta_d + \beta_m\beta_{\texttt{ALI}}) \times \texttt{BoaN}_{ct} + \beta_m\gamma_{\texttt{ALI}} \times \texttt{BoaN}_{ct} \times \texttt{Black}_c + \sum_c \texttt{FE}_c + \sum_t \texttt{FE}_t + \epsilon_{ct}. \tag{13}$$

Mapping the relationship with the coefficients recovered from the estimation yields:

$$\hat{\beta}_{\texttt{REC}} = \beta_d + \beta_m\hat{\beta}_{\texttt{ALI}} \tag{14}$$

$$\hat{\gamma}_{\texttt{REC}} = \beta_m\hat{\gamma}_{\texttt{ALI}} \tag{15}$$

Rearranging the terms enables to quantify the coefficients of the direct and indirect effects:

$$\beta_d = \hat{\beta}_{\texttt{REC}} - \frac{\hat{\gamma}_{\texttt{REC}}}{\hat{\gamma}_{\texttt{ALI}}}\hat{\beta}_{\texttt{ALI}} \tag{16}$$

$$\beta_m = \frac{\hat{\gamma}_{\texttt{REC}}}{\hat{\gamma}_{\texttt{ALI}}} \tag{17}$$

In counties where $\texttt{Black}_c = 0$, visual inspection of equation (13) indicates that the total treatment effect is equal to $\beta_d + \beta_m\beta_{\texttt{ALI}}$. Hence, the contribution of the mediated effect to the total treatment effect is equal to:

$$\frac{\beta_m\beta_{\texttt{ALI}}}{\beta_d + \beta_m\beta_{\texttt{ALI}}} = \frac{\hat{\gamma}_{\texttt{REC}}}{\hat{\gamma}_{\texttt{ALI}}} \times \frac{\hat{\beta}_{\texttt{ALI}}}{\hat{\beta}_{\texttt{REC}}} \tag{18}$$

In counties where $\texttt{Black}_c = 1$, the contribution is equal to:

$$\frac{\beta_m\beta_{\texttt{ALI}} + \beta_m\gamma_{\texttt{ALI}}}{\beta_d + \beta_m\beta_{\texttt{ALI}} + \beta_m\gamma_{\texttt{ALI}}} = \frac{\hat{\gamma}_{\texttt{REC}}}{\hat{\gamma}_{\texttt{ALI}}} \times \frac{\hat{\beta}_{\texttt{ALI}} + \hat{\gamma}_{\texttt{ALI}}}{\hat{\beta}_{\texttt{REC}} + \hat{\gamma}_{\texttt{REC}}} \tag{19}$$

Finally, if we consider the baseline sample with both types of counties, the contribution of the

mediated effect to the total treatment effect becomes:

$$\frac{\beta_m \beta_{\texttt{ALI}} + \beta_m \gamma_{\texttt{ALI}} \bar{\texttt{Black}}}{\beta_d + \beta_m \beta_{\texttt{ALI}} + \beta_m \gamma_{\texttt{ALI}} \bar{\texttt{Black}}} = \frac{\hat{\gamma}_{\texttt{REC}}}{\hat{\gamma}_{\texttt{ALI}}} \times \frac{\hat{\beta}_{\texttt{ALI}} + \hat{\gamma}_{\texttt{ALI}} \bar{\texttt{Black}}}{\hat{\beta}_{\texttt{REC}} + \hat{\gamma}_{\texttt{REC}} \bar{\texttt{Black}}} \tag{20}$$

where $\bar{\texttt{Black}}$ is the sample mean of the shifter.

### E.2.2 Mediation Analysis Approach: OLS and Reduced Form

This section reports OLS (Table E5) and reduced form (Table E6) evidence that complement the mediation analysis presented in the text.

Table E5: Alienation as a Mean to Reconciliation - OLS Estimates

| | Full Sample | | Free States | | Low Int. Civil War | |
|---|---|---|---|---|---|---|
| | (1) Alien. | (2) Reconc. | (3) Alien. | (4) Reconc. | (5) Alien. | (6) Reconc. |
| Birth of a Nation | 0.011* | 0.100*** | 0.000 | 0.027 | -0.007 | 0.055 |
| | (0.005) | (0.020) | (0.006) | (0.025) | (0.008) | (0.036) |
| Birth of a Nation * Black | 0.029*** | 0.092*** | 0.050*** | 0.129*** | 0.044*** | 0.180*** |
| | (0.009) | (0.024) | (0.009) | (0.029) | (0.013) | (0.046) |
| Observations | 102,772 | 102,772 | 63,899 | 63,899 | 36,441 | 36,441 |
| R-squared | 0.934 | 0.802 | 0.917 | 0.798 | 0.933 | 0.802 |
| Mediation Effect (% Total) | | 59.69 | | 77.25 | | 51.22 |

NOTE: The table reports the OLS estimates. The dependent variable is $\texttt{Supremacism}_{c,t}$ in Columns (1), (5) and (7), and $\texttt{Reconciliation}_{c,t}$ in Columns (2), (6) and (8). See Section 7 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all baseline counties in Columns (1) and (2), only counties located within free states in Columns (3) and (4), and only counties with a lower than median share of Civil War casualties in Columns (5) and (6). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). The variable $\texttt{Black}_c$ is an indicator variable taking a value of 1 for counties with a population share of African Americans above the 33rd percentile. All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E6: Alienation as a Mean to Reconciliation - Reduced Form Estimates

| | Full Sample | | Free States | | Low Int. Civil War | |
|---|---|---|---|---|---|---|
| | (1) Alien. | (2) Reconc. | (3) Alien. | (4) Reconc. | (5) Alien. | (6) Reconc. |
| Million Dollar Mystery | 0.017** | 0.140*** | 0.013 | 0.124*** | 0.000 | 0.060 |
| | (0.007) | (0.022) | (0.008) | (0.028) | (0.017) | (0.049) |
| Million Dollar Mystery * Black | 0.026*** | 0.079*** | 0.035*** | 0.088*** | 0.0437*** | 0.146*** |
| | (0.010) | (0.024) | (0.010) | (0.030) | (0.014) | (0.052) |
| Observations | 102,772 | 102,772 | 63,899 | 63,899 | 36,441 | 36,441 |
| R-squared | 0.934 | 0.802 | 0.917 | 0.799 | 0.933 | 0.801 |
| Mediation Effect (% Total) | | 53.29 | | 49.8 | | 61.61 |

NOTE: The table reports the RF estimates. The dependent variable is $\texttt{Supremacism}_{c,t}$ in Columns (1), (5) and (7), and $\texttt{Reconciliation}_{c,t}$ in Columns (2), (6) and (8). See Section 7 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all baseline counties in Columns (1) and (2), only counties located within free states in Columns (3) and (4), and only counties with a lower than median share of Civil War casualties in Columns (5) and (6). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). The variable $\texttt{Black}_c$ is an indicator variable taking a value of 1 for counties with a population share of African Americans above the 33rd percentile. All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# F A Pervasive Reconciliation

## F.1 Heterogeneous Effects in Confederate and Unionist States

In this section we explore whether the changes in attitudes and behaviors that *the Birth of a Nation* engendered in former Confederate states and former Unionist states are similar in size. The following tables presents two different coefficients that capture the effects of *the Birth of a Nation* in former Unionist and Confederate counties, estimated as per equation 7.

Table F1 replicates baseline results on $\texttt{Reconciliation}_{c,t}$, separating the effect between former Confederate states and former Union states. Table F2, Table F3, and F4 replicate the same exercise with Navy enlistment, naming patterns, supremacism and racial segregation as outcomes.

Table F1: Reconciliation Rhetoric in Former Confederate States and Former Union States

| Type of Rhetoric: | Reconciliation | | | Patriotic | | | Divisive | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation * Unionist | 0.164*** | | 0.605*** | 0.099*** | | 0.351*** | -0.065*** | | -0.254*** |
| | (0.017) | | (0.058) | (0.019) | | (0.062) | (0.01) | | (0.035) |
| Birth of a Nation * Confederate | 0.141*** | | 0.538*** | 0.091** | | 0.266*** | -0.049** | | -0.272*** |
| | (0.025) | | (0.064) | (0.038) | | (0.07) | (0.023) | | (0.046) |
| Million Dollar Mystery * Unionist | | 0.202*** | | | 0.123*** | | | -0.08*** | |
| | | (0.017) | | | (0.02) | | | (0.011) | |
| Million Dollar Mystery * Confederate | | 0.150*** | | | 0.062** | | | -0.088*** | |
| | | (0.023) | | | (0.027) | | | (0.018) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.801 | 0.802 | - | 0.933 | 0.933 | - | 0.981 | 0.981 | - |
| 1st stage F-Stat BoaN * Unionist | - | - | 168.3 | - | - | 168.3 | - | - | 168.3 |
| 1st stage F-Stat BoaN * Confederate | - | - | 160.1 | - | - | 160.1 | - | - | 160.1 |
| Dep. Var. Mean | .122 | .122 | .122 | -1.706 | -1.706 | -1.706 | -1.828 | -1.828 | -1.828 |
| Dep. Var. Std. Dev. | .788 | .788 | .788 | 1.421 | 1.421 | 1.421 | 1.68 | 1.68 | 1.68 |

NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\texttt{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\texttt{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\texttt{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county ($c$) in a particular month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. Unionist is an indicator variable that takes value 1 if the county belongs to a former Union state. Confederate is an indicator variable that takes value 1 if the county belongs to a former Confederate state. All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F2: Enlistment in Former Confederate States and Former Union States

| | Navy Enlistments | | | | | |
|---|---|---|---|---|---|---|
| | Jan 1913 - Nov 1918 | | | Jan 1913 - Mar 1917 | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation x Unionist | 0.042*** | | 0.076*** | 0.015*** | | 0.028*** |
| | (0.006) | | (0.015) | (0.004) | | (0.009) |
| Birth of a Nation x Confederate | 0.004 | | 0.023** | 0.003 | | 0.013 |
| | (0.006) | | (0.01) | (0.004) | | (0.008) |
| Million Dollar Mystery x Unionist | | 0.031*** | | | 0.009*** | |
| | | (0.007) | | | (0.003) | |
| Million Dollar Mystery x Confederate | | -0.001 | | | 0.001 | |
| | | (0.004) | | | (0.002) | |
| Observations | 73,059 | 73,059 | 73,059 | 52,479 | 52,479 | 52,479 |
| R-squared | 0.979 | 0.978 | 0.002 | 0.995 | 0.995 | 0.000 |
| 1st stage F-Stat BoaN * Unionist | - | - | 309 | - | - | 261.3 |
| 1st stage F-Stat BoaN * Confederate | - | - | 205.4 | - | - | 155.3 |
| Dep. Var. Mean | -10.088 | -10.088 | -10.088 | -10.107 | -10.107 | -10.107 |
| Dep. Var. Std. Dev. | .961 | .961 | .961 | .977 | .977 | .977 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5), and 2SLS (Columns 3 and 6) estimates. The dependent variable is the log share of enlistment in county $c$ at month-year $t$. See Section 5.1 for further details. The unit of observation is the county ($c$) in the month-year ($t$). The sample includes all months between January 1913 and November 1918 in Columns (1) to (3), and all months between January 1913 and March 1917 in Columns (4) to (6). Birth of a Nation is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable taking a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later. Unionist is an indicator variable that takes value 1 if the county belongs to a former Union state. Confederate is an indicator variable that takes value 1 if the county belongs to a former Confederate state. All regressions control for county and month-year fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F3: Naming Patterns in Former Confederate States and Former Union States

| | Enemy Name Index | | | | | |
|---|---|---|---|---|---|---|
| | Continuous | | | Dummy | | |
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS |
| Birth of a Nation * Unionist | 0.757*** | | 1.191*** | 0.025*** | | 0.036*** |
| | (0.233) | | (0.379) | (0.006) | | (0.0101) |
| Birth of a Nation * Confederate | -0.174 | | 0.582 | -0.013 | | 0.011 |
| | (0.342) | | (0.698) | (0.008) | | (0.018) |
| Million Dollar Mystery * Unionist | | 0.806*** | | | 0.0250*** | |
| | | (0.246) | | | (0.006) | |
| Million Dollar Mystery * Confederate | | 0.201 | | | 0.002 | |
| | | (0.425) | | | (0.011) | |
| Observations | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 | 95,034 |
| R-squared | 0.069 | 0.069 | - | 0.064 | 0.064 | - |
| 1st stage F-Stat BoaN * Unionist | - | - | 224.5 | - | - | 224.5 |
| 1st stage F-Stat BoaN * Confederate | - | - | 253.3 | - | - | 253.3 |
| Dep. Var. Mean | 44.07 | 44.07 | 44.07 | .36 | .36 | .36 |
| Dep. Var. Std. Dev. | 16.82 | 16.82 | 16.82 | .48 | .48 | .48 |

NOTE: The table reports OLS (Columns 1 and 4), reduced form (Columns 2 and 5) and 2SLS (Columns 3 and 6) estimates. The dependent variable is the Enemy-Sounding Name Index (`ENI`) of individual $i$ in Columns (1) to (3), and the binarized version of the `ENI` in Columns (4) to (6) (see Section 6.1 for further details). The unit of observation is the individual $i$ from county $c(i)$, observed in census $s(i)$ and born in year $y(i) \in [1910 - 1920]$. The sample includes all white native-born individuals born in year $y \in$ [1910-1920] and recorded in the 1920 and 1930 Censuses. Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). Unionist is an indicator variable that takes value 1 if the county belongs to a former Union state. Confederate is an indicator variable that takes value 1 if the county belongs to a former Confederate state. All regressions control for county, gender and (year of birth $\times$ census year) effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table F4: Alienation in Former Confederate States and Former Union States

| | Supremacism | | | Discrimination | | | Alienation | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) OLS | (2) RF | (3) 2SLS | (4) OLS | (5) RF | (6) 2SLS | (7) OLS | (8) RF | (9) 2SLS |
| Birth of a Nation * Unionist | 0.033*** | | 0.098*** | 0.079*** | | 0.223*** | 0.038*** | | 0.108*** |
| | (0.006) | | (0.018) | (0.013) | | (0.042) | (0.005) | | (0.017) |
| Birth of a Nation * Confederate | 0.016 | | 0.09*** | -0.006 | | 0.102* | 0.003 | | 0.064** |
| | (0.01) | | (0.03) | (0.022) | | (0.059) | (0.009) | | (0.026) |
| Million Dollar Mystery * Unionist | | 0.032*** | | | 0.086*** | | | 0.04*** | |
| | | (0.006) | | | (0.015) | | | (0.006) | |
| Million Dollar Mystery * Confederate | | 0.026* | | | 0.003 | | | 0.009 | |
| | | (0.014) | | | (0.026) | | | (0.011) | |
| Observations | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 | 102,772 |
| R-squared | 0.951 | 0.951 | - | 0.808 | 0.808 | - | 0.934 | 0.934 | - |
| 1st stage F-Stat BOAN * Unionist | - | - | 168.3 | - | - | 168.3 | - | - | 168.3 |
| 1st stage F-Stat BOAN * Confederate | - | - | 160.1 | - | - | 160.1 | - | - | 160.1 |
| Dep. Var. Mean | -4.576 | -4.576 | -4.576 | -4.247 | -4.247 | -4.247 | -.76 | -.76 | -.76 |
| Dep. Var. Std. Dev. | 1.178 | 1.178 | 1.178 | .971 | .971 | .971 | .671 | .671 | .671 |

NOTE: This table reports OLS (Columns 1, 4 and 7), reduced form (Columns 2, 5 and 8) and 2SLS (Columns 3, 6 and 9) estimates. The dependent variable is the salience of whiteness in the public discourse ($\texttt{Supremacism}_{c,t}$) in Columns (1) to (3), racial discrimination ($\texttt{Discrimination}_{c,t}$) in Columns (4) to (6) and the first principal component of $\texttt{Supremacism}_{c,t}$ and $\texttt{Discrimination}_{c,t}$, in Columns (7) to (9) (see Section 7 for further details). The unit of observation is the county ($c$) in the month-year ($t$). Birth of a Nation is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Sections 2.3 for details). Million Dollar Mystery is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise, transposed 231 days later (see Section 3.2 for further details). Unionist is an indicator variable that takes value 1 if the county belongs to a former Union state. Confederate is an indicator variable that takes value 1 if the county belongs to a former Confederate state. All regressions control for county, month-year and coverage percentile fixed effects. Standard errors are clustered at the county level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.