

RECONCILIATION NARRATIVES

–“The Birth of a Nation” after the US Civil War–

Elena Esposito* Tiziano Rotesi[†] Alessandro Saia[‡] Mathias Thoenig[§]

ONLINE APPENDIX For Online Publication

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. Section B presents in a compact form all baseline robustness exercises that are common across the main empirical exercises. Additional results and robustness analysis of the effect of exposure to Birth of a Nation on newspaper rhetoric, Navy enlistment, naming patterns and racial discrimination are reported in Sections C, D, E and F, in Section G we explore effects on marriage patterns.

Contents

A Context and Data	IV
A.1 Background	IV
A.2 Measuring Movie Diffusion From Historical Newspapers	VI
A.3 Comparison between ERST and Ang	XII
A.4 Instrumenting the Screening of The Birth of Nation: The Million Dollars Mystery	XVI
A.5 Construction of the Bag-of-Words	XVIII
A.5.1 Seed Words	XVIII
A.5.2 Co-occurring Words	XIX
A.5.3 Semantic Analysis by External Judges.	XXI
A.6 Principal Components Analysis (PCA): Additional Results	XXV
A.7 Outcome Variables: Descriptive Statistics	XXVII
A.7.1 Correlation Between Outcome Variables	XXVII
A.7.2 Time-Series Evolution	XXIX
A.7.3 Geographical Distribution	XXXII
A.7.4 Reconciliation Rhetoric: Lists of keywords	XXXIV
A.7.5 Enlistment in the Navy: Validation of the Measure	XXXVI
A.7.6 Popular Names and ENI	XXXVII

*University of Turin (ESOMAS Department); Collegio Carlo Alberto and CEPR; e.esposito@unito.it.

[†]Department of Economics, University of Lausanne; tiziano.rotes@unil.ch .

[‡]Department of Economics, University of Bologna; alessandro.saia@unibo.it.

[§]Department of Economics, University of Lausanne; University of Oxford; IMD Business School and CEPR; mathias.thoenig@unil.ch.

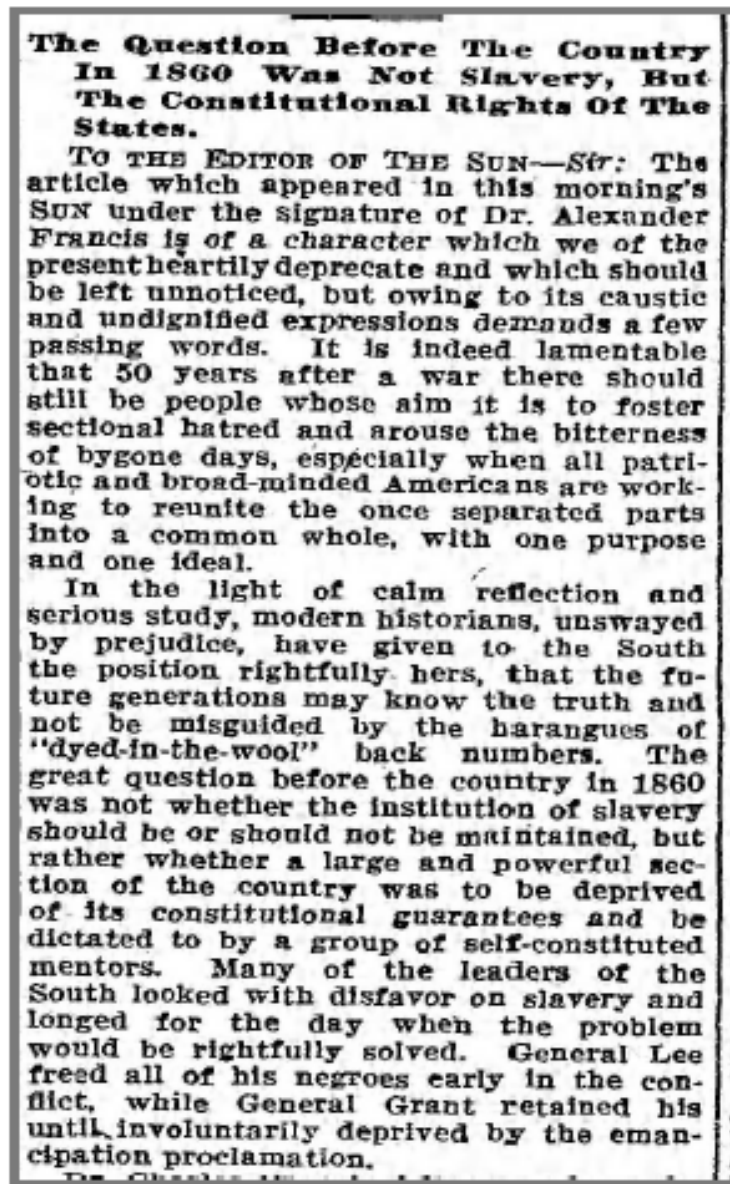
B	Main robustness exercises common across outcomes	XXXIX
B.1	Extended Estimation Sample	XXXIX
B.2	Alternative Measures of Local Screenings of <i>Birth of a Nation</i>	XLIV
B.2.1	Local screening using data from Ang (2020)	XLV
B.2.2	Local screening defined as Union of ERST and Ang	XLVII
B.2.3	Local screening defined as Intersection of ERST and Ang	XLIX
B.3	Alternative Measures of BON Exposure	LI
B.4	War-related Robustness	LVI
B.5	Instrumental Variable Strategy: Other Movies	LXIII
B.6	Instrumental Variable Strategy: Presence of Theaters	LXVIII
B.7	Falsification Exercises	LXXI
B.8	Accounting for Time-Varying Confounders	LXXV
B.9	Accounting for Spatial Correlation in the Errors	LXXVII
B.10	A Pervasive Reconciliation	LXXVIII
B.11	Heterogeneous Effect of BON	LXXIX
B.12	Spatial Spillovers	LXXXIII
C	The Birth of a Nation and the Public Debate: Extended Results	LXXXV
C.1	Alternative Construction of the Dependent Variable	LXXXV
C.2	Alternative Measures of Rhetoric	LXXXVIII
C.2.1	Alternative Lists of Keywords	LXXXVIII
C.2.2	Measurement of Patriotic and Divisive Rhetoric	XCI
C.3	Alternative Sample and Coverage	XCII
C.3.1	Alternative Ways to Control for Newspaper Coverage	XCII
C.3.2	Alternative Sample	XCIV
C.3.3	Word-Based Analysis	XCV
C.3.4	Characterizing Compliers	XCVI
D	Reconciliation and Patriotism: Extended Results	XCVIII
D.1	Alternative Sample	XCVIII
D.2	Enlistment Measured with Deaths of Infectious Disease	XCVIII
E	Reconciliation Culture and Names of the Former Enemies: Extended Results	C
E.1	Robustness Analysis	C
E.1.1	Alternative Measures of Names' Connotation	C
E.1.2	Alternative Sample	CI
E.1.3	Alternative Definitions of BON Exposure	CII
E.2	Collapsed Sample: County-Year Level Analysis	CIII
E.3	Alternative Definitions of Former Enemies	CIV
E.4	Accounting for Migration Flows	CV
E.5	Removing "Birth of a Nation" Characters Names	CVIII
E.6	Accounting for Additional FEs	CX
E.7	Falsification Exercises	CXII
E.7.1	Falsification Exercise - Black Individuals	CXII
E.7.2	Falsification Exercise - Older Cohorts	CXV
F	White Reconciliation and Racial Discrimination: Extended Results	CXIX
F.1	Alternative Sample	CXIX
F.2	Alternative Measures of Supremacism and Discrimination	CXXI
F.3	Alternative Ways to Control for Job Ads Supply	CXXIV

A Context and Data

A.1 Background

Tension between North and South was still present almost 50 years after the end of the Civil War. Figures A1 and A2 show two examples taken from newspapers at the time. Both extracts highlight the role of memory and the importance of slavery in defining the North/South cleavage.

Figure A1: Civil War Memories and North-South Resentment: the "True" Causes of the War



NOTE: Examples of persistent and divisive underlying tensions between North and South emerging on newspapers in the first decades 20th century. The Baltimore Sun, 6 September 1914. Source: newspapers.com

Figure A2: Civil War Memories and North-South Resentment: Sectional Hatred in Public Schools

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.

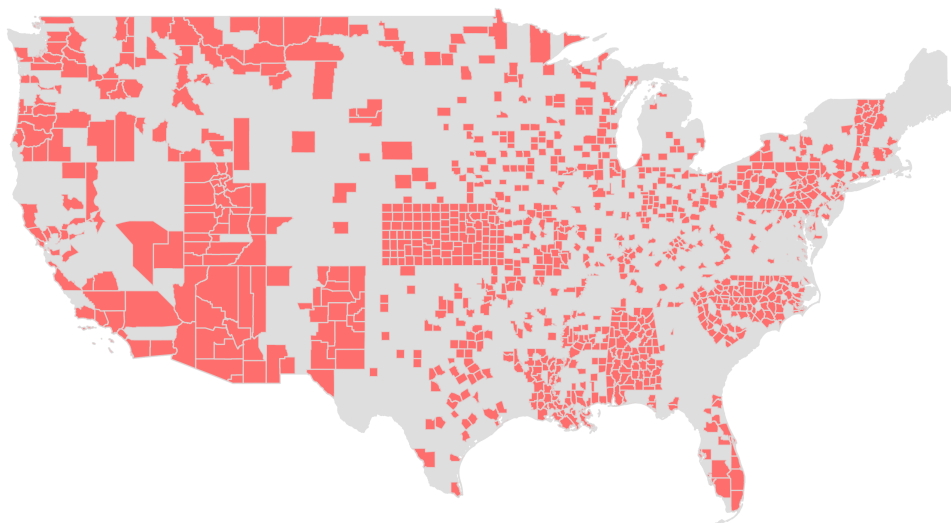
NOTE: Examples of persistent and divisive underlying tensions between North and South emerging on newspapers in the first decades 20th century. The Washington Times, 21 May 1912. Source: newspapers.com

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.¹

Figure A3: Newspapers Headquarters



NOTE: Red 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.

¹These 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, Tusculumbia, 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.

We do this in two ways. First, we only keep counties that had at least some coverage at the time of our initial data collection. The data collection for Birth of a Nation screenings (and for several patriotic, divisive, and neutral words such as "dollar") on *newspapers.com* was performed between the end of 2018 and the summer of 2019. Since *newspapers.com* is continuously digitizing new newspapers' pages, a small subset of counties that were not covered by newspapers at the time of the *first* data collection, i.e. targeting Birth of a Nation screenings, was instead covered by newspapers in later months.² To benchmark the sample of counties to our baseline and starting data collection of Birth of a Nation screenings, we impose a condition that fixes the sample to counties with newspapers' coverage available on the archive at the time of our first data collection (i.e. end of 2018 - summer of 2019). The condition we impose is to only include counties that had at least one newspaper page containing keywords in the initial set we collected.³ Second, we only keep counties/months that have at least one newspaper page digitized in the month. We identify them by retrieving information on newspaper coverage as provided by *newspapers.com*.⁴ To refine this criterion, we also 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 1,070 counties (965 outside Kansas). 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 56,044 items, from 1,837 newspapers, located across 844 counties. These hits include three main types of items: i) movie ads (see an example on the top left panel of Figure 1); ii) movie time table (see an example on bottom left panel of Figure 1); iii) general articles referring to the movie that are informative about a screening taking place in the county (see two examples on the right panel of Figure 1).

A large share of items retrieved are not related to any specific movie screenings but are general reviews of the movie or articles discussing the reactions to the movie across the country (see an example in Figure A4). It should be noted, on top of this, that the searching algorithm excludes the definite and indefinite articles and some hits might be completely unrelated to the movie. 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 ones in Figure A4 and Figure A5. This process allows us to obtain 14,421 "validated" screening records, from 866 newspapers located in 581 counties. To quantify the judges' precision we then double check 5% of the sample. We find that 90% of items are correctly assigned, with mistakes split between false positives (57% of the mistakes in the double checked sample) and false negatives (43%).

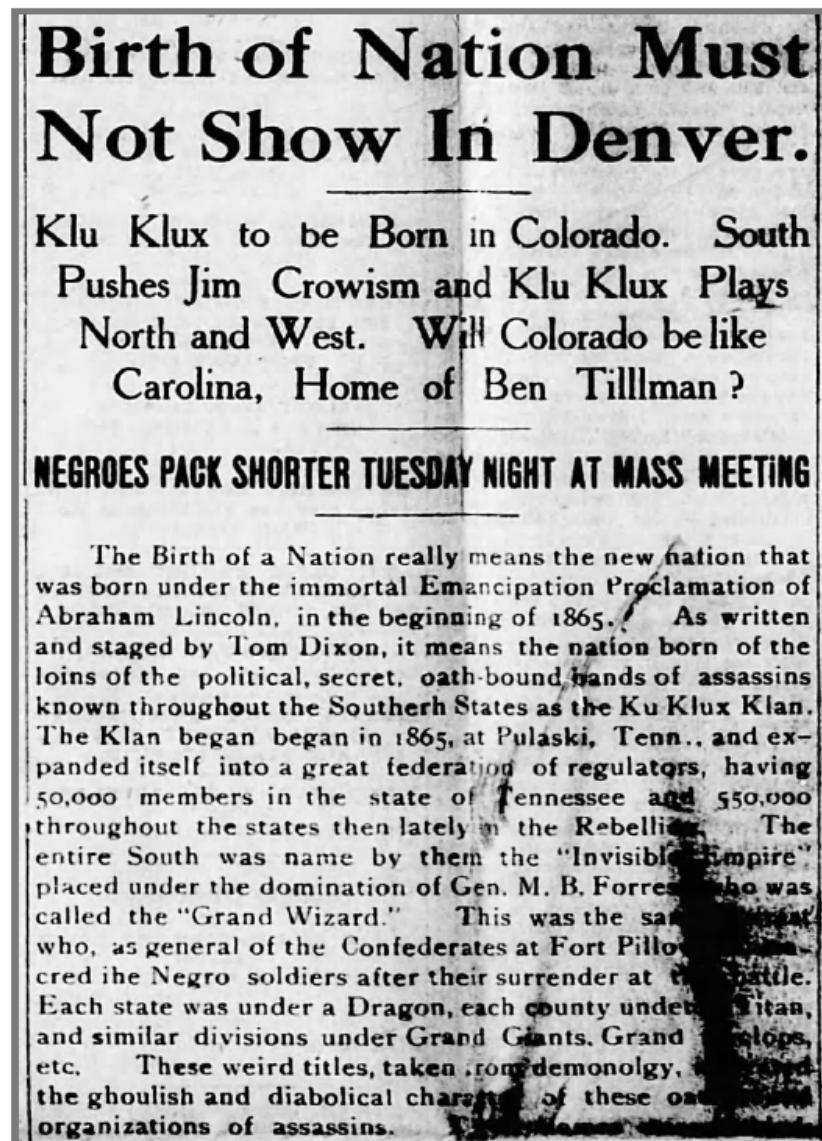
Finally, when considering treated counties as those with at least two verified screening records, we obtain a sample of 440 treated counties and 525 untreated counties (442 and 628 respectively if we include Kansas). Figure A6 displays the time-series evolution of the share of counties that screened *The Birth of a Nation* over time. Table A1 compares treated and untreated counties across demographic characteristics collected in the 1910 Census. Counties reached by the movie by the end of 1920 tend to have a lower share of illiterate population, a lower share of Black population, a

²Note that *newspapers.com* is constantly adding both new journals and new pages of journals that are already present in the archive. Our initial data collections dates back to the end of 2018 - summer of 2019, our final data collection dates back to the spring of 2022.

³This set includes keywords for "The Birth of a Nation", "Million Dollar Mystery", "What Happened to Mary", "Traffic in Souls", and "dollar", which is the most common neutral word we downloaded.

⁴<https://www.newspapers.com/browse/us>, accessed June 11 2022.

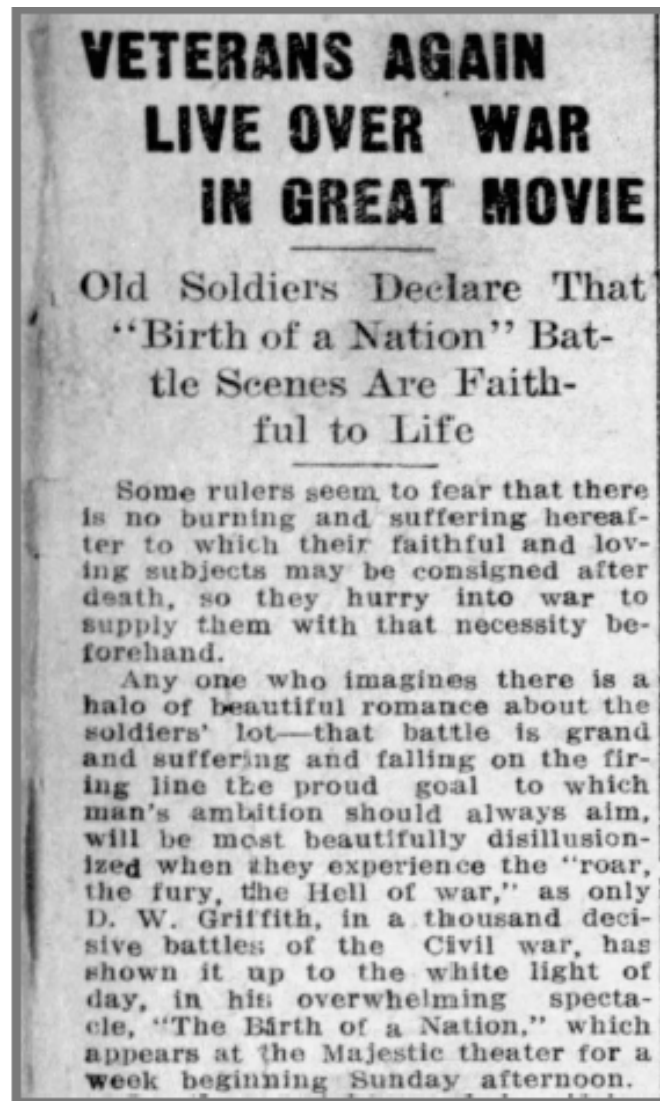
Figure A4: Newspaper Page with *The Birth of a Nation* - A False Positive



NOTE: An article about the movie *The Birth of a Nation*, with no information about screenings in the local theaters. Source: The Indianapolis News, 14 December 1915 [newspapers.com].

higher share of population living in cities and a higher share of foreign population.

Figure A5: Newspaper Page with *The Birth of a Nation* - A True Positive

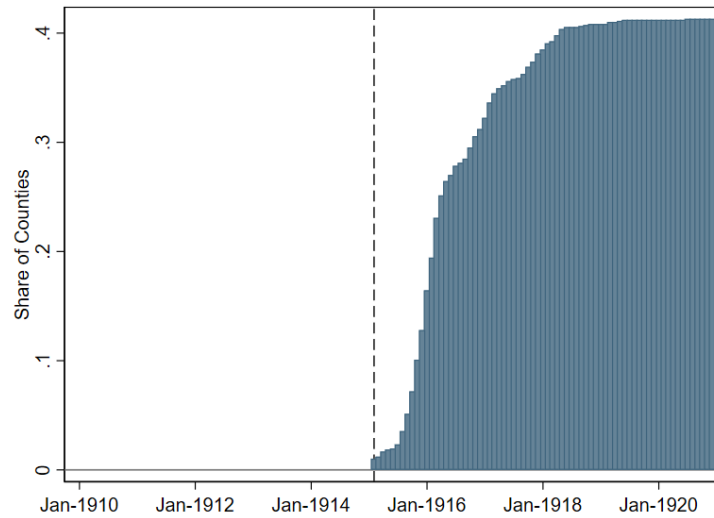


NOTE: An article about the movie *The Birth of a Nation* with information regarding a forthcoming screening of the movie at cinema Majestic. Source: La Crosse Tribune, 12 November 1915 [newspapers.com].

Table A1: Counties with and without BON

Variable	(1) Untreated Counties	(2) Treated Counties	(3) Difference
Share of Illiterate (10+) Population	0.130 (0.118)	0.078 (0.089)	-0.052 (0.007)
Share of Black Population	0.181 (0.236)	0.110 (0.188)	-0.072 (0.014)
Share of Males	0.523 (0.034)	0.521 (0.031)	-0.002 (0.002)
Share of Population in Cities (25k)	0.039 (0.158)	0.148 (0.290)	0.109 (0.015)
Share of Foreign Population	0.060 (0.084)	0.116 (0.104)	0.057 (0.006)
Observations	525	440	965

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 A6: *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 BON_{ct} as defined in the text.

The case of Kansas. Despite the state level ban that was introduced in Kansas, there is evidence of the presence of illegal screenings of the movie. Indeed, local movie owners sometimes decided to screen the movie incurring in legal consequences that were documented in local newspapers. For example, Figure A7 - left, displays evidence of one illegal screening that took place in Topeka in June 1917, violating the ban.

Moreover, the ban had the effect of stimulating advertisements of screening taking place in neighboring counties located on the other side of the state border. Figure A7 - right, presents a movie ad that was published on the Hiawatha Daily World, Brown county (Kansas), and refers to a screening in Falls City (Nebraska), located 17 miles from Hiawatha, therefore in a different county and state.

Figure A7: Screening in Kansas despite the ban and on the other side of the border

EXHIBIT "THE BIRTH OF A NATION"

State Fails to Stop Showing of Film, Altho L. M. and Roy Crawford Are Arrested.

Altho L. M. Crawford and Roy Crawford, theatrical agents, were arrested yesterday afternoon for violation of the motion picture censorship law, "The Birth of a Nation," was shown to a Topeka audience last night.

Attempts by John M. Hunt, assistant attorney general, to secure an injunction from the district court yesterday were blocked by the absence of Judge A. W. Dana and by the refusal of Judge George H. Whitcomb to interrupt a case in his division of the court in order to hear the injunction application.

Hunt refused to state whether he would attempt to secure an injunction today. The producers of "The Birth of a Nation," are acting under the assumption that he will not, and are going ahead with their plans.

The arrest of the Crawfords marks the first case under the provisions of the censorship law, which prohibits the advertising of pictures, permission of which to be shown has not been granted by the censorship board. Ex-Sheriff L. L. Kiene immediately went on their bond. The arrest of the theater proprietors does not prevent the showing of the picture, according to the law.

The Sherman Elliott people claim that under the law they can show the picture thirty days before they have to turn in the picture for re-examination by the censorship board.

Hunt secured a mandamus from the supreme court ordering the company to turn in the film for re-examination.

As yet it is undecided whether Kansas is to have thirty days of "The Birth of a Nation" or only the showing last night.

5000 Scenes
18000 People

D. W. Griffith's 8th Wonder of the World

The Birth of a Nation

Taken from Thomas Dixon's
"THE CLANSMAN"
Presented by Elliott & Sherman

GEHLING THEATRE, FALLS CITY, NEB.
FRIDAY and SATURDAY, FEB. 9 and 10

Twice Daily, 2:15 and 8:15.

Cities built up and then destroyed by fire. Ford's Theatre, Washington, reproduced to the smallest detail for the Lincoln tragedy.

This will be your last chance to see this great American play as it will not be brought here again.

20 Piece Symphony Orchestra 20

The Seating Capacity is Limited. Get Your Tickets at Once.
PRICES 25c, 50c, 75c and \$1.00.

3000 Horses
Cost \$500,000

Note: *On the left*, an article documenting an illegal screening of the movie the Birth of a Nation in Topeka, despite the ban from state authorities. The article was published on the Topeka Daily Capital, Topeka (KS) on June 16th, 1917. *On the right*, an ad documenting a screening of the movie *The Birth of a Nation* in a neighboring county with respect to the newspaper's headquarter. The ads was published on the Hiawatha Daily World, Brown county (Kansas), and refers to a screening in Falls City (Nebraska), located 17 miles from Hiawatha, therefore in a different county and state. Source: newspapers.com

A.3 Comparison between ERST and Ang

In this section we contrast our cross-sectional measure of local screenings of *The Birth of a Nation* at the county-level (denoted ERST hereafter) with the alternative measure (denoted Ang) collected in the companion paper by Ang (2020).⁵ We start by comparing their respective data collection strategy. Both are based on information related to local screenings of the movie that is retrieved mostly from archives of historical newspapers:

- From newspapers' archives, Ang (2020) looks at movie ads *only*. These ads are retrieved from multiple newspaper archives: *newspapers.com*, *newspaperarchive.com*, and Library of Congress' Historic American Newspapers database. He complements data from newspapers' archives with information from the movie industry periodical *The Moving Picture World*, where he visually inspects dispatches/reports of the film's roadshow.
- We use only one archive: *newspapers.com*. However, we extract from this source more information. Not only we look for movie ads, but also we recover evidence on local screenings from a variety of more "indirect" proofs by scraping a broad and heterogeneous body of newspapers' texts (e.g. articles, movie tables; more details are exposed in online Appendix Section A.2). These pieces, that we retain as proofs of screening when we code ERST, are not movie ads: Hence, they are not considered in the data collection strategy devised by Ang (2020). Overall, out of all screening records, such indirect proofs represent 76% of the pieces of evidence used in our construction and coding of ERST.

We now compare ERST and Ang and interpret their discrepancies at the light of their respective data collection strategies. At first sight, the two measures drastically differ in term of county coverage. Most of the empirical analysis in Ang (2020) is conducted in the cross-section; and Ang covers the full set of 3103 US mainland counties. We work with an unbalanced panel of 1070 counties that we are able to trace at the monthly-level over 1910-1920; and ERST is intended at covering only this sample.

Contrasting the county coverage of the two datasets does not really inform on the relative merits of the two measures. Indeed, among the 2033 counties that are not covered by ERST, 91% of them (1851 counties) are made of zeros in Ang (i.e. untreated). Where do those zeros come from? The rule used by Ang to code the zeros states that the absence of screening proofs must be interpreted as evidence that a county did not screen the movie. Hence, this way of dealing with missing information makes possible to extend the measure to the entire set of US mainland counties. However, this extra coverage does *not* mean that the sources used for constructing Ang have a richer informational content than the sources used for ERST.

Table A2: Treated Counties according to ERST and Ang

Measure	Ang		
	Not Treated	Treated	Total
ERST			
Not Treated	544	84	628
Treated	102	340	442
Total	646	424	1070

⁵In our two-way fixed effect setting, the treatment effect is captured by the coefficient of our main variable of interest BON_{ct} (see equation 1). Conceptually, BON_{ct} is the panel counterpart of the cross-sectional measure of screening, $ERST_c$. Formally, they relate to each other through this equation: $\max_t (BON_{ct}) = ERST_c$ where c is county and t is time.

The two-way Table A2 reports the cross-sectional distributions of ERST and Ang for our sample of 1070 counties. It reassuringly appears that the two measures coincide for 884 counties out of 1070, with 544 untreated units and 340 treated units according to both measures. Discrepancies between the two measures are observed for 186 counties (off-diagonal elements): 84 counties are treated in Ang and untreated in ERST; reversely, 102 counties are treated in ERST and untreated in Ang.

It appears that the discrepancies between the two measures do not result from different interpretations of the *same* pieces of information. Instead, they stem from the fact that we use *different* sources of information:

- i/ In the case of ERST, 92% of the validated screening proofs of the 102 off-diagonal counties are retrieved from the larger body of texts (articles, movie tables, etc...) that we examine in *newspapers.com*. It is likely that Ang did not document these screenings because they were not reported in the forms of ads (i.e. with an identifiable "black circle" logo), and/or they could not be detected by visual inspection. Quite importantly, all the screening records of these 102 counties have been double-checked by research assistants; this extra cleaning allows us to be assertive in claiming that these counties are "true positives" (i.e. treated in ERST and treated *de facto*). The examples of Sonoma (CA) and Dodge (NE) counties illustrate this point. For both counties, we recovered screening proofs from the archives that are not movie ads (see Figures A8 and A9); therefore, they were not used in the construction of Ang; however, they clearly indicate that local screening took place in both counties. This explains why they belong to the set of 102 counties that are treated in ERST and untreated in Ang.
- ii/ In the case of Ang, the coding of 73 counties out of the 84 off-diagonal treated counties exploits informational pieces recovered from the three additional sources that Ang uses (newspaperarchive.com, Library of Congress' Historic American Newspapers database, and the quarterly trade magazine The Moving Picture World).

Our overall interpretation of Table A2 is that ERST and Ang are highly congruent. Both measures map the spread of the movie across US counties in a consistent way. We believe that the credibility of our data work is strengthened by the fact that the two measures are based on different historical sources and were constructed independently by the two research teams. Hence, each measure offers a reliable cross-validation of the other one. It also appears that none measure clearly dominates the other: Each one identifies treated counties that are not detected by the other one.

A4. Making use of the two measures

The availability of ERST and Ang allows for important validation exercises. For this purpose, we consider more data-rich constructions of the treatment variable that combine information from both measures. There are several ways to proceed, each one with its own advantages.

- In the union approach, we define as treated all counties that are coded as treated by *at least one* of the two measures. Formally, in the cross-section, we can define the treatment variable as $\text{Union}_c = \max(\text{ERST}_c, \text{Ang}_c)$. From Table A2, we get that 526 out of 1070 counties in our sample ($=340+84+102$) are treated according to this variable. This way of combining information is suited if ERST and Ang are prone to false negatives. But it comes at the risk of increasing false positives if one of the measures is affected by them.
- In the intersection approach, we define as treated the counties that are coded as treated by *both* measures. Formally, in the cross-section, the treatment variable is equal to $\text{Inter}_c = \min(\text{ERST}_c, \text{Ang}_c)$. From Table A2, we get that 340 out of 1070 counties in our sample are treated according to this variable. By contrast with the previous approach, this way of combining information is well suited if ERST and Ang are prone to false positives. However, it

Figure A8: Excerpts of articles (not movie Ads) documenting local screenings: Sonoma county

**"THE BIRTH OF A NATION"
COMING NEXT SUNDAY NIGHT**

**Half Million Dollar Film Production in
Twelve Reels at Novelty Theatre.**

"The Birth of a Nation," (The Clansman), is coming to the Novelty Theatre next Sunday evening. This wonderful film has but recently been shown in San Francisco and is said to be the most stupendous and fascinating production ever created. It deals with the American civil war and the reconstruction period. It shows the causes that led up to this conflict and carries the spectator through the war and the revolution that followed, treating impartially the causes for which each side was struggling, and indicates the political conditions in the north and south during these struggles. It cost half a million dollars to produce this film, and 25,000 soldiers are shown in action. It will be shown but once in Cloverdale, so theatre-goers should not miss this opportunity given them. The twelve reels will be thrown on the screen, commencing promptly at 8 o'clock. Reserve your seats early.

**THE CLANSMAN
ON WEDNESDAY**

Wednesday and Thursday evenings the great masterpiece, "The Clansman" or "The Birth of a Nation," will be seen for the second and last time in this city and the price of admission will be lower than on its first appearance a year ago, so that it will be within the reach of all.

Those who have never seen this wonderful production should avail themselves of the opportunity, while many who have seen it before will go again, for it seems almost impossible to grasp all of the details of the great picture on one occasion. Thousands of persons and thousands of horses are used and the production is certainly a wonder as all will attest who have seen it.

At the Starland

Two machines are now installed at the Starland Theatre. On all big features both outfits will be used.

On Sunday, November 5, Geraldine Farrar will be shown in "Temptation." She is the same actress who was shown in "Carmen" recently.

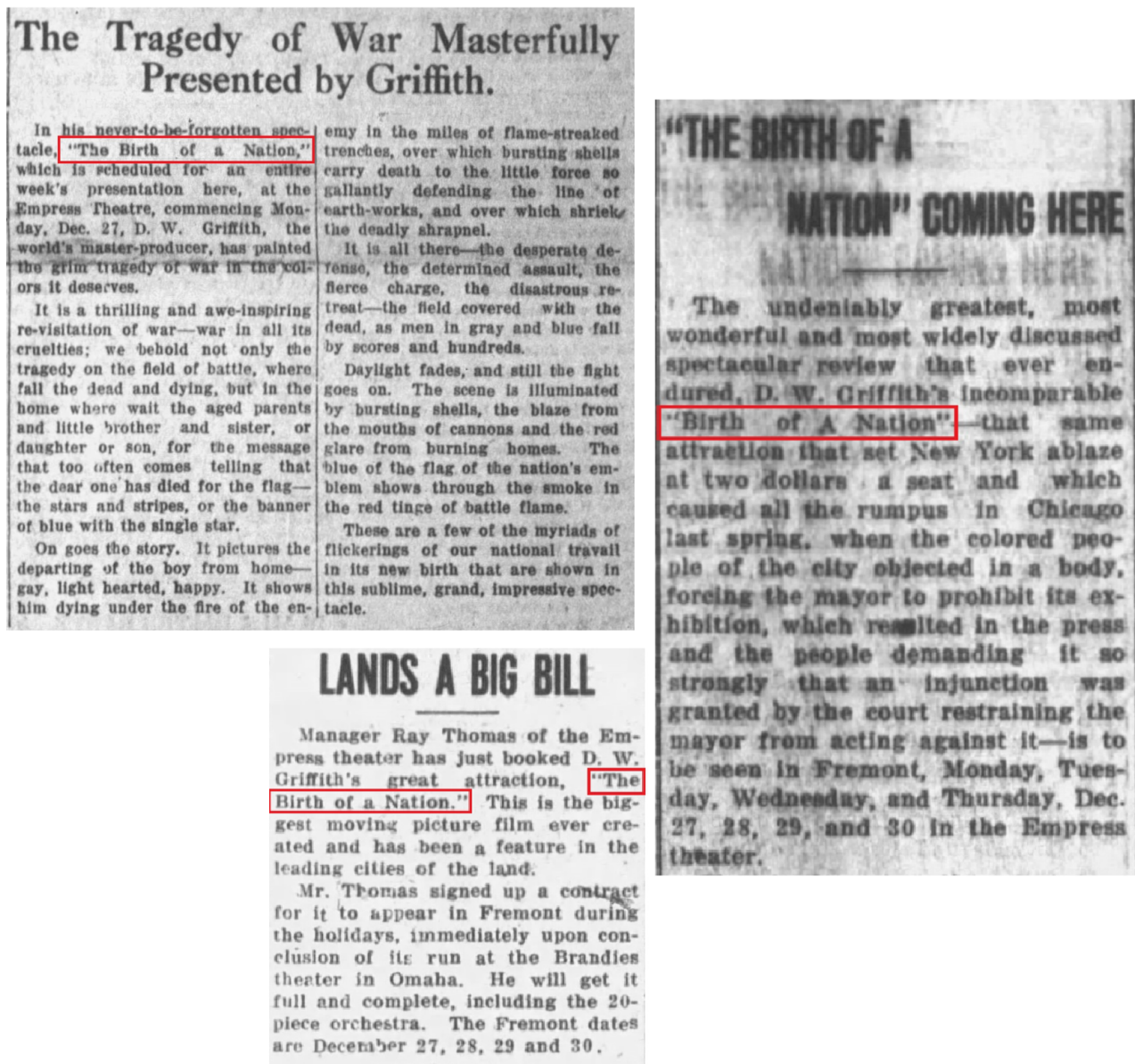
The "Clansman" or "A Birth of a Nation" will be shown Sunday, November 12. This is a big and wonderful feature in twelve reels.

NOTE: Three articles documenting screening of the movie the Birth of a Nation in Sonoma county (California). Source: *newspapers.com*.

increases the rate of false negatives because valid information on screenings is discarded as soon as it is lost in the data collection of one of the two measures.

In total, regarding our treatment variable, we are now equipped with 4 different measures of the local screening of Birth of a Nation: ERST, Ang, Union and Inter. In the new version of the paper, we use ERST as our baseline measure. This is a natural choice given our instrumental variable strategy.

Figure A9: Excerpts of articles (not movie Ads) documenting local screenings: Dodge county



NOTE: Three articles documenting screening of the movie the Birth of a Nation in Dodge county (Nebraska).
Source: *newspapers.com*.

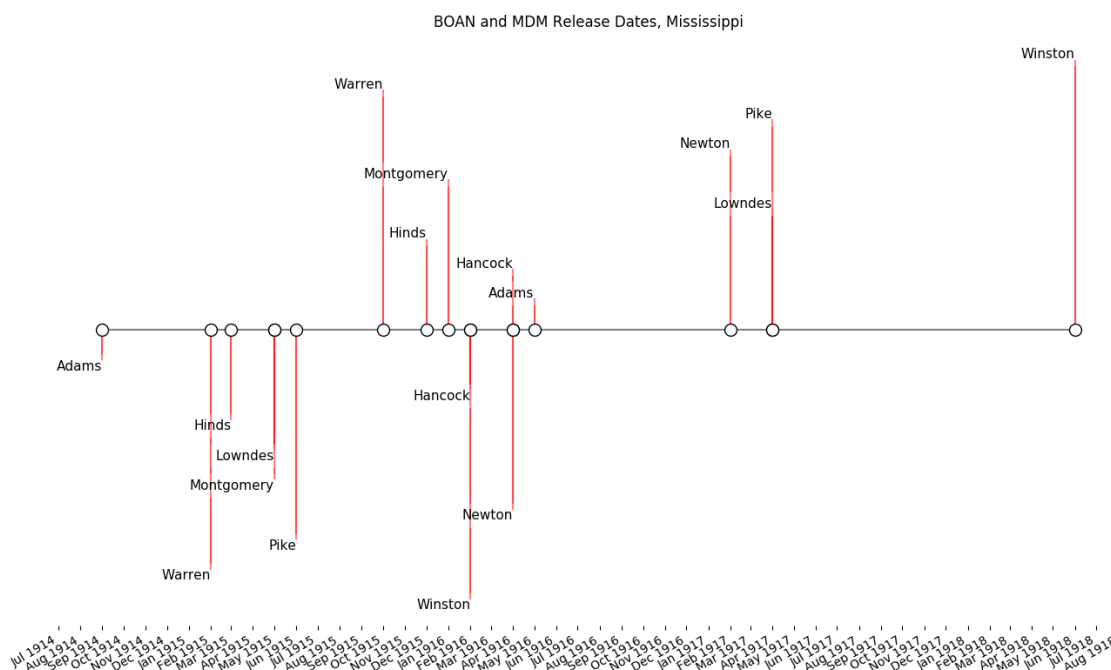
Indeed, our instrumental variable, that also measures the screening of a movie (i.e. Million Dollar Mystery), is based on the same historical archive and data collection strategy than ERST. For our baseline analysis, we opt for retaining consistency in the measurement of the screenings of the two movies. In our robustness analysis, we replicate all our baseline exercises with the 3 other variants of our treatment variable. Results are reported in Online Appendix B.2.

A.4 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 A10 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.

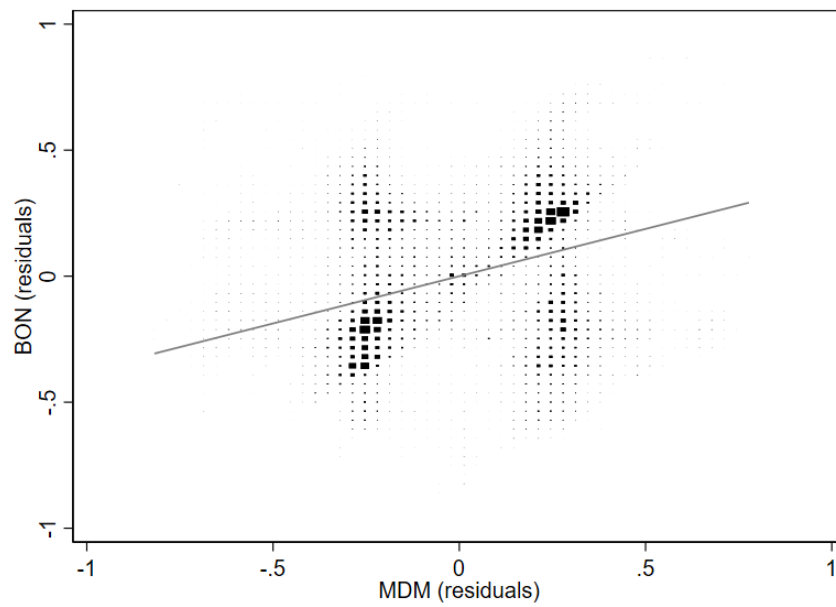
Figure A11 offers a graphical representation of the association between MDM and BON. The figure considers the residuals of the regression of BON_{ct} on MDM_{ct} obtained after filtering out the fixed effect structure we use in Equation (1). In the plane, larger rectangles correspond to a higher number of observations. Since residuals tend to distribute along the diagonal, the figure highlights that MDM_{ct} is indeed a strong predictor of BON_{ct} .

Figure A10: BON 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 BON.

Figure A11: Residuals



NOTE: The figure depicts the distribution of residuals of $BOND_{ct}$ and MDM_{ct} after filtering out month-year, county and coverage percentile fixed effects. The size of each rectangle is proportional to the number of observations in the corresponding part of the plane.

A.5 Construction of the Bag-of-Words

This section describes how we construct the two “bags” of keywords used for measuring patriotic/divisive opinions in historical newspapers.⁶ We summarize here the main steps of the procedure; all details are exposed in the next sections. Firstly, we select on a discretionary basis two starting lists of words. We ask external judges to evaluate and validate their semantic relevance as part of a *Patriotic* or *Divisive* rhetoric. We refer to these validated words as *seed* words. Secondly, relying on a large sample of texts representative of the language held in newspapers articles between 1904 and 1914, we identify words that are close to the seed words according to a co-occurrence metric. Thirdly, we ask external judges to assess the semantic relevance of the set of seed words augmented with their co-occurring words. Based on judges’ assessment, we construct a ranking of *Patriotic* and *Divisive* keywords. In our baseline analysis, the two bags of keywords are made of the top 20 keywords of each ranking.

The assessment of the semantic content of long lists of words by external judges is a key component of our procedure. We were consequently attentive at selecting judges with the right attitude, language proficiency and familiarity with the historical context. We recruited the judges using the platform MTurks. Only candidates complying with the following two criteria were hired: i) They had to be based in the United States; this requirement aims at increasing the likelihood of language proficiency and good knowledge of US history; ii) They had to successfully pass a practical test of performance in semantic analysis of words.⁷ In total, we hired 68 external judges.

A.5.1 Seed Words

On the basis of their relevance in the BON movie scripts (e.g. Figures AI and AII), in the tenets of the *Lost Cause* narrative, or in the speeches of famous historical figures advocating for reconciliation, we identify two starting lists of 12 *Patriotic* words and 17 *Divisive* words:⁸

Seed Words - Patriotic: “patriotic”; “united country”; “Americans together”; “American people”; “U.S. flag”; “Stars and Stripes”; “Americans”; “United States”; “The Star-Spangled Banner”; “reconciliation”; “American Revolution”; “fraternity”.

Seed Words - Divisive: “secession”; “civil war”; “slavery”; “confederates”; “sectionalism”; “Confederate Flag”; “Confederacy”; “Secession flag”; “Southerners”; “carpetbaggers”; “Northerners”; “yankee”; “Lost Cause”; “dixie”; “scalawags”; “bloody shirt”; “Stars and Bars”.

Semantic Analysis by External Judges. Given that the two previous lists are built on a discretionary basis, we ask external judges to validate their semantic content by assigning each word to

⁶With a slight abuse of terminology, we use the generic concepts of “words” or “keywords” to refer either to single words (e.g. “patriotic”) or to short combinations of words (e.g. 2-grams, [“united”, “country”]).

⁷The practical test was run as follows. Each candidate had to analyze a list of words including a mix of clearly patriotic and divisive terms. Inserted in the list were four terms that were *not* related to the patriotic and divisive themes. These words are “yellow fabric”, “thermometer”, “aquamarine”, “million dollar”. The practical test was considered passed if the candidate correctly spotted those four terms

⁸For the *Lost Cause*, the most relevant sources begin with the manifesto from Pollard (1866), continuing with famous speeches including, among others, Jefferson Davis’s speeches (such as the one to a group of veterans of the Army of the Tennessee in 1878) and Reverend Moses Drury Hoge speeches (such as the one at memorial day in Richmond 1875). For a commented review of several important sources, see Blight (2009). Also according to Blight (2009), the skeleton of the reconciliatory view takes shape from a series of speeches, oration and public addresses, including Wilson’s address in Gettysburg (1913), Horace Greeley’s campaign speeches (1872), Rutherford B. Hayes’s letter of acceptance as a Republican candidate (1872), and Oliver Wendell Holmes Jr. public address at the Keen Memorial Day (1884), among others. Out of this body of texts, we selected words that resonate intuitively with a divisive / reconciliatory message and either appear directly in the texts or are strictly related with terms that appear in the texts.

one of the three categories: *Patriotic*, *Divisive*, or a neutral/not applicable category N/A. In their set of instructions, judges were informed that they will be exposed to lists of words extracted from newspaper articles dating back to the aftermath of the US Civil War. The three categories were described to the judges in the following terms:

- Patriotic category: Words that typically appear in newspapers' articles for highlighting the ideas of reconciliation, patriotism and national unity around the time of the Civil War.
- Divisive category: Words that typically appear in newspapers' articles for highlighting the contrasts and fights between the Southern States and the Northern States around the time of the Civil War.
- N/A category: Words that do not relate to any of the previous categories.

Each word is randomly assigned to at least 40 judges in a blind way. The judges ignore the list from which the word comes from. The results are clearcut: All words belonging to the *Patriotic* list were assigned to the *Patriotic* category by the majority of judges; similarly, all words from the *Divisive* list were assigned to the *Divisive* category by the majority of judges. Moreover, the evaluations tended to be extremely consistent across judges: The consensus rate was equal to 88% for *Patriotic* words and 79% for *Divisive* words. For some words, a minority of judges opted for the N/A category, possibly revealing that they were not familiar with the terms; "scalawags" and "bloody shirt" are two examples of such words.

Overall, we interpret these results as a confirmation that the two starting lists of words embody semantic elements related to the Patriotic and Divisive rhetoric. Hereafter we refer to these validated words as "seed words". They constitute the set of words on which we build the rest of the procedure. Note that some of those seed words will not be retained at the final stage of the procedure when we construct the bags of keywords.

A.5.2 Co-occurring Words

In this step of the procedure, we go beyond the seed words and recover a large set of words that are potentially associated to the *Patriotic* and *Divisive* rhetoric in the language of the early twentieth century US. We devise a hands-off approach that identifies words co-occurring with the seed words in a large body of newspapers articles spanning the 1904-1914 period. Our metric of co-occurrence is borrowed from computational linguistics and is based on the Pointwise Mutual Information (PMI) score.

Newspaper Articles from Chronicling America. Having access to full texts is a pre-condition for performing the co-occurrence analysis. However, a limitation of our main data source, *Newspapers.com*, is that we can download only single words, not texts. For this reason, we instead use the archive *Chronicling America* which allows us to download full texts from millions of newspapers' pages. Given that we are interested in the language held in the period pre-dating the release of the movie, we download all texts available on the platform for the period 1904-1914, corresponding to a total of 1.7 million newspaper pages. To reduce the amount of computational resources needed for the text analysis, we randomly select a representative sample of 300,000 pages.

Our objective is to elicit co-occurrence patterns among words that belong to the *same* text or article, with the idea that semantic content is relatively coherent and stable within articles but may vary radically between articles. From *Chronicling America* we recover entire newspapers' pages. Clearly, each of these pages has potentially a very heterogeneous content as it may display articles on different topics, readers' letters, advertisements, etc. We consequently devise a simple method to isolate single articles out of the sample of newspapers' pages downloaded from *Chronicling America*: We

identify the starting phrase of each article whenever a line of text is composed with all uppercase letters; then we split the text accordingly. Note that we discard fragments of text shorter than 200 words or longer than 850, as this unusual size of articles might signal some errors in the identification of the starting line of the article. With this method, we manage to recover 1,174,469 articles. Finally, we excluded articles in foreign languages by discarding texts with less than 60% of words belonging to an English dictionary.⁹ Applying this criterion, we are left with 1,004,295 articles.

Data Cleaning and Text Processing. We implement a series of cleaning procedures to improve the readability of each article’s text, which tends to be low because the OCR algorithm does not always decipher correctly old and corrupted digitized pages. Breaks for new lines are often not indicated, typos are quite common, with the overall quality depending on the image resolution of the original digitized page. The cleaning procedure comprises several steps: (i) removing punctuation and digits; (ii) removing capitalization; (iii) tokenization; (iv) singularization; (v) removing short tokens of two or less characters (except ‘us’); (vi) tagging part of speech to retain only nouns, adjectives and verbs; (vii) removing stop words (including names of states, days of the week, and months); (viii) removing uncommon tokens (fewer than 50 occurrences). Note that we slightly deviate from what can be described as a standard procedure in natural language processing. First, in order to retain the specificity of the terms we are interested in, we do not stem the tokens. For instance, the word “southerner” and its stem “southern” have different meanings and are used in different contexts. Secondly, we add a cleaning step aimed at identifying bi-grams and words commonly split when going to a new line. This step is quite important since the lines in newspaper columns tend to be quite short, resulting in longer words being often split (the line break is rarely captured by the OCR algorithm).

2-grams. Our seed words are made of both single words and 2-grams (two consecutive words). In the same vein, we want our co-occurrence procedure to recover from the texts single words *and* 2-grams. This procedure aims to identify what are the most common 2-grams in the texts before performing the co-occurrence analysis. To this purpose, borrowing from computational linguistics, we compute the Pointwise Mutual Information (PMI) index of all 2-grams in the text—for a relevant and exhaustive review of this concept, see Manning and Schütze (1999):

$$\text{PMI}(x, y) = \log \left[\frac{\mathbb{P}(x, y)}{\mathbb{P}(x)\mathbb{P}(y)} \right]$$

where x and y are two single tokens, $\mathbb{P}(x, y)$ is the fraction of articles where both x and y appear and $\mathbb{P}(x)$ is the fraction of articles containing x .

We then define the set of most common 2-grams based on their PMI score and frequency. More precisely, we drop the 99% least common 2-grams containing two English words; then we select only those with the highest 20% PMI. We obtain a total of approximately 137 thousand 2-grams. Finally, in the texts, we substitute these most common 2-grams with their concatenated version: For example, “united country” becomes “unitedcountry”. We treat those concatenated 2-grams as single words in the next step of the analysis.

The Co-occurrence Method. In this step, we identify all words that tend to co-occur with the *Patriotic* and *Divisive* seed words.¹⁰ We start by calculating the PMI between each seed word and each

⁹Since the text was obtained from PDF through OCR processing, it contained many misspelled words and typos. The choice of the 60% cutoff permits us not to erroneously consider as foreign, English words with typos or mistakes.

¹⁰We exclude seeds that do not appear or appear only a handful of times in the sample of *Chronicling America* articles, e.g. Americans Together, Secession Flag, U.S. Flag.

word appearing in the representative sample of 1,004,295 articles. The number of combinations being very large, this is a computationally intensive exercise. Once we are equipped with this metric of co-occurrence, for each seed word, we build its set of closest neighbors: It is made of the 1000 words appearing in articles that have the highest PMI with this specific seed word. Finally, we define the set of co-occurring words as the ones that are neighbors of multiple seed words (at least one third of all the seeds).¹¹ The rationale is as follows: Whenever a word co-occurs with multiple seed words of a given category (patriotic or divisive), it is more likely to capture the semantic properties of the category. More concretely, we define the co-occurring *Patriotic* words as those belonging to at least 4 sets of closest neighbors of *Patriotic* seed words; for co-occurring *Divisive* words, we go up to 6 sets of closest neighbors of *Divisive* seed words.

Tables A3 and A4 report the lists of seed words (in bold) augmented with their co-occurring words for each semantic category (*Patriotic* and *Divisive*).¹² The column displaying integers represents the number of seed words each word is a closest neighbor of. Visual inspection reveals that, within each category, most of the words display an evident semantic commonality. This feature is indicative that our hands-off automatic procedure performs rather well in retrieving large sets of words with shared meaning. Interestingly, some seed words seem to be isolated, in the sense that they are closest neighbors of only few other seed words: “Americans together” is an illustrative example (it is a neighbor of no other seed word). Our interpretation is that these isolated seed words potentially capture a relevant but specific dimension of the semantic category.

A.5.3 Semantic Analysis by External Judges.

In this last step of the procedure, we construct the bags of keywords for the Patriotic and Divisive categories. Our co-occurrence method has the merits of being impartial and efficient at treating a very large body of texts. However, it offers only a coarse screening of the meaning of the words. Indeed, while most of the words identified by the co-occurrence method and displayed in Tables A3 and A4 share semantic properties, they do vary in how accurately they embody the Patriotic and Divisive rhetoric. For some words, the relationship with the other words from their category is not evident at all. We consequently supplement the co-occurrence method with a fine-grained analysis conducted by external judges; our aim is to build a ranking of the semantic properties among co-occurring words. The task of the judges is similar to the one described in Section A.5.1. They have to assign each word to one of the three categories: Patriotic, Divisive or n/a. Then, we aggregate information across the 68 judges by constructing for each word its semantic index:

$$\text{Semantic Index}_i = \left[\frac{\text{share}(\text{Patriotic})_i - \text{share}(\text{Divisive})_i}{1 + \text{share}(\text{N/A})_i} \right] \quad (1)$$

where $\text{share}(C)_i$ is the share of judges who assigns the word i to category $C \in \{\text{patriotic}; \text{divisive}; \text{N/A}\}$. This index measures in a compact way where the word locates in the *Divisive - Patriotic* dimension. It ranges from -1 to +1. The extreme values correspond to words that are unanimously assigned to the divisive and patriotic category respectively. By contrast, an index with an absolute value close to zero captures a semantic content that is ambiguous. Such a case happens when (i) $\text{share}(\text{N/A})$ is large, indicating that the word is neutral; and when (ii) the judges’ assessments balance equally between the patriotic and divisive categories.

We report the computed values of the semantic index for each word in Tables A3 and A4. Words have been ranked in decreasing order of the index in the first table (patriotic category) and in

¹¹We refrain to aggregate information across seed words by averaging their PMIs because of the bias towards combination of rare words. See for example Turney and Pantel (2010).

¹²In order to limit redundancies, we do not report words that are close permutations of other words already in the list. For example, we exclude the bi-gram “True Patriotism” because “Patriotism” is already in the list.

Table A3: Keywords Co-Occurring with Patriotic Seed Words - Semantic Index and Co-occurrence Score

Semantic Index	# Top 1000 Seeds	Keyword	Semantic Index	# Top 1000 Seeds	Keyword
1.00	4	American Flag	0.26	5	Marines Sailor
1.00	3	Patriotic	0.23	5	International Arbitration
1.00	4	Patriotism	0.20	4	Florence Nightingale
1.00	4	True Patriot	0.18	4	Infantry Band
0.95	6	United Country	0.16	4	International Relations
0.94	0	Americans Together	0.15	5	Union Veteran
0.92	5	National Hymn	0.14	4	Gave Recitation
0.92	5	National Salute	0.14	4	Landing Marines
0.92	6	Salute Flag	0.07	4	Union Veteran Legion
0.92	4	Liberty and Freedom	0.07	4	Navigation Law
0.91	1	American people	0.05	4	Subvention
0.87	0	U.S. Flag	0.04	4	Emilio Aguinaldo
0.86	4	Stars and Stripes	0.04	4	Literary Exercise
0.86	0	Americans	0.02	4	Perpetuate Memory
0.86	1	United States	0.00	5	Coastwise Vessel
0.86	4	The Star-Spangled Banner	0.00	4	General William H. Barry Garrison
0.85	8	Liberty and Equality	0.00	4	Neighboring Republic
0.85	5	Our Flag	0.00	4	Tricolor
0.85	4	Restoration of Peace	-0.06	4	Fantasie
0.81	1	Reconciliation	-0.07	4	Balder
0.77	4	Honor and Dignity	-0.07	4	Jingoism
0.72	4	Freedom of Speech	-0.07	4	Non-Interference
0.70	4	Salute of 21 Guns	-0.07	4	Occupation of Vera Cruz
0.68	4	Old Glory	-0.08	4	Tolls Repeal
0.60	4	Declaration of Independence	-0.11	5	Sister Republic
0.52	4	Salute Fired	-0.11	4	Conditions Have Arisen
0.50	4	Continental Congress	-0.13	4	Armed Force
0.47	4	Rousing Cheer	-0.20	4	Foreign Shore
0.43	4	Splendid Tribute	-0.20	4	Foreign Competitor
0.43	4	Flag Float	-0.24	4	Taxation Without Representation
0.37	4	Unveiling Ceremony	-0.28	4	Foreign Nation
0.36	2	American Revolution	-0.39	4	Missionary Ridge
0.35	4	Continental Army	-0.47	4	Imperialistic
0.32	1	Fraternity	-0.60	5	Foreign Invasion
0.31	4	Monument Unveiled	-0.63	5	Battle Flag
0.30	4	Formal Ceremony	-0.66	4	Armed Intervention
0.30	4	The Landing of the Pilgrims	-0.71	4	Armed Conflict
0.28	4	Merchant Marine	-0.76	4	Women of the Confederacy
0.26	5	Presence of Distinguished	-0.81	4	Confederate Flag
0.26	4	Sons of the Revolution	-0.92	4	Unpatriotic
0.26	4	Loyal Legion			

NOTE: This table reports the extended list of patriotic keywords. This list is obtained by merging the list of seeds with the list of keywords that co-occur with at least 4 seeds, as described in Subsection A.5.2. *Semantic Index* is defined in Equation (1) and aggregated judges' evaluations to measure how Patriotic a word is. *# Top 1000 Seeds* counts the number of seeds that have the keyword among their top 1000 co-occurring words, calculated using the PMI index. We highlight in bold the list of seed words.

increasing ordering in the second table (i.e. divisive category). We see that within each category the index spans the entire range from -1 to +1. This reflects the fact that our co-occurrence method is efficient at detecting regularities in associations of words but is not designed for discriminating between positive (e.g. synonyms) and negative (e.g. antonyms) associations. Importantly, seed words tend to have a high index (in absolute terms) confirming that our initial starting list of seed words was semantically relevant. Yet, some seed words are associated with a pretty low index.

Table A4: Keywords Co-Occurring with Divisive Seed Words - Semantic Index and Co-occurrence score

Semantic Index	# Top 1000 Seeds	Keyword	Semantic Index	# Top 1000 Seeds	Keyword
-1.00	6	Secede	-0.35	9	Southland
-1.00	6	Secessionist	-0.28	6	Impetuosity
-0.95	4	Slavery	-0.22	5	Bloody Shirt
-0.91	8	Secession	-0.19	6	Musket
-0.90	3	Civil War	-0.16	6	Crushed to Earth
-0.86	11	Confederates	-0.15	6	Blue and Gray
-0.82	7	Sectionalism	-0.14	6	Tattered
-0.81	7	Confederate Flag	-0.12	7	Breastwork
-0.79	6	Race Problem	-0.12	7	Magnolia
-0.77	13	Confederacy	-0.07	6	Battle-scarred
-0.77	0	Secession flag	-0.07	7	Chieftain
-0.76	6	Negro Domination	-0.07	6	Sickle
-0.74	8	General Lee	-0.05	6	Ingrained
-0.70	11	Southerners	-0.05	10	General Grant
-0.66	8	Picket Line	-0.04	9	Stars and Bars
-0.63	6	Battle Flag	-0.03	6	Federate
-0.62	6	Carpetbaggers	0.00	8	Wore the blue
-0.61	7	Northerners	0.04	8	Unveiled
-0.59	0	Yankee	0.19	6	Command a Regiment
-0.59	10	General Sickles	0.19	7	Hallowed
-0.59	10	Robert E. Lee	0.20	7	Crosses of Honor
-0.57	6	Wartime	0.24	7	Erect Monument
-0.57	3	Dixie	0.24	7	Handsome Monument
-0.57	8	Lost Cause	0.25	6	Charm of Manner
-0.57	3	Scalawags	0.26	6	The Emancipation Proclamation
-0.56	7	Sectional	0.41	6	Annual Reunion
-0.56	6	Southern Planters	0.44	6	Honorable Discharge
-0.53	13	Wore the Gray	0.47	6	Comrades in Arms
-0.50	8	Bull Run	0.57	7	Reconstruction Period
-0.47	6	Rancor	0.59	6	Soldier and statesman
-0.39	6	Missionary Ridge	0.71	9	Valor
-0.39	6	Southern Province	0.86	6	Stars and Stripes
-0.38	6	Abolitionist	0.95	9	United Country

NOTE: This table reports the extended list of divisive keywords. This list is obtained by merging the list of seeds with the list of keywords that co-occur with at least 6 seeds, as described in Subsection A.5.2. *Semantic Index* is defined in Equation (1) and aggregated judges' evaluations to measure how Patriotic a word is. *# Top 1000 Seeds* counts the number of seeds that have the keyword among their top 1000 co-occurring words, calculated using the PMI index. We highlight in bold the list of seed words.

This demonstrates the value added of the co-occurrence method that allows to recover words that were semantically more meaningful than the initial seed words (and not subject to a discretionary selection).

Our ultimate task consists in setting the criterion of inclusion in our bags of keywords. Here, a stringent limit relates to the number of keywords in each bag. Indeed, retrieving keywords from *Newspapers.com* for each county and month is extremely computationally- and time-consuming. As a consequence, in the baseline analysis, we select the top 20 *Patriotic* and top 20 *Divisive* words in the rankings based on the *Semantic Index*. In our robustness analysis of Section C.2.1, we propose alternative ways of constructing the bags of keywords to verify that the results do not rest on particular coding choices. The two bags of keywords used in our baseline empirical analysis are:

Bag of Keywords - Patriotic: "American Flag", "American People", "Americans", "Americans Together", "Liberty and Equality", "Liberty and Freedom", "National Hymn", "National Salute", "Our Flag"; "Patriotic", "Patriotism", "Reconciliation", "Restoration of Peace", "Salute Flag", "Stars and Stripes", "The Star-Spangled Banner", "True Patriot", "U.S. Flag", "United Country", "United States".

Bag of Keywords - Divisive: "Armed Conflict", "Armed Intervention", "Battle Flag", "Carpetbaggers", "Civil War", "Confederacy", "Confederate Flag", "Confederates", "General Lee", "Negro Domination", "Picket Line", "Race problem", "Secede", "Secession", "Secession Flag", "Secessionist", "Sectionalism", "Slavery", "Southerners", "Unpatriotic".

A.6 Principal Components Analysis (PCA): Additional Results

This section reports additional information related to the principal component analysis. Figures A12 and A13 refer to the set of patriotic and divisive keywords used in the baseline analysis, respectively. They report the share of variance explained by the 20 components and the factor loadings assigned to each keyword to obtain the first component.

Figure A12a shows that in the case of patriotic keywords, the first component accounts for approximately 70 percent of the variance. Figure A12b shows that relatively popular keywords as “Americans” and “United States” are assigned lower factor loadings. Figures A13a and A13b repeat the same analysis for the set of divisive keywords. In this case, the first component explains 84.9% of the variance. Among the divisive keywords, “Civil War” is assigned a lower factor loading.

Figures A14a and A14b display the fraction of variance explained by the different components for the extended set of patriotic and divisive keywords, respectively. In both cases, the fraction of variance explained by the first component is similar to the one observed for the baseline lists.

Figure A12: Baseline patriotic Keywords - Principal Component Analysis

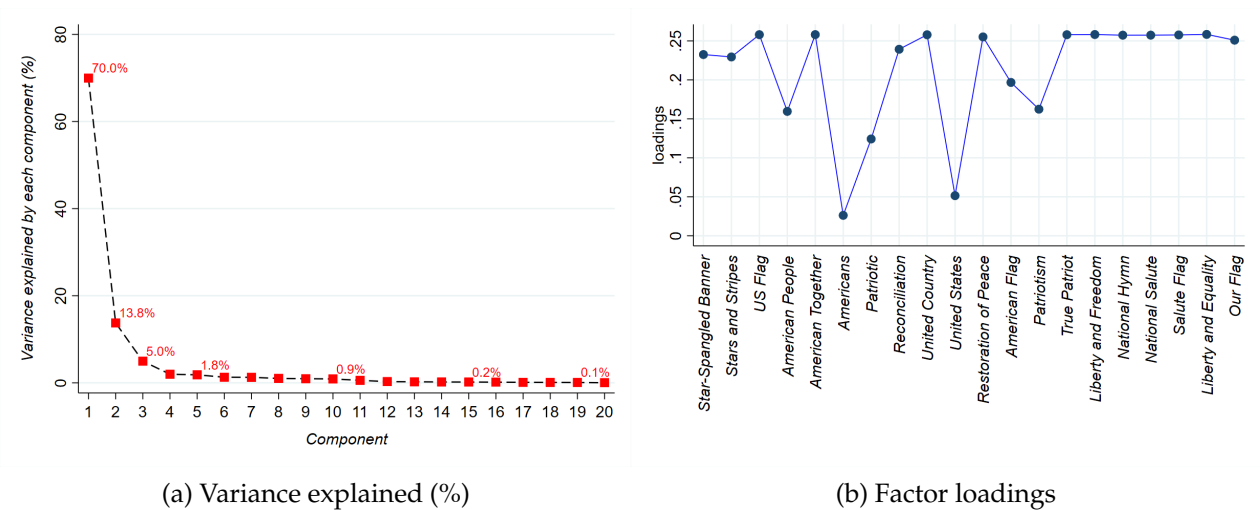


Figure A13: Baseline divisive keywords - Principal Component Analysis

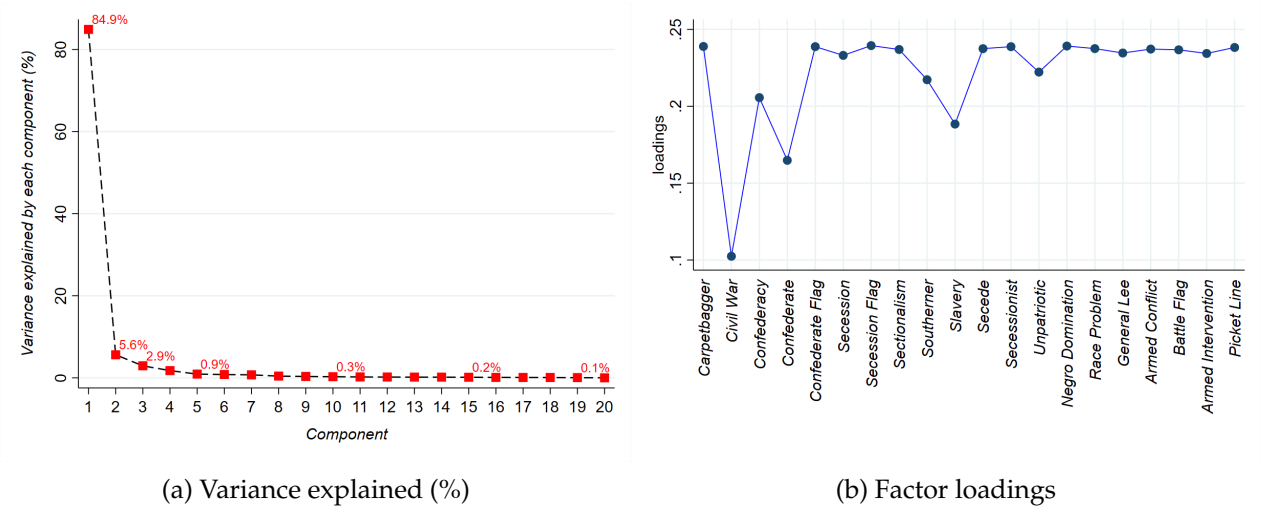
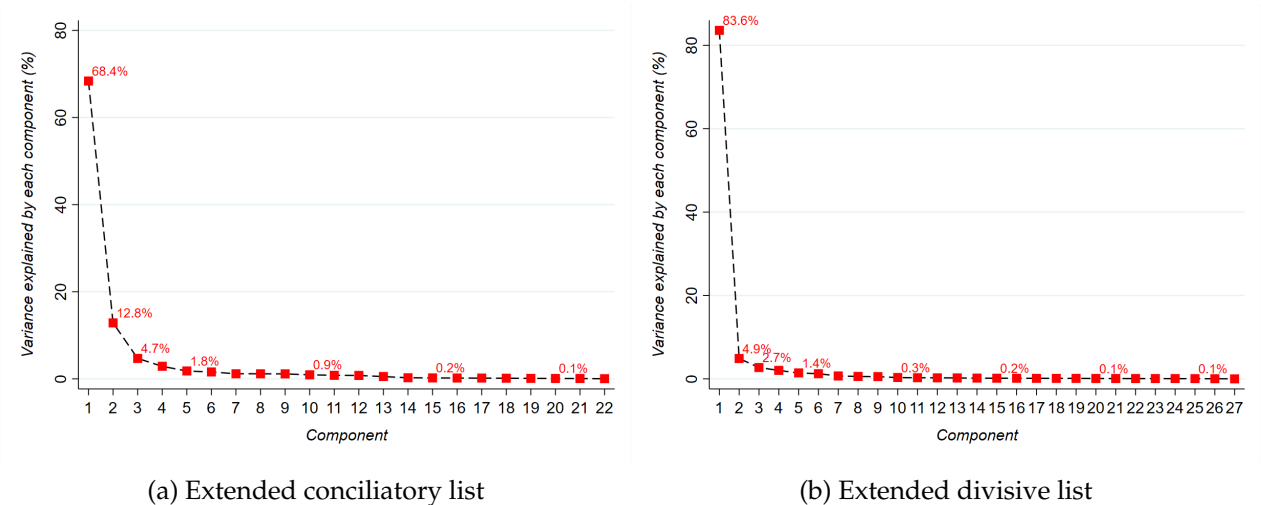


Figure A14: Extended list of keywords - Principal Component Analysis



A.7 Outcome Variables: Descriptive Statistics

This section reports descriptive statistics for the main outcome variables. Table A5 shows the number of observations, mean and standard deviation of our main outcome variables, for the sample of observations in baseline specifications. Subsection A.7.1 looks at how outcome variables are associated and how they correlate with demographic characteristics. Subsection A.7.2 and A.7.3 respectively show time series graphs and the geographical distribution of the main outcome variables. More details on the lists of words we use in the main analysis are provided in Subsection A.7.4. Finally, A.7.6 provides more information on the Enemy Name Index.

Table A5: Descriptive Statistics

	Observations	Mean	SD
Reconciliation	89,325	0.000	0.928
Enlistments Navy	62,968	0.034	0.181
Enemy Name Index	9,206	45.161	9.883
Supremacism	89,325	0.107	0.309
Discrimination	89,325	0.117	0.321

NOTE: The table reports descriptive statistics for the main outcome variables, calculated for the samples used in the baseline regressions. The outcome variables *Reconciliation*, *Navy Enlistments*, *Supremacism*, and *Discrimination* are observed at the county-month level. Data on the *Enemy Name Index* are aggregated at the county-year level.

A.7.1 Correlation Between Outcome Variables

Table A6 reports measures of association between the outcome variables. These are obtained by regressing one variable on the other, while keeping the same fixed effect structure used in the baseline analysis. These data show how the measure of reconciliation is strongly associated with the other outcome variables, suggesting that it is indeed a credible indicator of reconciliation. At the same time, the association between measures capturing behavioral aspects (Navy enlistments and the Enemy Name Index) as well as the two measures on discrimination and white supremacy is positive, but not perfect. This can be interpreted as suggesting how these outcome variables are capturing different aspects of this phenomenon.

Table A7 instead measures the association between outcome variables and a set of demographic characteristics obtained from the Census. Share of illiterate population, share of Black population, share of males, share of population in cities, and share of foreign populations are measured in 1910. Intermarriages instead uses data from 1910 and 1920. For each county, it reports the (log) number of *Former Enemies Households*, namely household where one two spouses was born in former Confederate states, while the other was born in former Unionist states. Since censuses do not report information on the year of marriage, data are aggregated at the (county $\times$ census year) level. It is remarkable how intermarriages between Northerners and Southerners is positively correlated with the outcome variables.

Table A6: Association between outcome variables - OLS regressions

	MATRICE				
	Reconciliation	Navy_Enlistments	Enemy_Name_Index	Supremacism	Discrimination
Reconciliation	1	0	0	0	0
r2	0	0	0	0	0
Obs	89325	0	0	0	0
Navy Enlistments	.038	1	0	0	0
r2	9.267	0	0	0	0
Obs	46457	62968	0	0	0
Enemy Name Index	.012	.02	1	0	0
r2	2.086	1.307	0	0	0
Obs	7128	5042	9206	0	0
Supremacism	.038	.011	.019	1	0
r2	10.819	1.187	.937	0	0
Obs	89325	46457	7128	89325	0
Discrimination	.032	.024	.041	.024	1
r2	8.189	2.055	2.11	4.364	0
Obs	89325	46457	7128	89325	89325

NOTE: The table reports beta coefficients from OLS regressions of the association between couples of outcome variables. Variables identifying a column are used as outcome variables, while variables identifying a row are used as explanatory variables. All regressions control for county FE, time FE (either month-year or year), and coverage percentile FE (if at least one of the two variables is obtained from *newspapers.com*). T-statistics are reported in parentheses. Values of the *Enemy Name Index* measure are collapsed at the county level for this analysis. The sample of observations varies according to the set of observations used in the main analysis. Whenever the samples do not coincide, we take the intersection. Finally, we aggregate monthly variables at the year level when considering the association with Names.

Table A7: Association between outcome variables and demographic characteristics

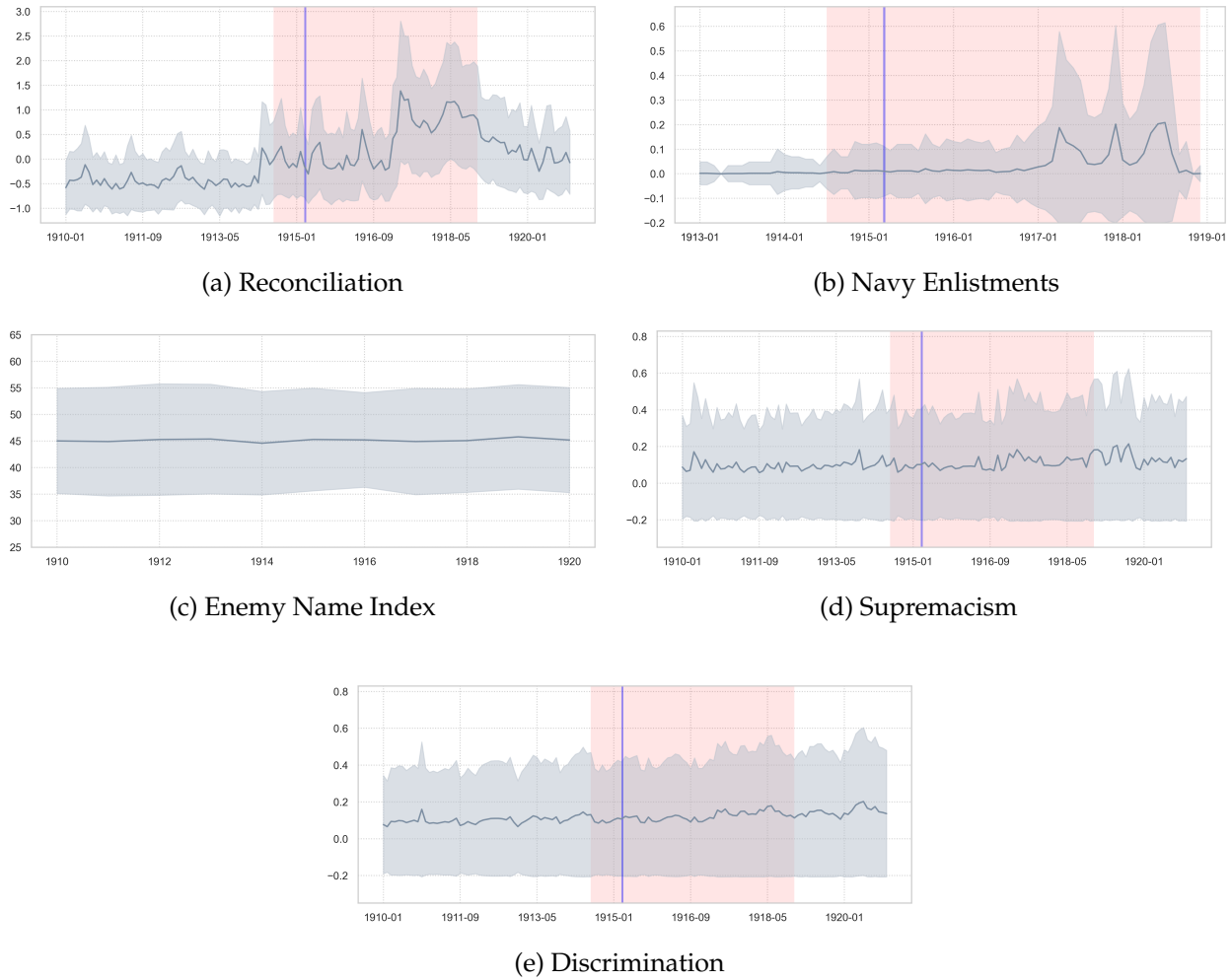
	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Navy Enlistments	Enemy Name Index	Supremacism	Discrimination
Share of Illiterate (10+)	-0.267 (-82.970)	-0.092 (-15.032)	0.234 (24.175)	-0.121 (-36.390)	-0.129 (-38.931)
Observations	89325	62968	9209	89325	89325
Share of Black	-0.260 (-80.450)	-0.062 (-9.664)	0.232 (23.649)	-0.068 (-20.503)	-0.066 (-19.789)
Observations	89325	62968	9209	89325	89325
Share of Males	0.042 (12.622)	-0.068 (-10.292)	-0.063 (-5.594)	-0.055 (-16.444)	-0.065 (-19.377)
Observations	89325	62968	9209	89325	89325
Share of Population in Cities (25k)	0.319 (100.508)	0.292 (61.821)	-0.047 (-6.426)	0.409 (133.894)	0.475 (161.208)
Observations	89325	62968	9209	89325	89325
Share of Foreign Population	0.335 (106.156)	0.195 (29.585)	-0.233 (-22.957)	0.202 (61.627)	0.217 (66.505)
Observations	89325	62968	9209	89325	89325
marr_std	0.123 (6.217)	0.000 (.)	0.075 (3.123)	0.337 (13.611)	0.355 (13.483)
Observations	1306	1306	1651	1306	1306

NOTE: The table reports beta coefficients from OLS regressions of the association between outcome variables and demographic characteristics. *Share of Illiterate*, *Share of Black*, *Share of Males*, *Share of Population in Cities*, and *Share of foreign population* are measured at the county level from the 1910 Census. *Intermarriages* is measured at the county level in 1910 and 1920, see the text for a description of how this variable is constructed. T-statistics are reported in parentheses. Values of the *Enemy Name Index* measure are collapsed at the county level for this analysis. The sample of observations varies according to the set of observations used in the main analysis. Finally, we aggregate monthly variables at the year level when considering the association with Names.

A.7.2 Time-Series Evolution

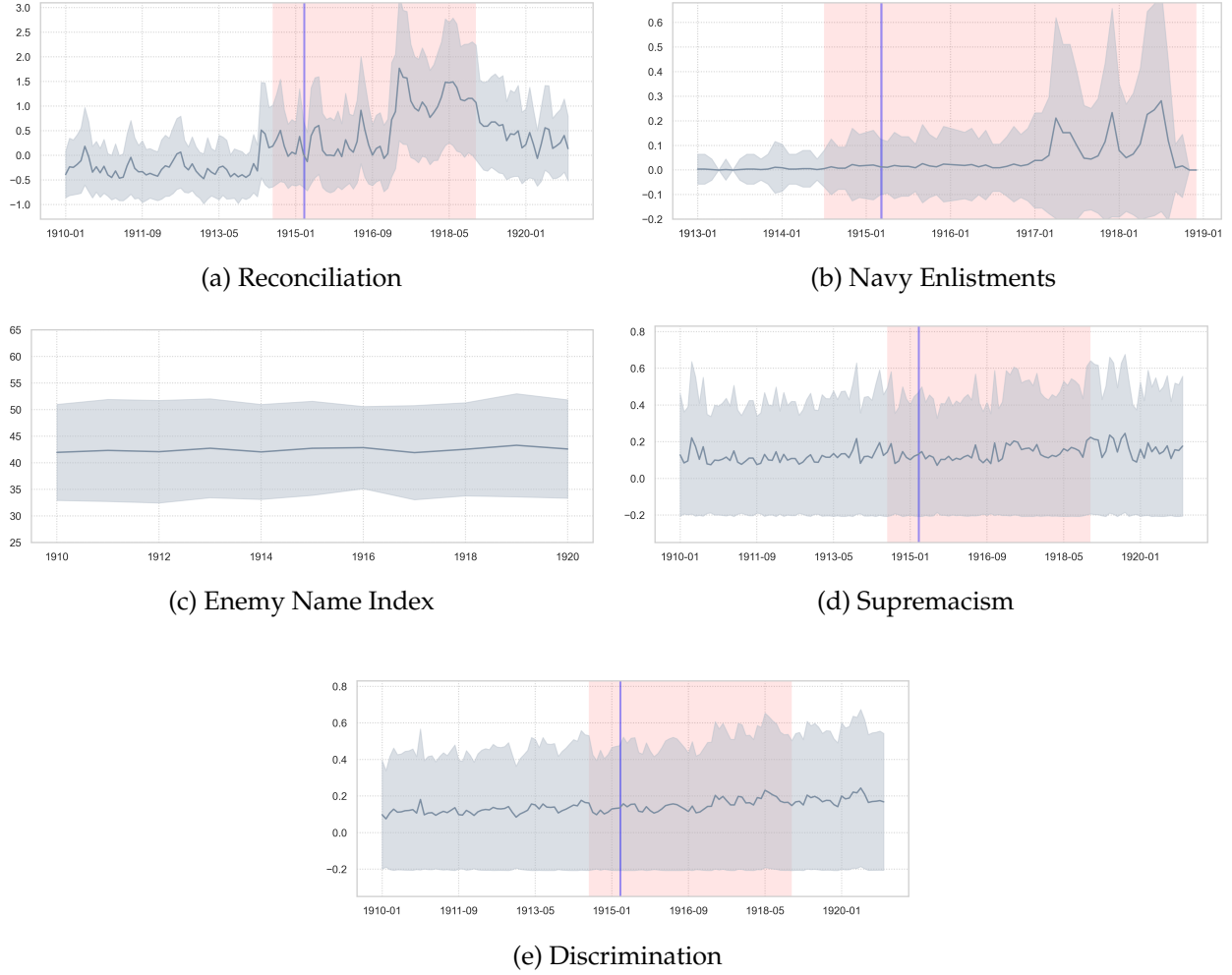
Figure A15, A16, and A17 depict the time series evolution of the outcome variables for the entire sample of counties, counties in Unionist states, and counties in Confederate states, respectively. Among the patterns that emerge, it is important to underline how the outbreak of the World War I in Europe in 1914 and the subsequent involvement of the United States in 1917 see an increase in the text-based measure of reconciliation. Also, former Confederate and Unionist states do not seem characterized by different trends in any of the outcome variables.

Figure A15: Outcome Variables - Time Series



NOTE: The Figure presents the time series evolution of *Reconciliation* (a), *Navy Enlistments* (b), *Enemy Name Index* (c), *Supremacism* (d), and *Discrimination* (e). Panels (a), (b), (d), and (e) report the monthly averages calculated for the entire set of counties (dark gray solid line). The interval in light gray is obtained by adding and subtracting 1 standard deviation, calculated at the monthly level. Panel (c) reports yearly quantities. We highlight in blue the initial release of “*The birth of a Nation*”. We highlight in red the months in which World War I was fought in Europe.

Figure A16: Outcome Variables - Time Series - Unionist States



NOTE: The Figure presents the time series evolution of *Reconciliation* (a), *Navy Enlistments* (b), *Enemy Name Index* (c), *Supremacism* (d), and *Discrimination* (e). Panels (a), (b), (d), and (e) report monthly averages calculated at the county level in former Unionist states (solid dark gray line). The interval in light gray is obtained by adding and subtracting 1 standard deviation, calculated at the monthly level. Panel (c) reports yearly quantities. We highlight in blue the initial release of "The birth of a Nation". We highlight in red the months in which World War I was fought in Europe.

Figure A17: Outcome Variables - Time Series - Confederate States

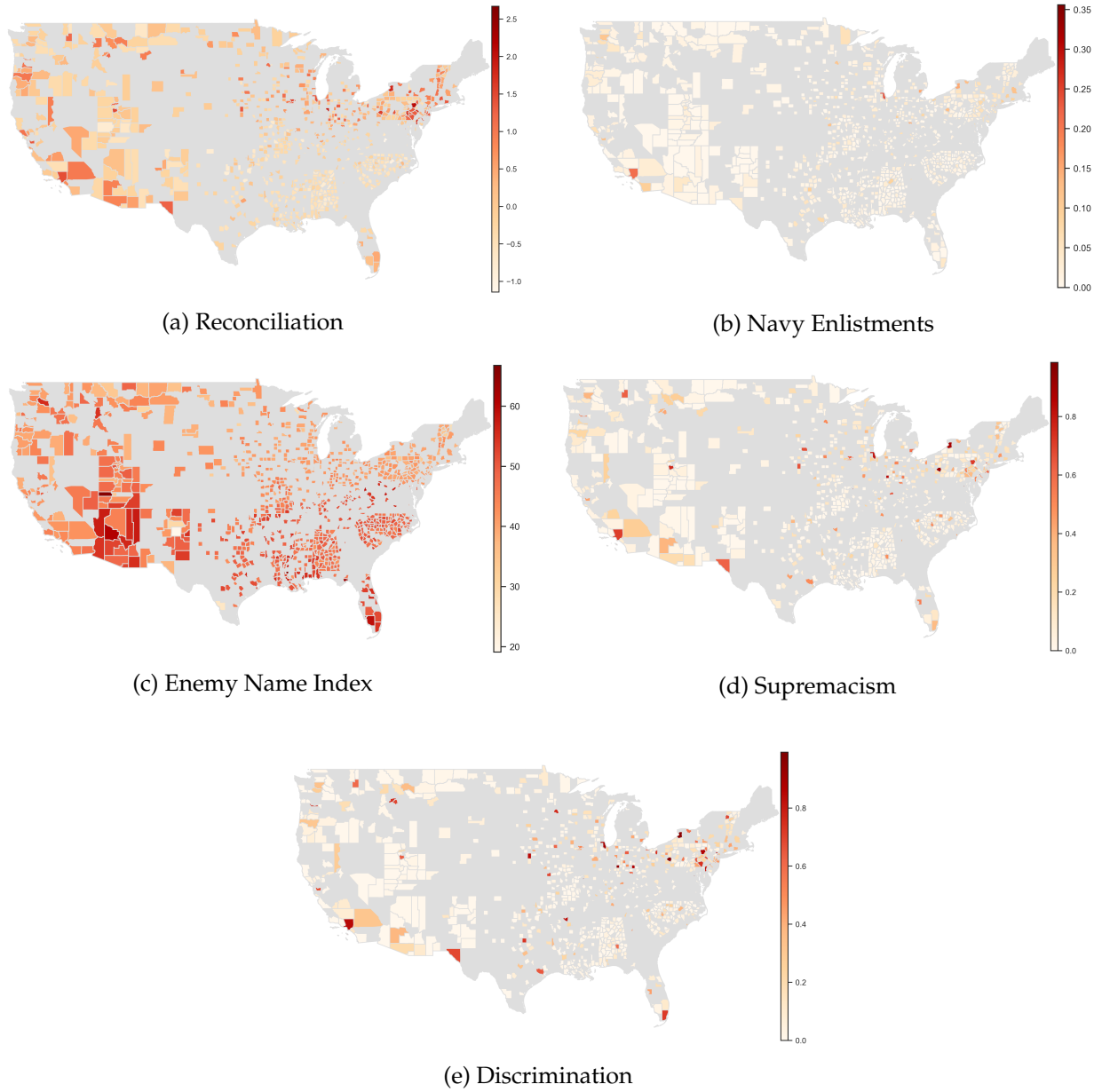


NOTE: The Figure presents the time series evolution of *Reconciliation* (a), *Navy Enlistments* (b), *Enemy Name Index* (c), *Supremacism* (d), and *Discrimination* (e). Panels (a), (b), (d), and (e) report monthly averages calculated at the county level in former Confederate states (solid dark gray line). The interval in light gray is obtained by adding and subtracting 1 standard deviation, calculated at the monthly level. Panel (c) reports yearly quantities. We highlight in blue the initial release of “*The birth of a Nation*”. We highlight in red the months in which World War I was fought in Europe.

A.7.3 Geographical Distribution

Figure A18 instead shows the geographical distribution of the outcome variables. The outcome variables appear spatially correlated. There also seem to be differences across former Union and Confederate states: the text-based reconciliation measure has lower average values in former Confederate states.

Figure A18: Outcome Variables - Maps



NOTE: The Figure presents the county level distribution of *Reconciliation* (a), *Navy Enlistments* (b), *Enemy Name Index* (c), *Supremacism* (d), and *Discrimination* (e). Darker shades of red are associated with higher average values. Counties colored in gray represent missing values.

A.7.4 Reconciliation Rhetoric: Lists of keywords

Tables A8 and A9 show descriptive statistics for the lists of words - patriotic and divisive - used in the baseline analysis. The keywords “Americans” and “United States” appear as outliers in terms of popularity when compared to the other words in the lists.

Table A8: Descriptive Statistics - List of Patriotic Keywords

	Number of pages		Frequency	
	(1) Mean	(2) SD	(3) Mean	(4) SD
Star-Spangled Banner	1.996	7.062	.007	.016
Stars and Stripes	2.246	6.899	.01	.024
US Flag	.016	.161	0	.001
American People	7.819	16.729	.034	.04
American Together	.076	.426	0	.002
Americans	126.29	254.554	.465	.215
Patriotic	17.159	46.606	.072	.097
Reconciliation	1.058	2.947	.003	.009
United Country	.083	.384	0	.003
United States	168.394	295.342	.749	.225
Restoration of Peace	.235	.913	.001	.004
American Flag	4.156	12.343	.014	.023
Patriotism	8.586	22.875	.038	.056
True Patriot	.168	.659	.001	.005
Liberty and Freedom	.184	.719	.001	.005
National Hymn	.231	.867	.001	.004
National Salute	.058	.39	0	.002
Salute Flag	.147	.81	.001	.004
Liberty and Equality	.106	.472	0	.003
Our Flag	.838	2.618	.004	.011

NOTE: The table reports descriptive statistics for the list of 20 patriotic keywords used in the baseline specification. Columns 1 and 2 refer to the monthly number of pages containing each word at the county level. Columns 3 and 4 report statistics on the frequencies, defined as the number of pages containing the word, over the total number of pages in the county-month.

Table A9: Descriptive Statistics - List of Divisive Keywords

	Number of pages		Frequency	
	(1) Mean	(2) SD	(3) Mean	(4) SD
Carpetbagger	.052	.302	0	.004
Civil War	11.37	23.561	.048	.046
Confederacy	1.811	5.416	.01	.022
Confederate	5.645	13.921	.035	.056
Confederate Flag	.095	.537	.001	.004
Secession	.426	1.374	.002	.007
Secession Flag	.001	.041	0	0
Sectionalism	.16	.689	.001	.004
Southerner	1.214	3.495	.005	.015
Slavery	2.692	5.909	.011	.02
Secede	.156	.591	.001	.004
Secessionist	.072	.503	0	.002
Unpatriotic	.692	2.576	.003	.01
Negro Domination	.026	.241	0	.002
Race Problem	.136	.607	.001	.004
General Lee	.283	.99	.002	.007
Armed Conflict	.129	.575	0	.003
Battle Flag	.183	.715	.001	.004
Armed Intervention	.218	1.043	.001	.004
Picket Like	.08	.511	0	.003

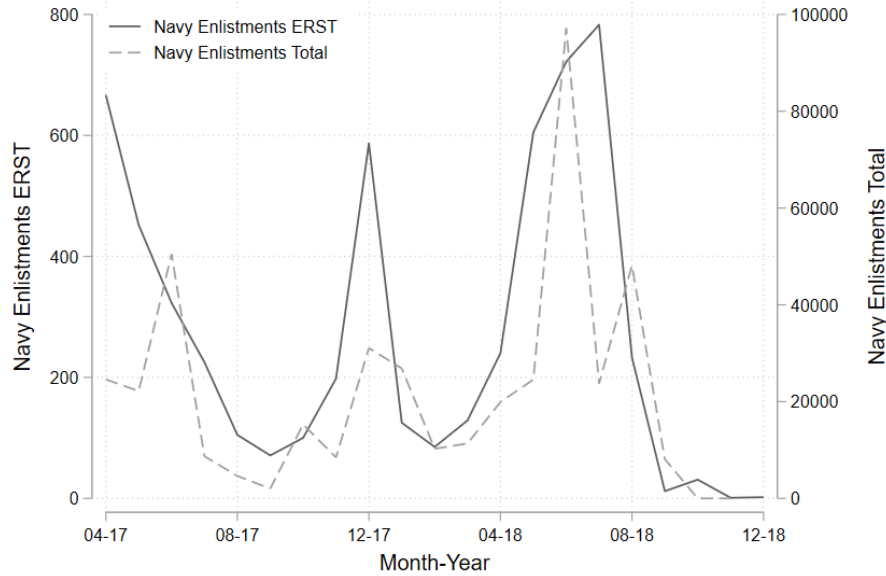
NOTE: The table reports descriptive statistics for the list of 20 divisive keywords used in the baseline specification. Columns 1 and 2 refer to the monthly number of pages containing each word at the county level. Columns 3 and 4 report statistics on the frequencies, defined as the number of pages containing the word, over the total number of pages in the county-month.

A.7.5 Enlistment in the Navy: Validation of the Measure

As explained in Section A., we use data on casualties (deaths) suffered by the US Navy in the First World War to proxy for time-varying county-level rates of enlistment in the Navy. The data come from the Bureau of Navigation, were digitized by G. Smith and retrieved at <https://www.naval-history.net/WW1NavyUS-CasualtiesAlphaA.htm>.

The dataset covers casualties of individuals who enlisted in the Navy between January 1913 and November 1918. Our strategy is to proxy for total enlistment in the Navy by looking at the enlistment of sailors who died during the war. Our approach, thus, 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, μ_c and ν_t . Then, abstracting from the +1 transformation, we get $\log \text{List}_{c,t} = \log \mu_c + \log \nu_t + \log \text{List}_{c,t}^{\text{tot}}$ where $\log \text{List}^{\text{tot}}$ 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 $\log \text{List}$ as a dependent variable yields the same coefficient of interest β as the ideal specification with the unobserved variable $\log \text{List}^{\text{tot}}$.

Figure A19: Total Enlistment and Proxy-based Enlistment



NOTE: The dashed line depicts total enlistment in the Navy between April 1917 and November 1918 as reported by the *Second Report of the Provost Marshal General to the Secretary of War on the Operations of the Selective Service System to December 20, 1918*. The solid line depicts the total number of enlisted soldiers by enlistment date, among those that died in service, relying on data from the Bureau of Navigation. See Appendix Section A.7.5 for further details.

As a way to validate our data source and show that our measure closely mirrors total enlistment, we aggregate our county-month figures on enlistment of deceased sailors ($\log \text{List}$) at the national level and contrast this time-series with official measures of total enlistment ($\log \text{List}^{\text{tot}}$) from the *Second Report of the Provost Marshal General to the Secretary of War on the Operations of the Selective Service System to December 20, 1918*, available from April 1917 to November 1918 (Table 79, Enlistments in the "Navy and Reserves" column). The two time-series, depicted in Figure A19, are correlated at 0.61 and co-move quite markedly over the entire period.

A.7.6 Popular Names and ENI

In this section, we document that, in the historical context of the US before the 1915 release of BON, the most popular first names tended to be the least enemy sounding. We also establish a list of typical Southern names and a list of names typical Northern names. To this purpose, we retrieve an exhaustive set of 5631 first names from the census data of the 1880 10% and 1900 5% Integrated Public Use Microdata Series (IPUMS) samples.¹³ For each name, we compute its frequency separately in the South (i.e. Confederate states) and in the North (i.e. Unionist states) using the definition of regions of footnote 43. Within each region, we compute the rank of each name in term of decreasing popularity as measured by its frequency (i.e. top ranks being associated with *more* frequent names). Feeding equation (3) with the measured frequencies, we compute for each name its Enemy Sounding Name Index (ENI). Note that the ENI is computed according to two possible perspectives, defined as whether the name holder is from a former Confederate state or from a former Unionist state. Finally, we also compute the rank of each name in term of increasing ENI (i.e. top ranks being associated with *less* Enemy Sounding names).

The (unreported) distribution of name popularity is extremely skewed, a standard pattern observed in other cultural contexts (see Algan et al. 2022): In the South, the top 1% most popular names (56 names) are held by 34% of the population (in the North, it amounts to 37%). The Spearman's rank correlation between popularity and ENI is positive (0.51 in the South and 0.42 in the North) and statistically highly significant (p-value below 0.01 in both cases). Hence, the data strongly support the view that, in each region, the most popular names tend to be the least Enemy Sounding. To show how to compute the ENI in practice, we list in Table A10 the 20 names with the lowest ENI among the top 1% most popular names in each region. We also report their frequencies in the South and in the North and their ENI. For example, Bettie was held by 0.26% of the population in the South and 0.04% of the population in the North; its ENI from the perspective of a Southern name holder is equal to 13.75 which makes it a very culturally distinctive name from this region.

As another way of putting the ENI in historical perspective, it is instructive to scrutinize the first names of US presidents elected between 1861 (Abraham Lincoln) and 1913 (Thomas Woodrow Wilson). Only 2 presidents came from Confederate states (Andrew Johnson and Thomas Woodrow Wilson). For both, the ENI indicates that their first names are clearly Confederate sounding. Out of the ten Northern presidents, the ENI identifies their first names as Unionist-sounding for 8 of them.¹⁴ While this exercise is only suggestive, it nonetheless reassures us that the ENI effectively captures a real geographical divide in the connotation of names.

¹³To reduce the impact of errors in transcription, we only consider names appearing more than 50 times.

¹⁴Thomas and Andrew have a Confederate ENI of 40.1 and 46.51 respectively. Among Northern presidents, Abraham, Chester and Theodore have Unionist ENI of 28.4, 27.8, and 29.3, respectively.

Table A10: List of Confederate and Unionist Names

<i>Name Holder from Confederate States</i>				<i>Name Holder from Unionist States</i>			
<i>Name</i>	<i>Frequency (%)</i>		<i>ENI</i>	<i>Name</i>	<i>Frequency (%)</i>		<i>ENI</i>
	<i>in South</i>	<i>in North</i>			<i>in North</i>	<i>in South</i>	
<i>Bettie</i>	0.26	0.04	13.75	<i>Harry</i>	0.67	0.15	17.87
<i>Lula</i>	0.30	0.07	19.98	<i>Grace</i>	0.34	0.10	23.29
<i>Sallie</i>	0.56	0.15	20.63	<i>Fred</i>	0.33	0.13	28.60
<i>Mollie</i>	0.25	0.08	23.67	<i>Charles</i>	1.61	0.66	29.12
<i>Mattie</i>	0.49	0.16	24.65	<i>Nellie</i>	0.42	0.18	29.53
<i>Charlie</i>	0.24	0.10	28.51	<i>Clara</i>	0.43	0.18	29.54
<i>Willie</i>	0.62	0.27	30.00	<i>Jacob</i>	0.32	0.15	31.15
<i>Nancy</i>	0.71	0.31	30.28	<i>Frank</i>	1.18	0.53	31.16
<i>Fannie</i>	0.44	0.19	30.34	<i>Edward</i>	0.78	0.38	32.88
<i>Martha</i>	0.99	0.55	35.62	<i>Anna</i>	0.71	0.36	33.94
<i>John W</i>	0.30	0.18	37.80	<i>Jennie</i>	0.47	0.24	34.24
<i>Robert</i>	0.80	0.51	38.75	<i>Clarence</i>	0.27	0.14	34.44
<i>Thomas</i>	1.00	0.67	40.20	<i>Bertha</i>	0.31	0.16	34.56
<i>Lucy</i>	0.35	0.25	41.20	<i>Arthur</i>	0.34	0.18	34.64
<i>Jane</i>	0.46	0.34	42.22	<i>Albert</i>	0.49	0.30	37.50
<i>James</i>	1.79	1.32	42.43	<i>George</i>	1.66	1.00	37.54
<i>Richard</i>	0.25	0.20	44.72	<i>Kate</i>	0.30	0.19	38.40
<i>Maggie</i>	0.43	0.35	44.84	<i>Florence</i>	0.40	0.25	38.72
<i>Julia</i>	0.35	0.29	45.06	<i>Daniel</i>	0.31	0.19	38.82
<i>Susan</i>	0.48	0.39	45.14	<i>Carrie</i>	0.34	0.22	39.50

Table A11: The table displays the 20 names with the lowest ENI, among the top 1% most popular names in each region. In the panel on the left, the most popular names and their respective ENI in former Confederate states, in the right panel, the most popular names and their respective ENI in former Unionist states. Frequencies measure the name absolute frequency in the population in the South (former Confederate states) and in the North (former Unionist states) of the country.

B Main robustness exercises common across outcomes

B.1 Extended Estimation Sample

As discussed in Section 2, the baseline estimation sample does not include observations from the (105) counties located in Kansas—the only state that legally enforced a ban of the movie BON. This section reports results obtained by including observations from Kansas. Results are displayed in Table B1 for all our main outcomes variables. Panel A displays OLS estimates, Panel B shows reduced form estimates obtained using MDM_{ct} as instrument, and Panel C contains 2SLS estimates. Overall, results are in line with the baseline coefficients reported in the main text. Table B2 displays the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Tables B3 and B4 report results of reduced form regressions that estimate the effect of MDM for the subsamples of counties located outside of Kansas (Panel A) and in Kansas (Panel B).¹⁵ In the former subsample, the coefficients are statistically significant at the conventional threshold in all specifications. In contrast, in the latter subsample, there are weak to null reactions to our MDM instrument. Only one out of five coefficients is statistically significant and its point estimate is much smaller than its counterpart estimated outside of Kansas (not only in absolute but also in relative terms when compared to the subsample mean). In other words, Kansas' counties that should have screened BON in absence of the ban (as predicted by MDM) did not experience a pervasive change in attitudes. While this result suggests that the enforcement of the ban in Kansas was quite effective, we believe that the interpretation should actually be more nuanced: As already discussed in Section C. (page 9) and in Online Appendix A.2, Kansas state authorities won the legal battle for the ban but evidence suggests that the population of Kansas remained partially exposed to the movie and its message. Indeed, there was a lot of resistance from local movie owners who appealed to the decision of the state and sometimes decided to screen the movie despite the ban (Stokes 2007). Figure A7, left panel, reports evidence of one illegal screening that took place in Topeka in June 1917, violating the ban. However, it is difficult to evaluate how pervasive illegal screenings had been because, to avoid sanctions, illegal screenings were probably not advertised on newspapers. Moreover, the political struggle that led to the ban raised such a turmoil in the state that the movie ended up receiving an immense media coverage. In order to meet this high demand for the movie from Kansans, movie owners from bordering counties in neighboring states intensively advertised screenings of the movie on Kansas' newspapers (see an example in Figure A7, right panel). Indeed, historians document that people traveled to watch Birth of a Nation to neighboring states, from Butters (2007): "It was an immense hit in Kansas City, and many who lived in the eastern part of Kansas crossed the state border to see the controversial film."

¹⁵Results in Panels B of Tables B3 and B4 are similar because the estimation sample is composed of only one state (Kansas). Hence, conditioning for month fixed effects or state-month fixed effects is equivalent.

Table B1: Extended estimation sample - Including Kansas

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.315 (0.023)	0.044 (0.005)	0.014 (0.006)	0.034 (0.005)	0.043 (0.005)
Panel B: Reduced Form					
Million Dollar Mystery	0.405 (0.022)	0.023 (0.004)	0.025 (0.006)	0.031 (0.005)	0.025 (0.005)
Panel C: 2SLS					
Birth of a Nation	1.363 (0.111)	0.076 (0.013)	0.052 (0.014)	0.103 (0.016)	0.084 (0.017)
1st Stage F-Stat	140	233	83	143	144
Observations	102,944	70,108	95,063	102,944	102,944

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. Estimates are obtained by adding counties located in Kansas to the baseline sample. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors are clustered at the county level.

Table B2: Extended estimation sample - Including Kansas - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.358 (0.023)	0.036 (0.005)	0.009 (0.006)	0.038 (0.006)	0.030 (0.006)
Panel B: Reduced Form					
Million Dollar Mystery	0.333 (0.024)	0.017 (0.004)	0.018 (0.007)	0.027 (0.005)	0.015 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.297 (0.109)	0.062 (0.013)	0.040 (0.016)	0.105 (0.019)	0.058 (0.021)
1st Stage F-Stat	123	213	81	123	123
Observations	102,232	69,904	95,061	102,232	102,232

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. Estimates are obtained by adding counties located in Kansas to the baseline sample. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacism in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

Table B3: Reduced form outside Kansas and in Kansas

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: Counties Outside Kansas					
Million Dollar Mystery	0.411 (0.025)	0.029 (0.005)	0.024 (0.006)	0.035 (0.005)	0.035 (0.006)
Dep. Var. Mean	-0.04	0.03	0.36	0.11	0.12
Dep. Var. Std. Dev.	0.93	0.18	0.48	0.31	0.32
Observations	89,325	62,968	91,612	89,325	89,325
Panel B: Counties in Kansas					
Million Dollar Mystery	0.285 (0.048)	0.004 (0.005)	0.007 (0.038)	0.004 (0.010)	-0.019 (0.012)
Dep. Var. Mean	0.26	0.02	0.30	0.10	0.08
Dep. Var. Std. Dev.	0.90	0.13	0.46	0.30	0.27
Observations	13,619	7,140	3,451	13,619	13,619

NOTE: The table reports OLS estimates. Estimates displayed in Panel A (B) are obtained using counties outside (in) Kansas. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B4: Reduced form outside Kansas and in Kansas - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: Counties Outside Kansas					
Million Dollar Mystery	0.336 (0.027)	0.019 (0.005)	0.018 (0.007)	0.031 (0.006)	0.021 (0.006)
Dep. Var. Mean	-0.04	0.03	0.36	0.10	0.11
Dep. Var. Std. Dev.	0.93	0.18	0.48	0.31	0.32
Observations	88,613	62,764	91,610	88,613	88,613
Panel B: Counties in Kansas					
Million Dollar Mystery	0.285 (0.048)	0.004 (0.005)	0.007 (0.038)	0.004 (0.010)	-0.019 (0.012)
Dep. Var. Mean	0.26	0.02	0.30	0.10	0.08
Dep. Var. Std. Dev.	0.90	0.13	0.46	0.30	0.27
Observations	13,619	7,140	3,451	13,619	13,619

NOTE: The table reports OLS estimates. Estimates displayed in Panel A (B) are obtained using counties outside (in) Kansas. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.2 Alternative Measures of Local Screenings of *Birth of a Nation*

In Section C. and in Appendix A.3 we discuss similarities and differences between our measure of local screening of *The Birth of a Nation* (denoted ERST hereafter) with the alternative measure (denoted Ang) collected in the companion paper Ang (2020).

This section contains estimates obtained with three alternative measures to identify the movie's distribution over time and space:

- 1 In Section B.2.1 we present results obtained using the measure of local screenings collected by Ang (2020).

We also consider more data-rich constructions of the treatment variable that combine information from both measures. We use several ways to combine information from both measures:

- 2 In the *union* approach, we define as treated all counties that are coded as treated by *at least one* of the two measures. Formally, we can define the treatment variable as $\text{Union}_{ct} = \max(\text{ERST}_{ct}, \text{Ang}_{ct})$. Corresponding results are displayed in Section B.2.2
- 3 In the *intersection* approach, we define as treated the counties that are coded as treated by *both* measures. Formally, the treatment variable is equal to $\text{Inter}_{ct} = \min(\text{ERST}_{ct}, \text{Ang}_{ct})$. Results obtained with this measure are reported in B.2.3.

In total, regarding our treatment variable, we are now equipped with 4 different measures of the local screening of Birth of a Nation: ERST, Ang, Union and Inter. Overall, all results are confirmed whatever measure of the treatment is used.

B.2.1 Local screening using data from Ang (2020)

Table B5: *The Birth of a Nation* - Ang measure of local screening

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation [Ang]	0.369 (0.024)	0.050 (0.006)	0.018 (0.006)	0.039 (0.005)	0.046 (0.006)
Panel B: 2SLS					
Birth of a Nation [Ang]	1.338 (0.109)	0.115 (0.019)	0.092 (0.028)	0.113 (0.017)	0.113 (0.019)
1st Stage F-Stat	136	139	24	139	142
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A) and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Birth of a Nation (Ang) is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section B.2 for additional 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B6: *The Birth of a Nation* - Ang measure of local screening - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation [Ang]	0.324 (0.026)	0.045 (0.006)	0.012 (0.006)	0.039 (0.006)	0.043 (0.006)
Panel B: 2SLS					
Birth of a Nation [Ang]	1.275 (0.135)	0.090 (0.020)	0.080 (0.035)	0.115 (0.022)	0.080 (0.022)
1st Stage F-Stat	84	93	20	84	85
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A) and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Birth of a Nation (Ang) is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section B.2 for additional 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Standard errors clustered at the county level. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample.

B.2.2 Local screening defined as Union of ERST and Ang

Table B7: *The Birth of a Nation* - Union measure of local screening

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation [Union]	0.370 (0.021)	0.042 (0.005)	0.018 (0.007)	0.035 (0.004)	0.035 (0.005)
Panel B: 2SLS					
Birth of a Nation [Union]	1.187 (0.085)	0.088 (0.014)	0.069 (0.019)	0.101 (0.015)	0.101 (0.017)
1st Stage F-Stat	181	234	58	182	186
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A) and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Birth of a Nation (Union) is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section B.2 for additional 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B8: *The Birth of a Nation* - Union measure of local screening - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation [Union]	0.320 (0.023)	0.038 (0.005)	0.010 (0.006)	0.035 (0.005)	0.029 (0.006)
Panel B: 2SLS					
Birth of a Nation [Union]	1.240 (0.120)	0.072 (0.016)	0.059 (0.025)	0.112 (0.021)	0.078 (0.022)
1st Stage F-Stat	100	144	40	101	101
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A) and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Birth of a Nation (Union) is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section B.2 for additional 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Standard errors clustered at the county level. Standard errors clustered at the county level. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample.

B.2.3 Local screening defined as Intersection of ERST and Ang

Table B9: *The Birth of a Nation* - Inter measure of local screening

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation [Inter]	0.412 (0.024)	0.051 (0.007)	0.018 (0.006)	0.042 (0.006)	0.050 (0.006)
Panel B: 2SLS					
Birth of a Nation [Inter]	1.219 (0.086)	0.097 (0.016)	0.060 (0.016)	0.104 (0.015)	0.103 (0.017)
1st Stage F-Stat	177	226	50	179	184
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A) and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Birth of a Nation (Inter) is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section B.2 for additional 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 B. for further details). Results displayed in Column 1 are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B10: *The Birth of a Nation* - Inter measure of local screening - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation [Inter]	0.369 (0.026)	0.045 (0.007)	0.011 (0.006)	0.043 (0.006)	0.046 (0.007)
Panel B: 2SLS					
Birth of a Nation [Inter]	1.150 (0.102)	0.072 (0.016)	0.048 (0.019)	0.104 (0.019)	0.072 (0.020)
1st Stage F-Stat	111	165	51	111	113
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A) and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Birth of a Nation (Inter) is an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section B.2 for additional 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 B. for further details). Results displayed in Column 1 are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.3 Alternative Measures of BON Exposure

This section presents results using alternative proxies of movie screenings. In our baseline analysis, we define BON_{ct} as an indicator variable that takes value 1 when a county reaches the threshold of 2 verified movie screening ads. Table B11 replicates baseline results with a threshold of 3, and Table B13 with a threshold of 5. Tables B12 and B14 display the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates). Increasing the threshold from 2 to 3 ads 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 BON when this truly happened, and p_0 as the probability that we observe a county to have screened BON when, in fact, BON 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:

$$\text{plim } \hat{\beta}_{IV} \equiv \left[\frac{\beta}{p_1 - p_0} \right] \quad (2)$$

In the presence of misclassification of the variable BON_{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 B11: Alternative Measures of Screening of the Movie: At Least 3 Ads

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation	0.415 (0.022)	0.047 (0.006)	0.021 (0.006)	0.039 (0.005)	0.039 (0.005)
Panel B: Reduced Form					
Million Dollar Mystery	0.418 (0.025)	0.028 (0.005)	0.025 (0.007)	0.035 (0.005)	0.037 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.071 (0.067)	0.075 (0.013)	0.048 (0.013)	0.090 (0.014)	0.093 (0.015)
1st Stage F-Stat	234	321	91	236	242
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. 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). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B12: Alternative Measures of Screening of the Movie: At Least 3 Ads - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation	0.378 (0.024)	0.042 (0.006)	0.013 (0.006)	0.043 (0.006)	0.033 (0.006)
Panel B: Reduced Form					
Million Dollar Mystery	0.346 (0.027)	0.018 (0.005)	0.017 (0.007)	0.031 (0.006)	0.023 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.089 (0.087)	0.056 (0.014)	0.039 (0.016)	0.097 (0.018)	0.072 (0.020)
1st Stage F-Stat	138	211	63	137	139
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. 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). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

Table B13: Alternative Measures of Screening of the Movie: At Least 5 Ads

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation	0.469 (0.023)	0.056 (0.006)	0.021 (0.006)	0.043 (0.005)	0.048 (0.006)
Panel B: Reduced Form					
Million Dollar Mystery	0.445 (0.026)	0.032 (0.005)	0.027 (0.007)	0.038 (0.006)	0.036 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.043 (0.059)	0.082 (0.013)	0.051 (0.013)	0.090 (0.013)	0.085 (0.015)
1st Stage F-Stat	276	326	95	276	284
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. 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). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B14: Alternative Measures of Screening of the Movie: At Least 5 Ads - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation	0.426 (0.024)	0.048 (0.007)	0.014 (0.006)	0.046 (0.007)	0.043 (0.007)
Panel B: Reduced Form					
Million Dollar Mystery	0.375 (0.028)	0.020 (0.005)	0.020 (0.007)	0.033 (0.006)	0.024 (0.007)
Panel C: 2SLS					
Birth of a Nation	1.028 (0.072)	0.060 (0.014)	0.043 (0.016)	0.090 (0.016)	0.067 (0.018)
1st Stage F-Stat	171	233	63	170	172
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. 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). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.4 War-related Robustness

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 perform two different exercises.

We re-estimate the baseline specifications on the subperiod strictly predating the US participation to the War. There are two main drawbacks related to this approach. On the one hand, shortening the time-windows of interest reduces the (post-treatment) support of our estimates, both at the intensive margin (for counties screening the movie before the war ends) and at the extensive margin (for counties screening the movie after the war ends). On the other hand, we must acknowledge the arbitrariness related to the definition of “participation to the War”. We identify two distinguishable breakpoints that shaped the US participation to the War: (i) when the US entered the War (April 6, 1917) and (ii) the battle of Cambrai (FR) when the American troops (11th Engineers) first participated in active fighting (November 30, 1917). The results of these robustness checks are reported in Tables B15 and B17. Tables B16 and B18 display the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Additionally, since the salience and popularity of WWI in the US could have varied over time and across counties between 1914 and 1918, we perform a robustness check where we explicitly measure attention in the press for WWI. In particular, we compute in any given county-month the *salience of WWI in the local public debate* using the frequencies of words related to conflict on European soil (“European War”, “World War” and “Great War”) in the corresponding county-month. Finally, we re-estimate the baseline regressions controlling for this measure of salience of WWI in local newspapers. Estimation results are reported in Table B19. The main virtue of this new robustness check, with respect to the previous two exercises, is that it enables us to control for the confounding effect of WWI without dropping observations from the estimation sample. Table B20 displays the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

All in all, the results of these robustness checks indicate that the the potential confounding effect of WWI is not driving our results.

Table B15: Sub-period strictly predating the entry of US in the War [April 1917]

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.326 (0.020)	0.016 (0.004)	0.010 (0.009)	0.015 (0.007)	0.011 (0.007)
Panel B: Reduced Form					
Million Dollar Mystery	0.313 (0.020)	0.009 (0.003)	0.013 (0.009)	0.019 (0.005)	0.015 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.100 (0.072)	0.031 (0.010)	0.029 (0.019)	0.066 (0.019)	0.050 (0.022)
1st Stage F-Stat	197	281	57	200	202
Observations	59,267	47,226	56,195	59,267	59,267

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The sample includes all months up to March 1917 (for names-related regressions the sample includes all cohorts up to the one born in 1916). The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B16: Sub-period strictly predating the entry of US in the War [April 1917] - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.279 (0.021)	0.017 (0.005)	-0.002 (0.010)	0.017 (0.009)	0.009 (0.009)
Panel B: Reduced Form					
Million Dollar Mystery	0.254 (0.022)	0.007 (0.003)	0.008 (0.009)	0.020 (0.006)	0.005 (0.007)
Panel C: 2SLS					
Birth of a Nation	1.185 (0.108)	0.027 (0.011)	0.023 (0.027)	0.093 (0.027)	0.023 (0.030)
1st Stage F-Stat	104	180	40	103	104
Observations	58,780	47,073	56,194	58,780	58,780

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The sample includes all months up to March 1917 (for names-related regressions the sample includes all cohorts up to the one born in 1916). The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

Table B17: Sub-period strictly predating first US participation in active fighting [December 1917]

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.459 (0.024)	0.032 (0.005)	0.013 (0.007)	0.031 (0.006)	0.027 (0.006)
Panel B: Reduced Form					
Million Dollar Mystery	0.406 (0.025)	0.017 (0.004)	0.018 (0.008)	0.029 (0.005)	0.022 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.284 (0.080)	0.050 (0.011)	0.037 (0.016)	0.092 (0.017)	0.068 (0.019)
1st Stage F-Stat	204	300	73	204	210
Observations	64,751	54,634	64,866	64,751	64,751

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The sample includes all months up to November 1917 (for names-related regressions the sample includes all cohorts up to the one born in 1917). The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B18: Sub-period strictly predating first US participation in active fighting [December 1917] - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation	0.400 (0.026)	0.029 (0.005)	0.004 (0.008)	0.033 (0.008)	0.023 (0.008)
Panel B: Reduced Form					
Million Dollar Mystery	0.324 (0.027)	0.012 (0.004)	0.015 (0.008)	0.027 (0.006)	0.009 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.332 (0.112)	0.042 (0.013)	0.036 (0.021)	0.109 (0.024)	0.039 (0.025)
1st Stage F-Stat	113	198	52	113	115
Observations	64,224	54,457	64,864	64,224	64,224

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The sample includes all months up to November 1917 (for names-related regressions the sample includes all cohorts up to the one born in 1917). The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

Table B19: Accounting for Time-Varying War-Related Controls

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.166 (0.017)	0.028 (0.005)	0.010 (0.007)	0.029 (0.004)	0.025 (0.004)
Panel B: Reduced Form					
Million Dollar Mystery	0.159 (0.019)	0.013 (0.005)	0.019 (0.007)	0.026 (0.005)	0.022 (0.006)
Panel C: 2SLS					
Birth of a Nation	0.677 (0.092)	0.043 (0.014)	0.054 (0.019)	0.099 (0.021)	0.085 (0.023)
1st Stage F-Stat	72	206	38	84	86
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. All columns also include time-varying flexible control for a measure of salience of WWI in local newspapers obtained using the log frequency of war-related keywords (i.e., "European War", "World War" and "Great War") observed in the corresponding county-month. Standard errors clustered at the county level.

Table B20: Accounting for Time-Varying War-Related Controls - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.147 (0.017)	0.026 (0.005)	0.003 (0.007)	0.033 (0.005)	0.021 (0.006)
Panel B: Reduced Form					
Million Dollar Mystery	0.128 (0.018)	0.009 (0.004)	0.014 (0.007)	0.025 (0.006)	0.011 (0.006)
Panel C: 2SLS					
Birth of a Nation	0.711 (0.122)	0.034 (0.016)	0.041 (0.023)	0.127 (0.030)	0.054 (0.031)
1st Stage F-Stat	46	152	43	52	53
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. All columns also include time-varying flexible control for a measure of salience of WWI in local newspapers obtained using the log frequency of war-related keywords (i.e., "European War", "World War" and "Great War") observed in the corresponding county-month. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.5 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 41% 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 33% of counties. We construct two additional instruments, $Traffic_{ct}$ and $Mary_{ct}$, mimicking the construction of our baseline instrument 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.

Tables B21 and B23 report 2SLS and reduced form estimates using as an instrument: $Mary_{ct}$ constructed based on the diffusion of the movie "*What Happened to Mary*" (with no check of judges), $Traffic_{ct}$ 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. Tables B22 and B24 display the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Table B21: Alternative Movies as Instrument: "*What Happened to Mary*"

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: Reduced Form					
What Happened to Mary	0.460 (0.026)	0.067 (0.007)	0.021 (0.006)	0.042 (0.006)	0.050 (0.007)
Panel B: 2SLS					
Birth of a Nation	1.094 (0.066)	0.147 (0.016)	0.038 (0.012)	0.101 (0.015)	0.120 (0.017)
1st Stage F-Stat	299	407	297	291	297
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports reduced form (in Panel A), and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Panel B and Panel C display RF and 2SLS results obtained using the indicator variable Mary_{ct} constructed based on the diffusion of the movie "*What Happened to Mary*" (with no check of judges). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B22: Alternative Movies as Instrument: "*What Happened to Mary*" - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: Reduced Form					
What Happened to Mary	0.389 (0.027)	0.048 (0.007)	0.011 (0.007)	0.040 (0.007)	0.041 (0.008)
Panel B: 2SLS					
Birth of a Nation	1.138 (0.089)	0.123 (0.018)	0.021 (0.014)	0.118 (0.022)	0.121 (0.024)
1st Stage F-Stat	157	239	203	155	155
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports reduced form (in Panel A), and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Panel B and Panel C display RF and 2SLS results obtained using the indicator variable Mary_{ct} constructed based on the diffusion of the movie "*What Happened to Mary*" (with no check of judges). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

Table B23: Alternative Movies as Instrument: "*Traffic in Souls*"

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: Reduced Form					
Traffic in Souls	0.485 (0.027)	0.058 (0.007)	0.015 (0.006)	0.039 (0.008)	0.057 (0.008)
Panel B: 2SLS					
Birth of a Nation	1.177 (0.071)	0.124 (0.015)	0.034 (0.013)	0.096 (0.018)	0.139 (0.020)
1st Stage F-Stat	340	417	131	331	343
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports reduced form (in Panel A), and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Panel B and Panel C display RF and 2SLS results obtained using the indicator variable Traffic_{ct} constructed based on the diffusion of the movie "*Traffic in Souls*" (with no check of judges). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B24: Alternative Movies as Instrument: "*Traffic in Souls*" - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: Reduced Form					
Traffic in Souls	0.402 (0.031)	0.040 (0.007)	0.006 (0.006)	0.033 (0.009)	0.049 (0.009)
Panel B: 2SLS					
Birth of a Nation	1.242 (0.104)	0.103 (0.017)	0.014 (0.015)	0.104 (0.028)	0.152 (0.029)
1st Stage F-Stat	162	223	135	161	163
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports reduced form (in Panel A), and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Panel B and Panel C display RF and 2SLS results obtained using the indicator variable Traffic_{ct} constructed based on the diffusion of the movie "*Traffic in Souls*" (with no check of judges). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.6 Instrumental Variable Strategy: Presence of Theaters

Ang (2020) proposes an instrumental variable strategy that relies on the presence of theaters in 1914 to instrument for whether BON was screened. Using Ang’s data on theaters, we adjust his instrument to our framework and construct a binary variable equal to 1 in counties with a movie theater in 1914 after the national release of Birth of a Nation in February 1915. In other terms, our new instrument corresponds to the variable: $\text{Theater}_c \times \text{Post BON}_t$, where Theater_c is an indicator of counties that had a theater in 1914 and Post BON_t is an indicator taking on value of 1 after February 1915. Table B25 presents reduced forms and 2SLS results with this new instrument. Table B26 repeats the same specifications with state-month fixed effects (state-year in the case of name-based estimates). The coefficient, in the reduced form analysis, allows to compare counties with a movie theater in 1914 and counties without a movie theater, looking at the change in outcomes before and after the national release of Birth of a Nation. Overall, we see that all our results based on our instrument (2SLS and reduced-form) are confirmed when using Ang’s instrument as an alternative. Finally, Table B27 shows the underlying relationship between BON, MDM and the new instrument based on theaters’ presence. Panel A displays the first stage and Panel B confirms that our instrument based on MDM and the new instrument highly correlates.

Table B25: Alternative Instruments: Using Presence of Theaters in the County

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: Reduced Form					
Theatre x Post-BON	0.356 (0.028)	0.036 (0.004)	0.028 (0.010)	0.028 (0.005)	0.034 (0.005)
Panel B: 2SLS					
Birth of a Nation	0.766 (0.052)	0.106 (0.011)	0.061 (0.027)	0.060 (0.010)	0.072 (0.010)
1st Stage F-Stat	323	409	13	327	331
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports reduced form (in Panel A), and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words’ log frequencies minus the first principal component of divisive words’ log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Panel A and Panel B display RF and 2SLS results obtained using the indicator variable $\text{Theater} \times \text{Post-BON}$ constructed using a variable for the presence of theaters in the county (Theater) and a variable (Post BON) that indicates the period after the official release of BON. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B26: Alternative Instruments: Using Presence of Theaters in the County - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: Reduced Form					
Theatre x Post-BON	0.275 (0.028)	0.027 (0.004)	0.017 (0.010)	0.021 (0.006)	0.019 (0.006)
Panel B: 2SLS					
Birth of a Nation	0.739 (0.070)	0.097 (0.013)	0.042 (0.028)	0.056 (0.015)	0.051 (0.015)
1st Stage F-Stat	164	219	23	166	165
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports reduced form (in Panel A), and 2SLS (in Panel B) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. for details). Panel A and Panel B display RF and 2SLS results obtained using the indicator variable $Theater \times Post-BON$ constructed using a variable for the presence of theaters in the county ($Theater$) and a variable ($Post-BON$) that indicates the period after the official release of BON. Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

Table B27: Relationship between BON, MDM and Theater x Post-BON

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: BON and Theatre x Post-BON					
Theatre x Post-BON	0.465 (0.026)	0.341 (0.017)	0.455 (0.122)	0.465 (0.026)	0.470 (0.026)
Panel B: MDM and Theatre x Post-BON					
Theatre x Post-BON	0.436 (0.035)	0.386 (0.025)	0.376 (0.119)	0.437 (0.035)	0.441 (0.035)
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS. In Panel A (Panel B) the dependent variable is Birth of a Nation (Million Dollar Mystery), an indicator variable that takes a value of 1 after the movie was screened in the county and 0 otherwise (see Section C. for details). The indicator variable Theater x Post-BON is constructed using a variable for the presence of theaters in the county (Theater) and a variable (Post BON) that indicates the period after the official release of BON. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

B.7 Falsification Exercises

In this section, we perform two placebo tests in support of the exclusion restriction assumption of the instrumental variable strategy. These falsification exercises leverage on the variations over respectively space and time of the screenings of MDM and BON; their shared goal is to show that the instrument (MDM) impacted attitudes towards reconciliation only through the treatment (BON). Overall, the analysis confirms that the treatment effect is really linked to the content of BON and does not come from exposure to any film circulated nationally which might have potentially created a sense of national bond regardless of the content of the film.

The first exercise rests on cross-county mismatches between the instrument and the treatment. Concretely, MDM screening is a strong, but not perfect, predictor of BON screening; several counties that were exposed to MDM did not screen BON (see Section B.). Hence, we can test whether MDM, as a falsified treatment, impacted attitudes in all counties or, instead, only in counties that also screened BON.

The test is designed in a very simple way. For each outcome variable, we estimate a modified version of the baseline Reduced Form specification (e.g. col. 2, panel A in Table 3) where the instrumental variable (transposed Million Dollar Mystery) is interacted with two indicator variables, Ever BON and Never BON, coding respectively for counties that screened at some point BON or never screened BON. The results are displayed in Table B28. A consistent pattern emerges across all columns: In counties that ever screened BON, the effect of MDM is statistically significant at the 1% threshold and quantitatively close to its baseline counterpart; by contrast, in counties that were not reached by BON during our sample period, the effect is statistically not different from zero at the 5% threshold and its magnitude is small. Overall, these findings provide further support to the validity of our exclusion restriction: The instrument impacted attitudes only through the screening of BON. Table B29 displays the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Table B28: Effect of exposure to MDM in BON-counties and in Never BON-counties

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Million Dollar Mystery $\times$ Ever BON	0.487 (0.026)	0.038 (0.006)	0.027 (0.007)	0.042 (0.006)	0.045 (0.007)
Million Dollar Mystery $\times$ Never BON	0.050 (0.035)	-0.004 (0.005)	0.004 (0.014)	-0.001 (0.005)	-0.010 (0.006)
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 B. for further details). Ever BON and Never BON, coding respectively for counties that screened at some point BON or never screened BON. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Our second falsification exercise exploits the differences in the national release dates of MDM (July 1914) and BON (February 1915). The idea is to show that MDM impacted attitudes toward reconciliation *only after* the release of BON, no effect being detected *before*. To some extent, this exercise resembles our event-study analysis (e.g. Figure 3) that already shows, in the data, the absence of

Table B29: Effect of exposure to MDM in BON-counties and in Never BON-counties - Accounting for state-period fixed effects

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Million Dollar Mystery x Ever BON	0.416 (0.028)	0.026 (0.005)	0.021 (0.007)	0.040 (0.006)	0.032 (0.007)
Million Dollar Mystery x Never BON	0.018 (0.035)	-0.006 (0.005)	-0.003 (0.014)	-0.005 (0.006)	-0.020 (0.007)
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 B. for further details). Ever BON and Never BON, coding respectively for counties that screened at some point BON or never screened BON. Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

any change in attitudes before treatment. However, the event-studies are based on the screening of BON (true treatment) whereas this exercise is based on the screening of MDM (falsified treatment). We operationalize our idea in the following way. We first define Ever MDM, an indicator function that codes for counties where MDM was screened at some point. Then, we build the dummy variables related to two non-overlapping time-windows: *interim* codes for the period elapsed between the official releases of MDM and BON; *post BON* codes for the period after the official release of BON. Considering dates of official release rather than dates of de facto screening of the movies alleviates potential concern related to the endogeneity of screenings. Finally, equipped with these new variables, we replicate for each outcome variable the baseline Reduced Form regression where the instrument is replaced by the interactions of Ever MDM with these two dummy variables. We expect that only the coefficient of the interaction with *post BON* loads positively; the interaction with *interim* should not be statistically different from zero at conventional levels. Note that the linear term Ever MDM is absorbed by the county fixed-effect.

This second exercise comes with two limitations. The main one is that the release of Million Dollar Mystery, and its diffusion throughout the territory, coincide with the outbreak of World War I in Europe. As explained in Section D., the war triggered a surge of patriotism in the US. Hence, during the interim period, the impact of MDM on attitudes potentially confounds with the one of war. A way to cope with this issue is to include the same set of control variables as the one used in Table B19. Another limitation pertains to the time structure of the data on names: The ENI varies at the yearly level because censuses do not report individuals' month of birth. As a consequence, and for this outcome variable only, we are forced to define the two non-overlapping time windows in a coarse way with *interim* coding for the year 1914 and *post BON* coding for all years after 1915. With these caveats in mind, we report the estimation results in Table B30. Our estimates show that the impact of MDM in fostering reconciliation is negligible before the official release of BON. By contrast, the impact is positive and statistically significant after the official release of BON in February 1915. Table B31 displays the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Table B30: Effect of MDM over different release windows

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Ever MDM x Post BON	0.180 (0.021)	0.016 (0.005)	0.029 (0.008)	0.027 (0.006)	0.021 (0.006)
Ever MDM x Interim	-0.015 (0.019)	-0.008 (0.003)	0.018 (0.012)	0.006 (0.009)	-0.004 (0.008)
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Ever MDM indicates counties that screened MDM at some point. We defined two indicator variables related to two non-overlapping time-windows: Interim codes for the period elapsed between the official releases of MDM and BON; Post BON codes for the period after the official release of BON. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. All columns include all additional war-related controls used in Table B19. Standard errors clustered at the county level.

Table B31: Effect of MDM over different release windows - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Ever MDM x Post BON	0.149 (0.020)	0.011 (0.004)	0.025 (0.008)	0.025 (0.006)	0.009 (0.007)
Ever MDM x Interim	0.006 (0.020)	-0.006 (0.003)	0.033 (0.014)	-0.002 (0.009)	-0.001 (0.009)
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). Ever MDM indicates counties that screened MDM at some point. We defined two indicator variables related to two non-overlapping time-windows: Interim codes for the period elapsed between the official releases of MDM and BON; Post BON codes for the period after the official release of BON. Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. All columns include all additional war-related controls used in Table B19. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.8 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, X_c : a lower share of male inhabitants, a lower share of illiterate population, a slightly lower share of Blacks, 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 BON_{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, 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 BON (at some point in time) and other counties. Table B32 contains regression results for specifications allowing the effect of X_c to vary at month-year level (or year level in the case of name-based estimates). Table B33 displays the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Table B32: Accounting for Time-Varying Confounders

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.287 (0.022)	0.008 (0.004)	0.012 (0.007)	0.031 (0.005)	0.025 (0.005)
Panel B: Reduced Form					
Million Dollar Mystery	0.318 (0.024)	0.005 (0.003)	0.020 (0.006)	0.029 (0.005)	0.024 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.136 (0.102)	0.016 (0.011)	0.048 (0.017)	0.103 (0.020)	0.083 (0.022)
1st Stage F-Stat	122	202	52	122	124
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. All columns also include time-varying flexible controls for the following county-level variables: share of male inhabitants, share of illiterate population, share of black individuals, share of people living in cities of more than 25 thousands inhabitants, and share of foreign population observed in the county. Standard errors clustered at the county level.

Table B33: Accounting for Time-Varying Confounders - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.279 (0.024)	0.014 (0.004)	0.010 (0.007)	0.031 (0.005)	0.025 (0.006)
Panel B: Reduced Form					
Million Dollar Mystery	0.283 (0.025)	0.005 (0.003)	0.018 (0.007)	0.027 (0.006)	0.016 (0.006)
Panel C: 2SLS					
Birth of a Nation	1.098 (0.111)	0.018 (0.012)	0.044 (0.017)	0.105 (0.023)	0.063 (0.023)
1st Stage F-Stat	100	176	81	100	101
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. All columns also include time-varying flexible controls for the following county-level variables: share of male inhabitants, share of illiterate population, share of black individuals, share of people living in cities of more than 25 thousands inhabitants, and share of foreign population observed in the county. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

B.9 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 B34 presents results with standard errors adjusted for two-dimensional spatial dependence, using cutoff thresholds of 50, 100, and 200 km and using standard errors clustered at the state level.

Table B34: Conley Standard Errors

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: OLS					
Birth of a Nation	0.400	0.041	0.017	0.036	0.037
	(0.021)	(0.005)	(0.006)	(0.005)	(0.005)
<i>State-Level Cluster</i>	0.032	0.007	0.005	0.006	0.005
<i>Spatial Cluster 50km</i>	0.007	0.003	0.006	0.003	0.003
<i>Spatial Cluster 100km</i>	0.008	0.003	0.006	0.003	0.003
<i>Spatial Cluster 200km</i>	0.011	0.004	0.006	0.004	0.003
Panel B: Reduced Form					
Million Dollar Mystery	0.410	0.029	0.024	0.035	0.035
	(0.025)	(0.005)	(0.006)	(0.005)	(0.006)
<i>State-Level Cluster</i>	0.031	0.006	0.008	0.006	0.007
<i>Spatial Cluster 50km</i>	0.007	0.002	0.006	0.003	0.003
<i>Spatial Cluster 100km</i>	0.008	0.003	0.006	0.003	0.003
<i>Spatial Cluster 200km</i>	0.010	0.003	0.006	0.004	0.003
Panel C: 2SLS					
Birth of a Nation	1.092	0.076	0.049	0.093	0.093
	(0.071)	(0.012)	(0.013)	(0.014)	(0.015)
<i>State-Level Cluster</i>	0.095	0.014	0.016	0.017	0.020
<i>Spatial Cluster 50km</i>	0.020	0.006	0.012	0.009	0.009
<i>Spatial Cluster 100km</i>	0.022	0.007	0.012	0.009	0.009
<i>Spatial Cluster 200km</i>	0.028	0.007	0.013	0.010	0.009

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. Estimates are obtained by adding counties located in Kansas to the baseline sample. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. 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 B. for further details). Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. The table reports standard errors clustered at the county level, standard errors clustered at the state level and adjusted for two-dimensional spatial dependence using cutoff thresholds of 50, 100, and 200km.

B.10 A Pervasive Reconciliation

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 5.

Table B35 replicates baseline results on our main outcome variables, separating the effect between former Confederate states and former Union states.

Table B35: *The Birth of a Nation* in Former Confederate States and Former Union States

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: OLS					
Birth of a Nation * Unionist	0.427 (0.025)	0.050 (0.007)	0.025 (0.007)	0.035 (0.005)	0.044 (0.006)
Birth of a Nation * Confederate	0.331 (0.033)	0.016 (0.008)	-0.016 (0.009)	0.040 (0.011)	0.018 (0.009)
Panel B: Reduced Form					
Million Dollar Mystery * Unionist	0.453 (0.029)	0.040 (0.006)	0.028 (0.007)	0.033 (0.006)	0.042 (0.007)
Million Dollar Mystery * Confederate	0.294 (0.039)	0.001 (0.005)	0.005 (0.011)	0.041 (0.010)	0.017 (0.009)
Panel C: 2SLS					
Birth of a Nation * Unionist	1.099 (0.070)	0.089 (0.013)	0.050 (0.013)	0.087 (0.013)	0.097 (0.015)
Birth of a Nation * Confederate	1.063 (0.096)	0.031 (0.014)	0.028 (0.022)	0.119 (0.022)	0.074 (0.021)
1st stage F-Stat BON * Unionist	233	366	88	234	240
1st stage F-Stat BON * Confederate	165	204	135	166	167
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS (in Panel A), reduced form (in Panel B), and 2SLS (in Panel C) estimates. Estimates are obtained by adding counties located in Kansas to the baseline sample. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. 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. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors are clustered at the county level.

B.11 Heterogeneous Effect of BON

In this section we investigate whether population characteristics affect the magnitude of the treatment effect. In Table B36 we estimate the baseline regressions allowing for the coefficient of the treatment effect to differ between counties with a share of African-American population (measured in 1910) lower or higher than the median. For all outcome variables, these two coefficients are similar in magnitude and precisely estimated in both sets of counties. The results suggest that, in counties with more African-Americans, the movie engendered a larger increase in white supremacy and racial discrimination. In Table B37, we repeat this exercise considering the share of illiterate population (measured in 1919) as the relevant dimension of heterogeneity. Again, the movie engendered significant and quantitatively similar effects independently of the literacy level. In Table B38 and B39 we look at newspaper circulation, measured respectively in terms of number of newspapers and size of readership (following Gentzkow, Shapiro, and Sinkinson 2011).¹⁶ Here again the treatment effect is stable along these dimensions of heterogeneity with some evidence that it is larger in places with higher newspapers' circulation. Possibly, this pattern is driven by a stronger echo chamber effect associated to the release of the movie in these counties. Last, in Table B40 we consider the role of pre-treatment social tensions as measured by our variable $Divisive_{c,1910}$ in 1910. For all outcome variables, the treatment effect is larger in places with lower initial levels of divisiveness: in places with higher levels of initial divisiveness, the screening of BON brings about an effect on reconciliation discourse that is half the size of the effect in less divisive counties. Moreover, in more divisive counties there is no statistically significant effect of the movie on enlistment. Interestingly, the impact on cultural convergence - measured by looking at naming choices - is the same in the two types of counties. Hence, cultural convergence might have permeated into traditionally hostile areas, even before any significant change in rhetoric could materialize.

Table B36: Share of Blacks in 1910 (2SLS)

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Birth of a Nation * Low Share of Blacks	1.044 (0.071)	0.064 (0.013)	0.055 (0.014)	0.076 (0.013)	0.089 (0.015)
Birth of a Nation * Above Share of Blacks	1.163 (0.084)	0.096 (0.017)	0.041 (0.014)	0.116 (0.017)	0.098 (0.019)
1st stage F-Stat BON * Low	253	351	120	253	260
1st stage F-Stat BON * High	204	267	90	206	209
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports 2SLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $Reconciliation_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($Supremacism_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($Discrimination_{c,t}$) in Column (5). 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 C. for details). Low Share of Blacks (High Share of Blacks) is an indicator variable that takes value 1 in counties that have a share of Black population below (above) the median in 1910. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

¹⁶Low Number of Dailies (High Number of Dailies) is an indicator variable equal to 1 in counties with a number of English daily newspapers below (above) the median over the period 1900-1915. Low Newspaper Circulation (High Newspaper Circulation) is an indicator variable equal to 1 in counties with total newspaper circulation below (above) the median; in this case, circulation is computed using the total circulation in the county-year and the estimated number of eligible presidential election voters.

Table B37: Illiteracy in 1910 (2SLS)

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Birth of a Nation * Low Illiteracy	1.072 (0.071)	0.089 (0.014)	0.055 (0.013)	0.086 (0.013)	0.100 (0.016)
Birth of a Nation * High Illiteracy	1.137 (0.087)	0.050 (0.015)	0.037 (0.015)	0.108 (0.019)	0.078 (0.018)
1st stage F-Stat BON * Low	255	371	131	256	263
1st stage F-Stat BON * High	180	239	86	181	183
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports 2SLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Low Illiteracy (High Illiteracy) is an indicator variable that takes value 1 in counties that have a share of illiterate population below (above) the median in 1910. Results displayed in Column 1 are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B38: Number of English-language daily newspaper (2SLS)

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Birth of a Nation * Low Number of Daily Newspapers	0.841 (0.120)	0.010 (0.023)	0.052 (0.019)	0.076 (0.017)	0.051 (0.018)
Birth of a Nation * Above Number of Daily Newspapers	1.106 (0.065)	0.084 (0.012)	0.048 (0.013)	0.094 (0.014)	0.095 (0.015)
1st stage F-Stat BON * Low	92	109	137	93	94
1st stage F-Stat BON * High	202	375	91	206	209
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports 2SLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Low Number of Dailies (High Number of Dailies) is an indicator variable variable that takes a value of 1 in counties that have a value of number of English-language daily newspapers below (above) the median in the period 1900-1915. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B39: Total Newspaper Circulation (2SLS)

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Birth of a Nation * Low Newspaper Circulation	0.865 (0.117)	0.014 (0.021)	0.054 (0.018)	0.069 (0.017)	0.054 (0.019)
Birth of a Nation * Above Newspaper Circulation	1.112 (0.065)	0.086 (0.012)	0.047 (0.013)	0.095 (0.013)	0.096 (0.015)
1st stage F-Stat BON * Low	100	128	138	101	101
1st stage F-Stat BON * High	212	381	95	215	218
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports 2SLS estimates. (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Low Newspaper Circulation (High Newspaper Circulation) is an indicator variable variable that takes a value of 1 in counties that have a value of total newspaper circulation (computed using the total circulation in the county-year and the estimated number of eligible presidential election voters) below (above) the median in the period 1900-1915. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B40: Initial Levels of $\text{Divisive}_{c,1910}$ (2SLS)

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Birth of a Nation * Low Divisive_c	1.029 (0.069)	0.074 (0.012)	0.049 (0.013)	0.089 (0.014)	0.084 (0.015)
Birth of a Nation * Above Divisive_c	0.612 (0.166)	-0.013 (0.036)	0.047 (0.020)	0.064 (0.026)	0.026 (0.026)
1st stage F-Stat BON * Low	53	144	89	53	54
1st stage F-Stat BON * High	24	29	105	24	24
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports 2SLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Low $\text{Divisive}_{c,1910}$ (High $\text{Divisive}_{c,1910}$) is an indicator variable variable that takes a value of 1 in counties that have a value of $\text{Divisive}_{c,1910}$ below (above) the median in 1910. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

B.12 Spatial Spillovers

Table B41 displays the results of an augmented version of our baseline specification: Besides our treatment variable, that codes for the local screening of the movie in the county, we include a measure of distant screening taking place in neighboring counties using three distance cutoff to define the neighboring counties.

Estimates presented in Table B41 suggest that attitudes weakly react to screening taking place in neighboring counties. For four outcome variables we indeed detect no spatial spillover. Only in the case of the text-based measure of reconciliation, we see that the coefficient of distant screening is statistically significant; nevertheless, its magnitude is significantly smaller than the coefficient of local screening. More importantly, for all outcome variables, the inclusion of distant screening barely affects the coefficient of local screening with respect to its counterpart in the baseline analysis. This stability of the treatment effect makes us conclude that screening in neighboring counties is unlikely to induce a meaningful bias in our baseline analysis. Table B42 displays the estimates obtained with state-month fixed effects (state-year in the case of name-based estimates).

Table B41: Spatial Spillovers

	(1) Reconciliation	(2) Enlistment	(3) ENI	(4) Supremacism	(5) Discrimination
Panel A: 30km Radius					
Birth of a Nation	0.393 (0.021)	0.040 (0.005)	0.016 (0.006)	0.036 (0.005)	0.036 (0.005)
Birth of a Nation within 30km radius	0.163 (0.053)	0.035 (0.015)	0.005 (0.010)	0.007 (0.011)	0.022 (0.013)
Panel B: 40km Radius					
Birth of a Nation	0.397 (0.021)	0.041 (0.005)	0.016 (0.006)	0.036 (0.005)	0.036 (0.005)
Birth of a Nation within 40km radius	0.080 (0.029)	0.005 (0.007)	0.008 (0.007)	-0.002 (0.006)	0.013 (0.007)
Panel C: 50km Radius					
Birth of a Nation	0.394 (0.021)	0.041 (0.005)	0.016 (0.006)	0.037 (0.005)	0.036 (0.005)
Birth of a Nation within 50km radius	0.088 (0.024)	0.003 (0.005)	0.008 (0.007)	-0.006 (0.005)	0.007 (0.006)
Observations	89,325	62,968	91,612	89,325	89,325

NOTE: The table reports OLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Birth of a Nation within radius is an indicator variable that takes a value of 1 if BON has been screened in a county within in a given radius. In Panel A (B) [C] [D] we display results obtained considering a 20km radius (30) [40] [50]. Results displayed in Column (1) are obtained using the set of controls used in Panel A of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel A of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel A of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel A in Table 6. Standard errors clustered at the county level.

Table B42: Spatial Spillovers - Accounting for state-period fixed effects

	(1)	(2)	(3)	(4)	(5)
	Reconciliation	Enlistment	ENI	Supremacism	Discrimination
Panel A: 30km Radius					
Birth of a Nation	0.354 (0.023)	0.036 (0.005)	0.009 (0.006)	0.036 (0.006)	0.031 (0.006)
Birth of a Nation within 30km radius	0.097 (0.062)	0.024 (0.014)	-0.009 (0.009)	0.017 (0.012)	0.007 (0.015)
Panel B: 40km Radius					
Birth of a Nation	0.356 (0.023)	0.036 (0.005)	0.009 (0.006)	0.037 (0.006)	0.031 (0.006)
Birth of a Nation within 40km radius	0.007 (0.033)	0.007 (0.007)	-0.002 (0.007)	-0.001 (0.006)	-0.000 (0.008)
Panel C: 50km Radius					
Birth of a Nation	0.356 (0.023)	0.036 (0.005)	0.009 (0.006)	0.037 (0.006)	0.031 (0.006)
Birth of a Nation within 50km radius	0.015 (0.027)	0.002 (0.005)	0.007 (0.007)	-0.008 (0.005)	-0.002 (0.007)
Observations	88,613	62,764	91,610	88,613	88,613

NOTE: The table reports OLS estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Column 1), an indicator function taking value 1 if at least one person was enlisted in county c and month-year t (Column 2), the binarized version of the ENI (Column 3), an indicator variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Column (4), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Column (5). 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 C. for details). Birth of a Nation within radius is an indicator variable that takes a value of 1 if BON has been screened in a county within in a given radius. In Panel A (B) [C] [D] we display results obtained considering a 20km radius (30) [40] [50]. Results displayed in Column (1) are obtained using the set of controls used in Panel C of Table 3. Results displayed in Column (2) are obtained using the set of controls used in Panel B of Table 4. Results displayed in Column (3) are obtained using the set of controls used in Panel B of Table 5. Results displayed in Columns (4) and (5) are obtained using the set of controls used in Columns (1) and (3) of Panel B in Table 6. Due to the inclusion of state-period fixed effects, number of observation may be reduced as singleton observations are dropped from the sample. Standard errors clustered at the county level.

C The Birth of a Nation and the Public Debate: Extended Results

C.1 Alternative Construction of the Dependent Variable

We perform several exercises related to alternative constructions of the dependent variable. We change the scale of the +1 transformation in the construction of log frequencies by considering a +0.01 transformation in Table C1 and a +0.1 transformation in Table C2. In Table C3, we look at the inverse hyperbolic sine transformation. In Table C4, the dependent variable is the total sum of frequencies across all keywords (in place of the PCA).

Table C1: Robustness of the Logarithmic Transformation - $\log(0.01 + \text{Freq}_{i,ct})$

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	0.371 (0.031)		1.004 (0.093)	0.301 (0.022)		0.767 (0.068)	-0.070 (0.022)		-0.237 (0.058)
Million Dollar Mystery		0.376 (0.035)			0.288 (0.025)			-0.089 (0.022)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	1.187	1.187	1.187	2.217	2.217	2.217	2.569	2.569	2.569
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (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 $\log(0.01 + \text{Freq}_{i,c,t})$. 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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

Table C2: Robustness of the Logarithmic Transformation - $\log(0.1 + \text{Freq}_{i,c,t})$

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	0.360 (0.024)		0.984 (0.075)	0.316 (0.023)		0.779 (0.069)	-0.044 (0.021)		-0.204 (0.058)
Million Dollar Mystery		0.369 (0.028)			0.292 (0.026)			-0.077 (0.021)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	1.114	1.114	1.114	2.885	2.885	2.885	3.341	3.341	3.341
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (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 $\log(0.1 + \text{Freq}_{i,c,t})$. 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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

Table C3: 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	0.394 (0.022)		1.084 (0.072)	0.341 (0.024)		0.844 (0.072)	-0.053 (0.018)		-0.239 (0.051)
Million Dollar Mystery		0.406 (0.025)			0.317 (0.027)			-0.090 (0.019)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	0.986	0.986	0.986	3.545	3.545	3.545	3.973	3.973	3.973
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of the sine hyperbolic transformation of patriotic words' frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of the sine hyperbolic transformation of divisive words' frequencies, $\text{Divisive}_{c,t}$ (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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

Table C4: 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	1.997 (0.121)		5.410 (0.387)	1.688 (0.141)		4.058 (0.412)	-0.309 (0.093)		-1.352 (0.264)
Million Dollar Mystery		2.029 (0.138)			1.522 (0.156)			-0.507 (0.096)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	6.269	6.269	6.269	18.061	18.061	18.061	21.570	21.570	21.570
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the sum of patriotic words' frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the sum of divisive words' frequencies, $\text{Divisive}_{c,t}$ (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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

C.2 Alternative Measures of Rhetoric

C.2.1 Alternative Lists of Keywords

This section replicates results presented in Table 3 on the effect of *The Birth of a Nation* on reconciliation rhetoric presenting robustness exercises on the lists of words. Table C5 shows results obtained using the list of *Patriotic* and *Divisive* seed words. Table C6 uses the top 15 keywords of the lists, as opposed to the top 20 keywords in the baseline. Finally, Table C7 shows that results are not sensitive to the exclusion of “Americans” and “United States” patriotic keywords. These two keywords, as shown in Table A8, appear with a significantly higher frequency than the other words in the list. Last, Figure C1 allows to compare estimates on measures of reconciliation rhetoric constructed with subsets of words, and namely retaining the top # keywords based on the Semantic Index of Tables A3 and A4 (see Section A. for details).

Table C5: Alternative Lists of Keywords - Seed words

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	0.326 (0.027)		0.848 (0.079)	0.257 (0.034)		0.553 (0.094)	-0.068 (0.021)		-0.294 (0.059)
Million Dollar Mystery		0.318 (0.029)			0.208 (0.036)			-0.110 (0.021)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	1.838	1.838	1.838	2.341	2.341	2.341	3.502	3.502	3.502
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words’ log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words’ log frequencies, $\text{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county (c) in a particular month-year (t). The *Patriotic* list includes the patriotic seed words: “patriotic”; “united country”; “Americans together”; “American people”; “U.S. flag”; “Stars and Stripes”; “Americans”; “United States”; “The Star-Spangled Banner”; “reconciliation”; “American Revolution”; “fraternity”. The *Divisive* list includes the divisive seed words: “secession”; “civil war”; “slavery”; “confederates”; “sectionalism”; “Confederate Flag”; “Confederacy”; “Secession flag”; “Southerners”; “carpetbaggers”; “Northerners”; “yankee”; “Lost Cause”; “dixie”; “scalawags”; “bloody shirt”; “Stars and Bars”. 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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

Table C6: Alternative Lists of Keywords - Top 15

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	0.347 (0.019)		0.953 (0.064)	0.305 (0.022)		0.746 (0.065)	-0.042 (0.016)		-0.207 (0.045)
Million Dollar Mystery		0.357 (0.022)			0.280 (0.024)			-0.078 (0.016)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	0.896	0.896	0.896	3.123	3.123	3.123	3.519	3.519	3.519
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

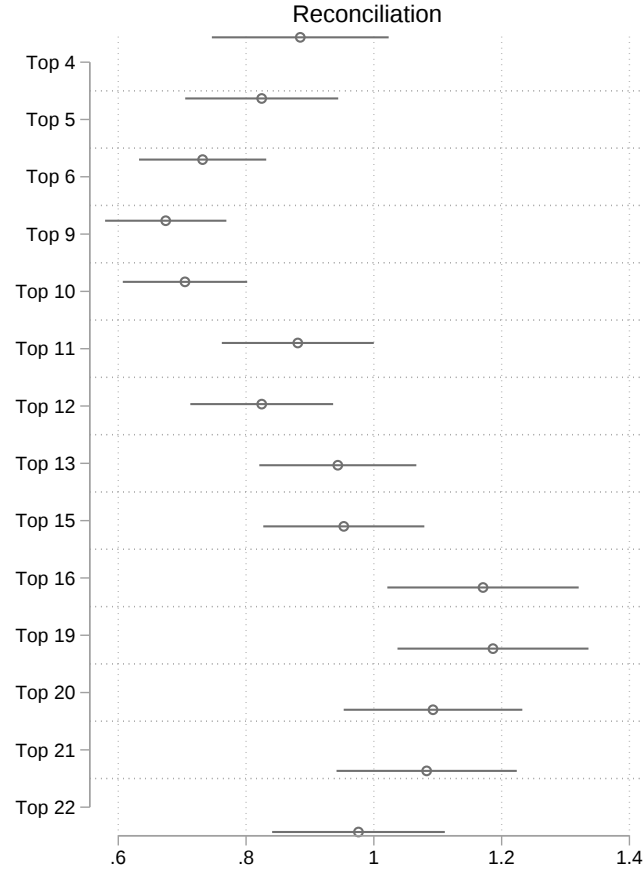
NOTE: The table reports OLS (Columns 1, 4, and 7), reduced form (Columns 2, 5, and 8), and 2SLS (Columns 3, 6, and 9) estimates. The dependent variables are the first principal component of patriotic words' log frequencies minus the first principal component of divisive words' log frequencies, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county (c) in a particular month-year (t). The *Patriotic* list includes 15 words with the highest value of IndexPatriotic constructed using judges' valuations: "patriotic", "united country", "Americans together", "American people", "U.S. flag", "Stars and Stripes", "Americans", "United States", "American flag", "patriotism", "true patriot", "liberty and freedom", "national hymn", "national salute", "salute flag". The *Divisive* list includes 15 words with the lowest value of IndexPatriotic constructed using judges' valuations: "General Lee", "Secession flag", "Confederacy", "Confederate Flag", "sectionalism", "confederates", "civil war", "secession", "slavery", "secede", "secessionist", "unpatriotic", "negro domination", "race problem", "armed conflict". 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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

Table C7: Alternative Lists of Keywords - Exclude common words

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	0.401 (0.021)		1.098 (0.070)	0.345 (0.022)		0.855 (0.068)	-0.056 (0.017)		-0.243 (0.047)
Million Dollar Mystery		0.412 (0.025)			0.321 (0.026)			-0.091 (0.017)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Std. Dev.	0.910	0.910	0.910	3.736	3.736	3.736	4.121	4.121	4.121
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (Columns 7 to 9). The unit of observation is the county (c) in a particular month-year (t). The *Patriotic* list includes the same keywords as the baseline list, with the exclusion of "Americans" and "United States". The *Divisive* list includes the same keywords as the baseline list. 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 C. 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 B. for further details). All regressions control for county, month-year, and coverage percentile fixed effects. Standard errors clustered at the county level.

Figure C1: Measures of Reconciliation Constructed with Subsets of Keywords



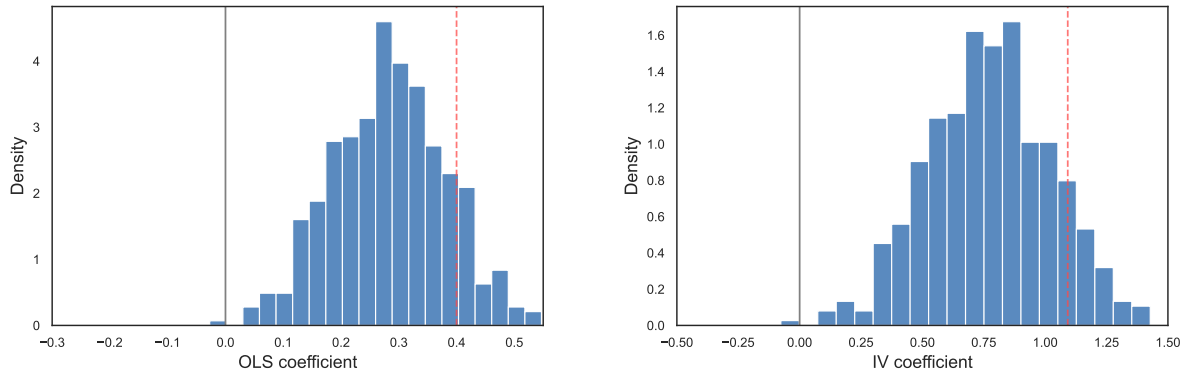
NOTE: The figure plots coefficients for the effect of BON_{ct} on various measures of $Reconciliation_{c,t}$. Specifications correspond to the one in Table 3, Panel A, Column 3. The measures of $Reconciliation_{c,t}$ are constructed retaining the top # keywords based on the Semantic Index of Tables A3 and A4 (see Section A. for details). Note that, since several words share the same Semantic Index and cannot be ranked, certain bundles (e.g. Top 2) cannot be well defined and are therefore excluded.

C.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 analysis. We then reconstruct 500 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 500 different versions of the variable $\text{Reconciliation}_{c,t}$.

Figure C2 summarizes results of 500 regressions that have, as outcome, $\text{Reconciliation}_{c,t}$ measured in our 500 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 $\text{Reconciliation}_{c,t}$ provide results fully in line with our baseline one: *The Birth of a Nation* increases references to a reconciliation rhetoric.

Figure C2: Conciliatory and Divisive List- Permutation Exercise



NOTE: The figures above each display 500 coefficients of the variable $\text{BON}_{c,t}$ regressed on 500 alternative measures of $\text{Reconciliation}_{c,t}$ (see the text for further details). On the left panel, we report OLS estimates. On the right panel, 2SLS estimates.

C.3 Alternative Sample and Coverage

C.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 3 in the main text - we flexibly control for newspaper coverage by accounting for the fixed effects of coverage percentiles. Table C8 replicates Table 3 in the main text without accounting for the fixed effects of coverage percentile. In Table C9, instead, we account for decile fixed effects of coverage.

Table C8: 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	0.427 (0.022)		1.145 (0.072)	0.061 (0.074)		0.148 (0.222)	-0.367 (0.076)		-0.997 (0.228)
Million Dollar Mystery		0.436 (0.026)			0.056 (0.085)			-0.380 (0.086)	
1st Stage F-Stat	-	-	223	-	-	223	-	-	223
Dep. Var. Std. Dev.	0.928	0.928	0.928	3.741	3.741	3.741	4.121	4.121	4.121
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (Columns 7 to 9) (see Section C. 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 C. 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 B. for further details). All regressions control for county and month-year. Standard errors clustered at the county level.

Table C9: 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	0.409 (0.022)		1.109 (0.072)	0.291 (0.028)		0.675 (0.081)	-0.118 (0.024)		-0.434 (0.073)
Million Dollar Mystery		0.419 (0.025)			0.255 (0.031)			-0.164 (0.026)	
1st Stage F-Stat	-	-	220	-	-	220	-	-	220
Dep. Var. Std. Dev.	0.928	0.928	0.928	3.741	3.741	3.741	4.121	4.121	4.121
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (Columns 7 to 9) (see Section C. 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 C. 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 B. for further details). All regressions control for county, month-year, and coverage deciles fixed effects. Standard errors clustered at the county level.

C.3.2 Alternative Sample

Our baseline results focus on counties with at least 25% of monthly coverage. Table C10 replicates baseline results imposing no restriction based on coverage.

Table C10: 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	0.402 (0.021)		1.091 (0.070)	0.345 (0.023)		0.847 (0.070)	-0.056 (0.017)		-0.244 (0.047)
Million Dollar Mystery		0.411 (0.025)			0.319 (0.026)			-0.092 (0.017)	
1st Stage F-Stat	-	-	221	-	-	221	-	-	221
Dep. Var. Std. Dev.	0.923	0.923	0.923	3.748	3.748	3.748	4.125	4.125	4.125
Observations	91,124	91,124	91,124	91,124	91,124	91,124	91,124	91,124	91,124

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, $\text{Reconciliation}_{c,t}$ (Columns 1 to 3), the first principal component of patriotic words' log frequencies, $\text{Patriotic}_{c,t}$ (Columns 4 to 6), and the first principal component of divisive words' log frequencies, $\text{Divisive}_{c,t}$ (Columns 7 to 9) (see Section C. 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 C. 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 B. 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.

C.3.3 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 baseline lists. We add to this list, 80 neutral words, selected to be unrelated to patriotic and divisive words. To investigate the heterogeneous effects of BON_{ct} on Patriotic and Divisive keywords, we estimate the following specification:

$$\# \text{ of word}_{ict} \equiv \beta_0 + \beta_1 BON_{ct} * Divisive_i + \beta_2 BON_{ct} * Concil._i + FE_{ct} + FE_{ic} + \epsilon_{ict}$$

where i stands for word, c for county and t for month. $\# \text{ of word}_{ict}$ measures the number of pages containing the word i in county c at month t . 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. 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 occurrence of words is the same for all the words in our lists). The coefficient β_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 β_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 C11 presents OLS and ITT results for the word-based analysis. The exposure of Birth of a Nation increases the relative frequency of Patriotic words, while decreasing the relative frequency of Divisive words.

Table C11: Word-Based Regression - Conciliatory and Divisive versus Neutral Words

	(1) OLS	(2) OLS	(3) OLS	(4) RF	(5) RF	(6) RF
Birth of a Nation * Patriotic Word	4.769 (0.398)				4.769 (0.398)	
Million Dollar Mystery * Patriotic Word		4.498 (0.402)				4.498 (0.402)
Birth of a Nation * Divisive Word			-1.906 (0.611)		-1.906 (0.611)	
Million Dollar Mystery * Divisive Word				-2.233 (0.589)		-2.233 (0.589)
Observations	9,021,825	9,021,825	9,021,825	9,021,825	10808325	10808325

NOTE: The table reports OLS (Columns 1, 2 and 3) and reduced form (Columns 4, 5, 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 C. 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 B. 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 III. 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.

C.3.4 Characterizing Compliers

The 2SLS estimates reported throughout our analysis represent the local average treatment effect (LATE) among *compliers*, namely the sub-population of counties/month for which the screening of BON is associated to the screening of MDM. The impact of BON in this sub-population might differ from the impact of BON on the population of counties at large (ATE). In this section we investigate the characteristics of those compliers.

Following Angrist and Pischke (2009), we first compute the share of compliers among the treated county/months in the following way:

$$\text{Share of Compliers} = \frac{\mathbb{P}[\text{MDM} = 1](\mathbb{E}[\text{BON}|\text{MDM} = 1] - \mathbb{E}[\text{BON}|\text{MDM} = 0])}{\mathbb{P}[\text{BON} = 1]} \quad (3)$$

where $\mathbb{P}[\text{MDM} = 1]$ is the share of counties that screened MDM, $\mathbb{P}[\text{BON} = 1]$ is the share of treated, and $(\mathbb{E}[\text{BON}|\text{MDM} = 1] - \mathbb{E}[\text{BON}|\text{MDM} = 0])$ corresponds to the first stage estimate. We report these statistics in Table C12 for our baseline estimation sample (row 1); for sub-samples that exclude the first year(s) of data (rows 2-5); for the cross-sectional sample (row 6). In the baseline sample, compliers represent 35% of the treated population, which is a sizable share. This share remains large even when excluding the first year(s) of data where both instrument (MDM) and endogenous variable (BON) take the value of 0 by construction. Looking at the cross-sectional dimension only, the share of compliers amounts to 54%. We conclude from this exercise that the LATE refers to a sub-population of fairly large size.

Table C12: Share of compliers in Panel Sub-Samples and Cross-Section

	$\mathbb{P}[\text{MDM} = 1]$	$\mathbb{P}[\text{BON} = 1]$	First Stage	Share of Compliers
Full Sample	.266	.251	.375	.354
1911-1920	.293	.276	.368	.347
1912-1920	.325	.307	.359	.339
1913-1920	.367	.346	.339	.32
1914-1920	.419	.396	.3	.284
Cross Section	.487	.555	.476	.543

NOTE: The table computes the share of compliers for different samples of our analysis (last column). $\mathbb{P}[\text{MDM} = 1]$ is the probability that the variable *MillionDollarMystery* takes on the value of 1, $\mathbb{P}[\text{BON} = 1]$ is the proportion of treated. The third column reports first stage estimates.

As a further step, we examine the distribution of compliers' pre-treatment characteristics for the following observables (all from 1910 Census): Share of illiterate population, share of Blacks, share of population living in cities above 25 thousand inhabitants, and share of foreign-born population. Following Angrist and Pischke (2009), we report in Table C13 the first-stage coefficients' estimates for counties above and below the median value of the characteristic under consideration. We also report in the bottom row the relative likelihood for a complier's characteristic to be above this median value.

All in all, we see that compliers are more likely to be counties with a large share of their population living in cities (> 25000 inhabitants) and to host a higher share of foreign-born population. By contrast, non-compliers are rural areas with more African-Americans and illiterates. Actually, a similar pattern emerges when we look at the distribution strategy of the movie. The distributors of BON have targeted counties following the distribution paths of the previous most successful movies, as a way to maximize revenue. For the pricing strategy, distributors of BON decided to charge unusually high prices for the time, up to 2 dollars per seat (Lenning 2004) ("for at the time

Table C13: Compliers - Demographic Characteristics

VARIABLES	(1) Illiteracy	(2) Black	(3) City	(4) Foreign
Million Dollar Mystery * Low Illiteracy	0.419 (0.027)			
Million Dollar Mystery * High Illiteracy	0.307 (0.035)			
Million Dollar Mystery * Low Share Black		0.414 (0.028)		
Million Dollar Mystery * High Share Black		0.325 (0.034)		
Million Dollar Mystery * Low Share City			0.318 (0.028)	
Million Dollar Mystery * High Share City			0.558 (0.027)	
Million Dollar Mystery * Low Share Foreign				0.237 (0.037)
Million Dollar Mystery * High Share Foreign				0.462 (0.026)
Observations	89,325	89,325	89,325	89,325

NOTE: The table reports first stage regression estimates across county demographic characteristics. The unit of observation is the county month-year. Each regression accounts for county, month-year, and coverage percentile fixed effects. The dependent variable is the binary variable Birth of a Nation, the explanatory variable of interest is Million Dollar Mystery. Low Illiteracy (High Illiteracy) is an indicator variable that takes value 1 in counties that have a share of illiterate population below (above) the median in 1910. Low Share of Blacks (High Share of Blacks) is an indicator variable that takes value 1 in counties that have a share of Black population below (above) the median in 1910. Low Share City is an indicator variable that takes value 1 in counties that have a share of population living in cities of at least 25k inhabitants below (above) the median. Low Share Foreign is an indicator variable that takes value 1 in counties that have a share of foreign population below (above) the median. The last row reports the ratio of the first stage in the sub-population (above the median) to the overall first stage.

full course suppers at leading restaurants were \$1.00 and workers in New York's garment industry were making from \$10 to \$15 a week. Thus a \$2 admission was the equivalent of almost an entire day's pay" (Lenning, 2004, page 119)). Therefore, among the counties that also screened MDM, the counties that could "comply" with this new pricing strategy were more likely to be more urban and educated, with higher potential incomes.

D Reconciliation and Patriotism: Extended Results

D.1 Alternative Sample

Table 4 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 estimates on the same sample of counties of the newspapers analysis (i.e. counties that have newspaper coverage for at least 25% of months in our sample). Results are confirmed.

Table D1: BON and Enlistments - Alternative Newspaper Coverage

	Navy Enlistments					
	Jan 1913 - Aug 1918			Jan 1913 - Mar 1917		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.039 (0.005)		0.081 (0.015)	0.016 (0.005)		0.032 (0.011)
Million Dollar Mystery		0.026 (0.005)			0.009 (0.003)	
1st Stage F-Stat	-	-	209	-	-	197
Dep. Var. Mean	0.04	0.04	0.04	0.01	0.01	0.01
Dep. Var. Std. Dev.	0.19	0.19	0.19	0.10	0.10	0.10
Observations	53,448	53,448	53,448	40,086	40,086	40,086

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 A. 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 C. 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 B. 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.

D.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 D2 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 D2: BON and Enlistments - Only Individuals who Died of Infectious Diseases

	Navy Enlistments					
	Jan 1913 - Aug 1918			Jan 1913 - Mar 1917		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.035 (0.005)		0.064 (0.011)	0.012 (0.003)		0.020 (0.007)
Million Dollar Mystery		0.024 (0.004)			0.006 (0.002)	
1st Stage F-Stat	-	-	329	-	-	281
Dep. Var. Mean	0.03	0.03	0.03	0.01	0.01	0.01
Dep. Var. Std. Dev.	0.16	0.16	0.16	0.07	0.07	0.07
Observations	62,968	62,968	62,968	47,226	47,226	47,226

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 . See Section A. 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 C. 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 B. 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.

E Reconciliation Culture and Names of the Former Enemies: Extended Results

E.1 Robustness Analysis

E.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 E1, we replicate Table 5 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 E1: *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	0.346 (0.153)		0.971 (0.334)	0.009 (0.006)		0.028 (0.012)
Million Dollar Mystery		0.485 (0.170)			0.014 (0.006)	
1st Stage F-Stat	-	-	100	-	-	100
Dep. Var. Mean	46.51	46.51	46.51	0.40	0.40	0.40
Dep. Var. Std. Dev.	13.53	13.53	13.53	0.49	0.49	0.49
Observations	124,440	124,440	124,440	124,440	124,440	124,440

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 A. 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 C. 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 B. 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.

E.1.2 Alternative Sample

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 E2 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 E2: *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	0.446 (0.247)		1.887 (0.657)	0.014 (0.007)		0.059 (0.018)
Million Dollar Mystery		0.747 (0.254)			0.024 (0.007)	
1st Stage F-Stat	-	-	46	-	-	46
Dep. Var. Mean	44.02	44.02	44.02	0.36	0.36	0.36
Dep. Var. Std. Dev.	16.73	16.73	16.73	0.48	0.48	0.48
Observations	81,461	81,461	81,461	81,461	81,461	81,461

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 A. 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 C. 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 B. 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.

E.1.3 Alternative Definitions of BON Exposure

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 E3 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 tends to be larger.

Table E3: *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	0.659 (0.225)		1.838 (0.513)	0.017 (0.006)		0.055 (0.013)
Million Dollar Mystery		0.839 (0.222)			0.025 (0.006)	
1st Stage F-Stat	-	-	94	-	-	94
Dep. Var. Mean	44.16	44.16	44.16	0.36	0.36	0.36
Dep. Var. Std. Dev.	16.81	16.81	16.81	0.48	0.48	0.48
Observations	91,612	91,612	91,612	91,612	91,612	91,612

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 A. 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 C. 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 B. 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.

E.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 E4 replicates baseline results aggregating our measure at the county-year level. In the first three columns, the outcome variable 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 E4: *The Birth of a Nation* and Naming Patterns - County-Year Sample

	Average			Enemy Name Index Mode			Dummy		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS	(7) OLS	(8) RF	(9) 2SLS
Birth of a Nation	0.987 (0.347)		2.235 (0.760)	1.110 (0.389)		2.576 (0.853)	0.019 (0.009)		0.047 (0.021)
Million Dollar Mystery		1.065 (0.363)			1.228 (0.408)			0.023 (0.010)	
1st Stage F-Stat			375			375			375
Dep. Var. Mean	45.16	45.16	45.16	45.22	45.22	45.22	0.39	0.39	0.39
Dep. Var. Std. Dev.	9.88	9.88	9.88	10.79	10.79	10.79	0.27	0.27	0.27
Observations	9,206	9,206	9,206	9,206	9,206	9,206	9,206	9,206	9,206

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 individuals 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 C. 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 B. for further details). All regressions control for county and year of birth fixed effects. Standard errors are clustered at the county level.

E.3 Alternative Definitions of Former Enemies

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 E5 replicates our baseline analysis using this alternative definition of former enemies. Results are, again, unchanged.

Table E5: 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	0.634 (0.269)		1.827 (0.586)	0.020 (0.007)		0.052 (0.015)
Million Dollar Mystery		0.877 (0.281)			0.025 (0.007)	
1st Stage F-Stat	-	-	63	-	-	63
Dep. Var. Mean	43.92	43.92	43.92	0.36	0.36	0.36
Dep. Var. Std. Dev.	17.97	17.97	17.97	0.48	0.48	0.48
Observations	76,445	76,445	76,445	76,445	76,445	76,445

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 A. 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 C. 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 B. 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.

E.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. We 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 E6).

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 E7 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).

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 E8, 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.

Table E6: 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	0.553 (0.306)		1.978 (0.624)	0.020 (0.009)		0.044 (0.018)
Million Dollar Mystery		0.985 (0.319)			0.022 (0.009)	
1st Stage F-Stat	-	-	90	-	-	90
Dep. Var. Mean	44.21	44.21	44.21	0.36	0.36	0.36
Dep. Var. Std. Dev.	16.57	16.57	16.57	0.48	0.48	0.48
Observations	46,765	46,765	46,765	46,765	46,765	46,765

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 A. 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 C. 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 B. for further details). All regressions control for county, gender and year of birth fixed effects. Standard errors are clustered at the county level.

Table E7: 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	0.605 (0.233)		1.780 (0.489)	0.017 (0.006)		0.053 (0.014)
Million Dollar Mystery		0.884 (0.241)			0.027 (0.007)	
1st Stage F-Stat	-	-	87	-	-	87
Dep. Var. Mean	43.92	43.92	43.92	0.36	0.36	0.36
Dep. Var. Std. Dev.	16.72	16.72	16.72	0.48	0.48	0.48
Observations	87,199	87,199	87,199	87,199	87,199	87,199

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 A. 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 C. 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 B. 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.

Table E8: 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	0.615 (0.227)		1.636 (0.468)	0.017 (0.006)		0.046 (0.013)
Million Dollar Mystery		0.817 (0.235)			0.023 (0.006)	
1st Stage F-Stat	-	-	88	-	-	88
Dep. Var. Mean	44.15	44.15	44.15	0.36	0.36	0.36
Dep. Var. Std. Dev.	16.81	16.81	16.81	0.48	0.48	0.48
Observations	89,796	89,796	89,796	89,796	89,796	89,796

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 A. 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 C. 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 B. 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.

E.5 Removing "Birth of a Nation" Characters Names

Tables E9 and E10 display results obtained removing from the analysis individuals with names of BON characters. In Table E9 we exclude individuals with the exact name of characters from the movie.¹⁷ In Table E10 we present the results obtained using a more flexible approach, where we exclude names that are similar to names of BON characters (for instance, we exclude Philip as it contains the name of Phil). Corresponding estimates are similar to the ones provided in the main text.

Table E9: Excluding Individuals with "Birth of a Nation" characters names

	Enemy Name Index					
	Continuous			Dummy		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.628 (0.228)		1.731 (0.476)	0.017 (0.006)		0.049 (0.013)
Million Dollar Mystery		0.864 (0.237)			0.024 (0.006)	
1st Stage F-Stat	-	-	90	-	-	90
Dep. Var. Mean	44.16	44.16	44.16	0.36	0.36	0.36
Dep. Var. Std. Dev.	16.81	16.81	16.81	0.48	0.48	0.48
Observations	91,612	91,612	91,612	91,612	91,612	91,612

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 A. 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 C. 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 B. 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.

¹⁷The following names were considered as "BON-related": Elsie, Flora, Benjamin, Ben, Margaret, Lydia, Austin, Silas, Gus, Jeff, Phil, Tod, Wade, Duke.

Table E10: Excluding Individuals with "Birth of a Nation" characters names - Coarse name identification

	Enemy Name Index					
	Continuous			Dummy		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.656 (0.232)		1.777 (0.486)	0.018 (0.006)		0.050 (0.013)
Million Dollar Mystery		0.886 (0.244)			0.025 (0.007)	
1st Stage F-Stat	-	-	90	-	-	90
Dep. Var. Mean	44.15	44.15	44.15	0.37	0.37	0.37
Dep. Var. Std. Dev.	16.91	16.91	16.91	0.48	0.48	0.48
Observations	89,405	89,405	89,405	89,405	89,405	89,405

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 A. 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 C. 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 B. 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.

E.6 Accounting for Additional FEs

This section presents estimates that control for first (E11) and last name fixed effects (E12) of the respondent. Including first name fixed effects allows to control for all characteristics that are common to families naming their newborn in the same way. Moreover, as long as last-names are somehow informative of family histories, the inclusion of last name fixed effects should partially account for family specific characteristics.

Table E11: Accounting for Last-Name FE of the Respondent

	Enemy Name Index					
	Continuous			Dummy		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.639 (0.277)		1.742 (0.561)	0.018 (0.008)		0.051 (0.015)
Million Dollar Mystery		0.873 (0.287)			0.026 (0.008)	
1st Stage F-Stat	-	-	85	-	-	85
Dep. Var. Mean	44.27	44.27	44.27	0.37	0.37	0.37
Dep. Var. Std. Dev.	16.80	16.80	16.80	0.48	0.48	0.48
Observations	79,999	79,999	79,999	79,999	79,999	79,999

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 A. 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 C. 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 B. for further details). All regressions control for county, gender, (year of birth $\times$ census year) and last-name fixed effects. Standard errors are clustered at the county level.

Table E12: Accounting for First-Name FE of the Respondent

	Enemy Name Index					
	Continuous			Dummy		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.307 (0.188)		1.189 (0.393)	0.011 (0.006)		0.036 (0.012)
Million Dollar Mystery		0.591 (0.193)			0.018 (0.006)	
1st Stage F-Stat	-	-	88	-	-	88
Dep. Var. Mean	44.15	44.15	44.15	0.37	0.37	0.37
Dep. Var. Std. Dev.	16.88	16.88	16.88	0.48	0.48	0.48
Observations	88,582	88,582	88,582	88,582	88,582	88,582

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 A. 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 C. 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 B. for further details). All regressions control for county, gender, (year of birth $\times$ census year) and first-name fixed effects. Standard errors are clustered at the county level.

E.7 Falsification Exercises

E.7.1 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 E13 and E14 display results obtained using as dependent variable the *ENI*, Enemy Name Index, indicator computed using Black individuals.

In Table E13, 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 E14 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 look 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 E13: 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	-0.642 (0.473)		-0.299 (0.978)	-0.028 (0.016)		-0.009 (0.032)
Million Dollar Mystery		-0.151 (0.494)			-0.005 (0.016)	
1st Stage F-Stat	-	-	80	-	-	80
Dep. Var. Mean	49.27	49.27	49.27	0.53	0.53	0.53
Dep. Var. Std. Dev.	15.46	15.46	15.46	0.50	0.50	0.50
Observations	14,031	14,031	14,031	14,031	14,031	14,031

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 A. 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 C. 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 B. 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.

Table E14: Falsification Exercise on Black Newborns - ENI based on white Individuals

	Enemy Name Index					
	Continuous			Dummy		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	RF	2SLS	OLS	RF	2SLS
Birth of a Nation	-0.136 (0.469)		-0.983 (0.969)	0.005 (0.015)		-0.034 (0.030)
Million Dollar Mystery		-0.496 (0.481)			-0.017 (0.015)	
1st Stage F-Stat	-	-	84	-	-	84
Dep. Var. Mean	45.54	45.54	45.54	0.42	0.42	0.42
Dep. Var. Std. Dev.	16.45	16.45	16.45	0.49	0.49	0.49
Observations	16,525	16,525	16,525	16,525	16,525	16,525

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 A. 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 C. 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 B. 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.

E.7.2 Falsification Exercise - Older Cohorts

This section presents evidence providing support to our empirical strategy by means of a falsification exercise constructed by 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 E15, E16, E17, and E18 show that no pre-existing trend in naming patterns 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 E15: 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	-0.421 (0.227)		-0.710 (0.458)	-0.010 (0.006)		-0.017 (0.013)
Million Dollar Mystery		-0.360 (0.231)			-0.008 (0.007)	
1st Stage F-Stat	-	-	99	-	-	99
Dep. Var. Mean	43.66	43.66	43.66	0.35	0.35	0.35
Dep. Var. Std. Dev.	16.47	16.47	16.47	0.48	0.48	0.48
Observations	82,999	82,999	82,999	82,999	82,999	82,999

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 A. 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 C. 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 B. 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.

Table E16: 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	-0.116 (0.213)		-0.953 (0.422)	0.000 (0.007)		-0.012 (0.013)
Million Dollar Mystery		-0.481 (0.210)			-0.006 (0.007)	
1st Stage F-Stat	-	-	95	-	-	95
Dep. Var. Mean	43.32	43.32	43.32	0.34	0.34	0.34
Dep. Var. Std. Dev.	16.07	16.07	16.07	0.47	0.47	0.47
Observations	80,349	80,349	80,349	80,349	80,349	80,349

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 A. 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 C. 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 B. 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.

Table E17: 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	-0.209 (0.138)		-0.491 (0.292)	-0.007 (0.004)		-0.016 (0.009)
Million Dollar Mystery		-0.247 (0.142)			-0.008 (0.004)	
1st Stage F-Stat	-	-	112	-	-	112
Dep. Var. Mean	43.46	43.46	43.46	0.34	0.34	0.34
Dep. Var. Std. Dev.	15.47	15.47	15.47	0.47	0.47	0.47
Observations	227,107	227,107	227,107	227,107	227,107	227,107

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 A. 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 C. 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 B. 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.

Table E18: 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	-0.203 (0.135)		-0.359 (0.279)	-0.008 (0.004)		-0.002 (0.009)
Million Dollar Mystery		-0.182 (0.142)			-0.001 (0.004)	
1st Stage F-Stat	-	-	127	-	-	127
Dep. Var. Mean	44.17	44.17	44.17	0.35	0.35	0.35
Dep. Var. Std. Dev.	14.96	14.96	14.96	0.48	0.48	0.48
Observations	197,044	197,044	197,044	197,044	197,044	197,044

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 A. 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 C. 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 B. 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.

F White Reconciliation and Racial Discrimination: Extended Results

F.1 Alternative Sample

Table 6 in the main text focuses on counties with at least 25% of monthly newspaper coverage. In Table F1, we replicate estimates with no restriction based on coverage and show that baseline results are confirmed. In Table F2, we show that, even when restricting the sample to only county-months with a positive number of hits for the word "Americans" and for the keywords for identifying job ads ("help" and "wanted"), results are retained.

Table F1: BON and Racism towards African Americans: No Restriction on Coverage

	Supremacism			Discrimination		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.036 (0.005)		0.093 (0.014)	0.037 (0.005)		0.093 (0.015)
Million Dollar Mystery		0.035 (0.005)			0.036 (0.006)	
1st Stage F-Stat	-	-	223	-	-	227
Dep. Var. Mean	0.11	0.11	0.11	0.12	0.12	0.12
Dep. Var. Std. Dev.	0.31	0.31	0.31	0.32	0.32	0.32
Observations	91,124	91,124	91,124	91,124	91,124	91,124

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 variable for the presence of white supremacy in the public discourse ($\text{Supremacism}_{c,t}$) in Columns (1) to (3), and an indicator variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$) in Columns (4) to (6). See Section VI. 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 C. 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 B. for further details). All regressions control for county and month-year. Results in Columns (1) to (3) includes percentile fixed effects based on the frequency of pages that contain the word *Americans* in the corresponding county-month. Results in Columns (3) to (6) includes frequency of job-ads percentile fixed effects. Results presented in Panel B also includes State-Year FEs. Standard errors are clustered at the county level.

Table F2: BON and Racism towards African Americans: Alternative Restriction on Coverage

	Supremacism			Discrimination		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.036 (0.005)		0.099 (0.014)	0.037 (0.005)		0.098 (0.017)
Million Dollar Mystery		0.036 (0.005)			0.036 (0.006)	
1st Stage F-Stat	-	-	195	-	-	194
Dep. Var. Mean	0.11	0.11	0.11	0.13	0.13	0.13
Dep. Var. Std. Dev.	0.31	0.31	0.31	0.33	0.33	0.33
Observations	85,964	85,964	85,964	82,264	82,264	82,264

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 first principal component of an indicator variable for the presence of white supremacy in the public and an indicator variable for racial discrimination in the labor market. See Section VI. 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 C. 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 B. for further details). All regressions control for county and month-year. Results in Columns (1) to (3) includes coverage percentile fixed effects. Results in Columns (4) to (6) includes percentile fixed effects based on the frequency of pages that contain the word *Americans* and frequency of job-ads percentile fixed effects observed in the corresponding county-month. Standard errors are clustered at the county level.

F.2 Alternative Measures of Supremacism and Discrimination

This section replicates results on supremacism and racial discrimination in the labor market (Table 6 in the main text) using alternative strategies for the construction of the outcome variables.

Our baseline measure of $\text{Supremacism}_{c,t}$ proxies for white supremacism targeting newspaper pages referring to “white Americans.” As a way to identify other facets of supremacism, we further look at the presence of references to “Aryan” and “True Americans.” On the one hand, the term “Aryan”, taken directly from the script of *The Birth of a Nation* (see Figure 4), is likely to identify more extreme forms of supremacism. On the other hand, “True Americans” can be seen at the same time as a patriotic expression and as a nuanced form of supremacism. Results presented in Table F3 indicate that exposure to the movie affected all these different dimensions of supremacism.

Table F3: Alternative Measures of Supremacism

	Aryan			True Americans		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.009 (0.004)		0.019 (0.012)	0.101 (0.008)		0.256 (0.025)
Million Dollar Mystery		0.007 (0.004)			0.096 (0.009)	
1st Stage F-Stat	-	-	217	-	-	218
Dep. Var. Mean	0.11	0.11	0.11	0.31	0.31	0.31
Dep. Var. Std. Dev.	0.31	0.31	0.31	0.46	0.46	0.46
Observations	89,325	89,325	89,325	89,325	89,325	89,325

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 are two alternative measures on the presence of white supremacism in the public discourse ($\text{Supremacism}_{c,t}$) based on the presence in newspaper pages of the words “Aryan” (in Columns (1) to (3)) and “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 C. 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 B. for further details). Regressions in Columns (1) to (3) control for county and month-year and coverage percentile fixed effects. Results in Columns (4) to (6) includes percentile fixed effects based on the frequency of pages that contain the word *Americans* in the corresponding county-month Standard errors are clustered at the county level.

Turning to racial discrimination in the labor market, in our baseline analysis, we measure racial discrimination by collecting newspaper ads addressed to “White Only” and the keywords “help” + “wanted.” Table F4 displays results obtained using alternative keywords to identify racial discrimination in local newspaper. In Columns (1) to (3), we use as outcome of interest an indicator variable taking a value of 1 if in the corresponding county-month there were pages containing the keywords “White Only” (without imposing any other condition). This approach has the advantage of targeting a very wide range of discriminating ads, but this advantage comes at the cost of a large number of false positives, i.e. all “White Only” retrieved from pages that are not job market ads. As an alternative approach, Columns (4) to (6) and Columns (7) to (9) display results obtained using information on pages containing the keywords “white Men Wanted” and “Wanted White”, respectively. While these keywords minimize the number of false positives, we loose all discriminatory

job ads that do not contain these specific terms.

Supremacism and labor market discrimination are correlated phenomena. To construct a variable combining the variation from both dimensions, we take as dependent variable the 1st component of the principal components of the variables Supremacism_{ct} and $\text{Discrimination}_{ct}$. Results are reported in Table F5.

Table F4: Alternative Measures of Racial Discrimination

	White Only			White Men Wanted			Wanted White		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS	(7) OLS	(8) RF	(9) 2SLS
Birth of a Nation	0.035 (0.006)		0.107 (0.019)	0.008 (0.002)		0.020 (0.005)	0.018 (0.006)		0.045 (0.017)
Million Dollar Mystery		0.040 (0.007)			0.008 (0.002)			0.017 (0.006)	
1st Stage F-Stat	-	-	217	-	-	217	-	-	217
Dep. Var. Mean	0.28	0.28	0.28	0.01	0.01	0.01	0.17	0.17	0.17
Dep. Var. Std. Dev.	0.45	0.45	0.45	0.12	0.12	0.12	0.37	0.37	0.37
Observations	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325	89,325

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 are three alternative measures of racial discrimination in the labor market ($\text{Discrimination}_{ct}$), based on the presence in newspaper pages of the words “White Only” (in Columns (1) to (3)) “White Men Wanted” (in Columns (4) to (6)) and “Wanted White” (in Columns (7) to (9)). 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 C. 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 B. for further details). All regressions control for county and month-year and coverage percentile fixed effects. Standard errors are clustered at the county level.

Table F5: BON and Racism against African Americans: PCA measure of Supremacism and Discrimination

	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.139 (0.016)		0.340 (0.047)	0.160 (0.017)		0.411 (0.053)
Million Dollar Mystery		0.128 (0.018)			0.154 (0.020)	
1st Stage F-Stat	-	-	217	-	-	218
Dep. Var. Std. Dev.	1.16	1.16	1.16	1.16	1.16	1.16
Observations	89,325	89,325	89,325	89,325	89,325	89,325

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 1st component of the principal component analysis of the variables Supremacism_{ct} and $\text{Discrimination}_{ct}$. See Section VI. 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 C. 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 B. for further details). All regressions control for county and month-year. Specifications in Columns (1) to (3) include percentile fixed effects based on the frequency of newspaper pages in the corresponding county-month. Specifications in Columns (3) to (6) include percentiles fixed effects of the frequency of the term "Americans" and the frequency of the keywords identifying job ads ("help" and "wanted"). Standard errors are clustered at the county level.

F.3 Alternative Ways to Control for Job Ads Supply

As discussed in the main text, we are not able to measure the total number of job ads but we can proxy for it counting the number of pages that contain the words “help” + “wanted” . This is consistent with the approach used to define our baseline measure of labor market discrimination. This section replicates our baseline results using two alternative specifications to control for the total number of job ads in the county-month. In our baseline table – Table 6 in the main text – we flexibly control for labor demand shocks and labor market tightness by accounting for the percentiles’ fixed effects associated with the frequency of pages with job ads (where pages with job ads are identified with the keywords “help” and “wanted”, as described above). In Table F6, we present robustness exercises where (instead of percentiles FEs): i) we include the frequency of pages with job ads as a linear control; and i) we include the deciles’ fixed effects associated with the frequency of pages with ads.

Table F6: BON and Racial Discrimination: Alternative Job-Ads Supply Controls

	Supremacism			Discrimination		
	(1) OLS	(2) RF	(3) 2SLS	(4) OLS	(5) RF	(6) 2SLS
Birth of a Nation	0.038 (0.005)		0.095 (0.015)	0.037 (0.005)		0.093 (0.015)
Million Dollar Mystery		0.036 (0.006)			0.036 (0.006)	
1st Stage F-Stat	-	-	223	-	-	223
Dep. Var. Mean	0.12	0.12	0.12	0.12	0.12	0.12
Dep. Var. Std. Dev.	0.32	0.32	0.32	0.32	0.32	0.32
Observations	89,325	89,325	89,325	89,325	89,325	89,325

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 variable for racial discrimination in the labor market ($\text{Discrimination}_{c,t}$). See Section VI. 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 C. 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 B. for further details). All regressions control for county and month-year fixed effects. Results in Columns (1) to (3) includes frequency of job-ads decile fixed effects. Results in Columns (4) to (6) includes the frequency of job-ads pages observed in the corresponding county-month. Standard errors are clustered at the county level.

G Intermarriages

One of the central element of *Birth of a Nation*'s plot is the romance between two Northerner and two Southerner characters. Importantly, the movie ends with the double wedding (i.e. between Margaret Cameron and Phil Stoneman and between Ben Cameron and Elsie Stoneman) that represents the ultimate metaphor of the reconciliation between the two former enemies (from the script "*The aftermath. At the sea's edge, the double honeymoon. Dare we dream of a golden day when the bestial War shall rule no more*").

The frequency of marriages involving North-South partners in each county could also be investigated as a proxy for measuring national reconciliation. In our empirical setting, we would ideally need information on the date of marriage and the spouses' place of origin. In practice, to our knowledge, the most comprehensive source of information on marriages is the US Census, which does not report information on the date of marriages. This means that we can construct a measure of marriages between North-South partners, but only at census-year level (i.e. in 1900, 1910 and 1920). As a consequence, we are not able to explore month-to-month nor year-to-year county variation in North-South marriages. We can only compare changes at the census-year level (i.e. decade level) in the spirit of a long-difference exercise: We compare changes in the frequency of North-South marriages in 1910 and 1920, between counties that screened and counties that did not screen *The Birth of a Nation* (instrumented by *The Million Dollar Mystery*).

More concretely, we compute, for each county in the 1900, 1910 and 1920 censuses, the (log) number of *Former Enemies Households*, namely households where the head of the household was born in former Confederate states and the spouse was born in former Unionist states (and vice-versa).¹⁸ Since censuses do not report information on the year of marriage, the information is aggregated at the (county $\times$ census year) level. Thus, for each county, we have three observations (in 1900, 1910, 1920): At most, only one of those is considered exposed to the screening of the movie (that was released in 1915). We then regress the number of Former Enemies Households on our treatment variable (adjusted for the ten-year time structure of the data):

$$\log(\text{nb Former Enemies HH}_{ct}) = \alpha + \beta \times \text{BON}_{ct} + \gamma_c + \gamma_t + \delta_t \times \text{Number HH}_{ct} + \epsilon_{ct}$$

where c is a county observed in census t and γ_c and γ_t stand for county and census-year fixed effects, respectively. The treatment variable BON_{ct} codes for the post-screening period and is equal to 1 in census-year 1920 if the movie was screened in county c at some point during period 1915-1920 and zero otherwise. Because population size might have changed dramatically over the course of the two decades, we control in a flexible way (i.e. a time varying coefficient δ_t) for the total number of households observed in each county-census-year.

Estimation results are displayed in Table G1. The first three specifications correspond to OLS (Column 1), reduced form (Column 2) and 2SLS (Column 3). The last three columns report the results obtained with a Poisson estimator: Indeed, this non-linear model is particularly well-suited for count data like the one under consideration here; note that its instrumented version is estimated with a control function approach. In all columns, the coefficient of interest is positive and statistically significant at conventional thresholds. Point estimates in columns (1) and (3) of Table G1 imply that airing the movie increased the county-level number of marriages between northerners and southerners partners by 8 and 13 percent respectively. Overall these results are consistent with the evidence presented in the main text and provide an additional and insightful piece of confirmation of our hypothesis that the large-scale diffusion of *the Lost Cause* narrative fostered cultural reconciliation between North and South.

¹⁸We use the data on US-born white individuals from the 1% sample of the Integrated Public Use Microdata Series (IPUMS) of the 1900, 1910 and 1920 censuses.

Table G1: *The Birth of a Nation* and Marriage Patterns

	<i>log (# Former Enemy Households)</i>			<i># Former Enemy Households</i>		
	(1) OLS	(2) RF	(3) 2SLS	(4)	(5) Poisson Model	(6)
Birth of a Nation	0.098 (0.037)		0.158 (0.077)	0.145 (0.066)		0.375 (0.149)
Million Dollar Mystery		0.082 (0.040)			0.198 (0.074)	
1st Stage F-Stat	-	-	305	-	-	-
Dep. Var. Mean	0.69	0.69	0.69	3.00	3.00	3.00
Dep. Var. Std. Dev.	0.84	0.84	0.84	5.47	5.47	5.47
Observations	2,586	2,586	2,586	1,916	1,916	1,916

NOTE: The table reports OLS (Column 1), reduced form (Column 2) and 2SLS (Column 3) estimates. Columns 4, 5 and 6 displays results from Poisson regressions. The instrumented version in Column 6 is estimated with a control function approach. The dependent variable is the log number of Former Enemy Households in columns 1-3 and the number in columns 4-6. The sample includes all white native-born householders recorded in the 1900, 1910 and 1920. 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 C. of the manuscript 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 B. of the manuscript for further details). All regressions control for county, census-year fixed effects and flexible controls for the number of householders observed. Standard errors are clustered at the county level.

References

- Ang, D. (2020). The Birth of a Nation: Media and Racial Hate. *HKS Working Paper No. RWP20-038*.
- Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics*. Princeton university press.
- 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.
- Butters, G. R. (2007). *Banned in Kansas: Motion picture censorship, 1915-1966*. University of Missouri Press.
- Gentzkow, M., J. M. Shapiro, and M. Sinkinson (2011). The effect of newspaper entry and exit on electoral politics. *American Economic Review* 101(7), 2980–3018.
- Lennig, A. (2004). Myth and Fact: The Reception of "The Birth of a Nation". *Film History* 16(2), 117–141.
- Manning, C. and H. Schütze (1999). *Foundations of Statistical Natural Language Processing*. MIT press.
- of Navigation of the Navy Department, N. B. (1920). *Officers and Enlisted Men of the United States Naval Service Who Died During the World War*. US Navy.
- Pischke, S. (2007). Lecture notes on measurement error. *London School of Economics, London*.
- 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.
- Turney, P. D. and P. Pantel (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research* 37, 141–188.