Is or Are: The "United States" in Nineteenth-Century Print Culture

Bryan Santin, Daniel Murphy, Matthew Wilkens

Is or Are: The "United States" in Nineteenth-Century Print Culture

Bryan Santin, Daniel Murphy, and Matthew Wilkens

Among scholars of the early United States, it is a historiographical commonplace that, beginning in the late eighteenth century, the phrase *United States* began a century-long shift from the grammatical plural to the singular. For several of the most recognized of these historians, the American Civil War functions as an explanatory cause for this move from "are" to "is."[1] The connection between language and nationhood has a structural antecedent in scholarship on the founding era. As Trish Loughran argues in *The Republic in Print* (2007), many scholarly accounts of the early republic foreground print culture as a "central and centralizing agent in the processes of American nation formation."[2] The unambiguous role of print culture in the development of a national identity, Loughran implies, is less an interpretation than a logical presupposition underpinning these accounts. This structuring device informs how scholars have read the Civil War as a transformational event in the history of American print culture and, by extension, the reimagination of American nationhood through a shared civic language. But much of the existing scholarship on this topic remains at the level of anecdote because of a deep-seated assumption about the kind of nation that was forged by the large-scale traumatic violence of the war.

In literary studies, there has been a corollary assumption that literature, and American fiction in particular, either caused or indicated such a shift. Both historical and literary scholars have been invested in this narrative of fracture and unification. In scrutinizing this organizing narrative, we find evidence that demands a more dynamic model to understand literature's historical role in cultural semiotics at the systemic level. By using digital techniques, a computational method, and a corpus-level perspective, we examine the role American fiction played in this cultural shift at the level of language. Based on our findings, we can claim with some confidence that American fiction published between 1790 and 1875 referred to the "United States" as an unambiguously plural entity at a high (though declining) rate relative to other print media such as newspapers. To put this claim in perspective, we found that the

rate of plural usage in American fiction published between 1851 and 1875 was actually higher than the rate of plural usage in the Confederate-aligned Richmond *Dispatch* newspaper during the war years. This is true despite the *Dispatch*'s strong political disposition toward the South's preferred conception of the nation as a collection of largely autonomous states. Our results provide a window into the evolution of Americans' cultural self-understanding in the nineteenth century and suggest importantly divergent rates of cultural change across different scales of print production; a single author or newspaper can—and did—shift outlook rapidly, but fiction writing and news reporting as fields of representation evolve much more slowly, even in the wake of so significant an event as the Civil War. This fact in turn suggests that the popular and influential narrative of the war's role in "settling" the singularity of the nation provides a misleading frame through which to grasp the war itself, its effects in the later nineteenth century, and the cultural impact of punctual events in general. Our results support, we believe, ongoing reconsiderations of the war's long cultural reach across the Reconstruction period in particular and the subsequent century and a half of American literary and social development, including work by scholars such as David Blight and Bruce Baker.[3]

It is important to understand how embedded the "is/are" unification story is in existing accounts of the Civil War. Even when historians do not overtly mention the grammatical singularization of the United States, the familiar narrative they tell of sociopolitical crisis and reconstitution is lucidly summarized by gesturing to this recognizable trend among historians of treating the United States as a grammatically singular entity. For instance, in her critically acclaimed *This Republic of Suffering* (2008), a history of the collective trauma of death during the Civil War, Drew Gilpin Faust never makes use of the "United States is" versus "United States are" comparison. But she does claim that the bodies of Civil War dead "became the focus of an imagined national community for the reunited states."[4] Here, Faust could just as easily have echoed James McPherson's assertion in his seminal history of the Civil War era, *Battle Cry of Freedom* (1988), that American society was transformed by the war from a "Union" into a "nation," and hence from a plural "are" to a singular "is."[5] Faust's and McPherson's accounts of the Civil War—and of the kinds of nationalism the event both reflected and brought about—are meaningfully different, but their histories are structurally homologous in their tacit commitment to a causally republican historiography. They both tend to treat language use accordingly as a metric of a newly forged national sensibility. McPherson ends his rich and attentive history with an explanatory anecdote along these lines,

claiming, categorically, that "before 1861 the two words 'United States' were rendered as a plural noun: 'the United States are a republic.' The war marked a transition of the United States to a singular noun."[6]

Two years after the publication of McPherson's Pulitzer Prize–winning book, another prominent American Civil War historian, Shelby Foote, made virtually the same claim in Ken Burns's 1990 PBS documentary, *The Civil War*. Echoing McPherson, Foote stated that the Civil War abruptly changed the "United States" from a plural noun to a singular one, claiming that one way for contemporary Americans to understand "what the war accomplished" was to recognize how radically it transformed American English and national identity: "It made us an 'is.'"[7] More recently, Sean Wilentz revives this claim in his massive, Bancroft Prize–winning *The Rise of American Democracy* (2005). The United States emerged from the war a "transformed nation," Wilentz writes, "no longer referred to as the Union or, as in the common prewar parlance, in the plural—the United States are—but in the singular—the United States is."[8] Across a range of recent and established Civil War histories, and in the wider cultural context those histories have helped shape, the is/are figure persists as an explanatory trope. In the present study, we want to investigate this figure and attempt to understand what it means and has meant for the way Americans think of the relationship between language and nationalization in the nineteenth-century republic.

While this common assertion about the grammatical singularization of the United States in the nineteenth century appears in some of the most celebrated pieces of Civil War scholarship, none of these historians provides much in the way of direct evidence for this particular claim. In McPherson's and Wilentz's books, the claim is not so much a quantitative assertion as an illuminating anecdote in support of larger points about Abraham Lincoln's two inaugural addresses and Gettysburg address, respectively. In fact, the singularization of the United States has a long history of functioning as a dramatic explanatory trope for the sociopolitical impact of the Civil War. As the ex-Confederate soldier and classics scholar Basil Lanneau Gildersleeve wrote in 1909: "It was a point of grammatical concord that was at the bottom of the Civil War—'United States are,' said one, 'United States is,' said another."[9] From our perspective, this way of thinking about language use and the Civil War, or measuring the Civil War's cultural effects via changes in language use over time, has not been a site of corrigible inquiry but is instead something that reflects a long-standing presupposition about the relationship of language and nationhood. For our purposes here, we are less concerned with attempting to amend or revisit specific

accounts of Civil War nationalism and more interested in trying to understand, from a large-data perspective, the nature of the relationship between language use, historical events, and national sensibility. In keeping with recent work on memorial practices and cultural evolution in other areas, we find that these forms changed more slowly and in more complicated ways after the war than the widely deployed story of "is" and "are" suggests.[10]

Even before the "is/are" question served as an anecdote in postbellum America, though, some of the highest elected leaders of the recently defeated Confederacy debated the question with fierce partisan vigor. In his quasi-presidential memoir *The Rise and Fall of the Confederate Government* (1881), ex–Confederate president Jefferson Davis invoked the Founding Era to argue that "those who would understand the true principles of the Constitution can not afford to lose sight of the essential plurality of idea invariably implied in the term 'United States.'"[11] According to Davis, the Constitution's theoretical basis in the concept of federalism proved that the Union "recognizes the distinct integrity of its members [i.e., states], not as fractional parts of one great unit, but as component units of an association."[12] In an even earlier apologia for Southern secession, ex–Confederate vice president Alexander H. Stephens prefigured Davis's claim that the inherent plurality of the United States was rooted in a proper understanding of American federalism. In his two-volume treatise *A Constitutional View of the Late War between the States* (1868), Stephens contended that the Civil War was not really a fight over the question of "African Subordination" but a fight "between the supporters of a strictly Federative Government on the one side, and a thoroughly National one, on the other."[13] What is important to understand, then, is how the grammatical singularization of the "United States" has served as convenient supplemental evidence for a familiar historiographical treatment of the Civil War, even though this evidence has remained largely unexamined.

As this brief history attests, the "is/are" question has significant implications not only for the Civil War era but also for American political history more generally. To date, the only systemic analysis of the issue finds that the shift from plural to singular is more complicated than previously imagined, at least with regard to the US Supreme Court. In his 2008 article "Supreme Court Usage and the Making of an 'Is,'" Minor Myers examined the grammatical singularity or plurality of the "United States" in Supreme Court opinions issued between 1790 and 1919. Categorizing court opinions by decade, Myers found a dramatic increase in singular usage in the 1860s, lending tentative support to the claim that the Civil War functioned as a catalyst for this increase (see fig. 1). Myers's data also show, however, that plural usage nevertheless continued
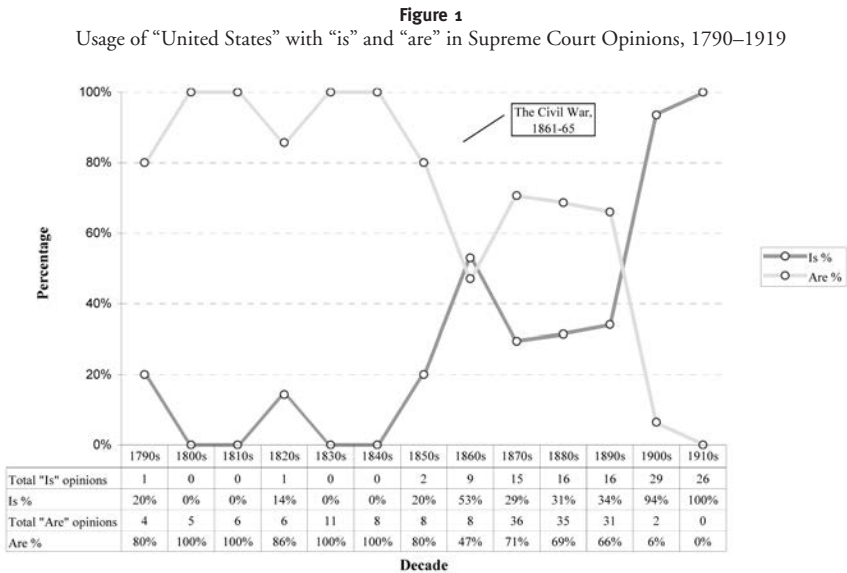
Figure 1

Usage of "United States" with "is" and "are" in Supreme Court Opinions, 1790–1919

	1790s	1800s	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s
Total "Is" opinions	1	0	0	1	0	0	2	9	15	16	16	29	26
Is %	20%	0%	0%	14%	0%	0%	20%	53%	29%	31%	34%	94%	100%
Total "Are" opinions	4	5	6	6	11	8	8	8	36	35	31	2	0
Are %	80%	100%	100%	86%	100%	100%	80%	47%	71%	69%	66%	6%	0%

Decade

Figure 1.

Singular and plural usage of "the United States" in US Supreme Court opinions, 1790–1919. Minor Myers, "Supreme Court Usage and the Making of an 'Is,'" *Greenbag 2D* 11.4 (2008): 460.

to outpace singular usage throughout the 1870s, 1880s, and 1890s. "Only in the beginning of the twentieth century," Myers concluded, "did singular usage achieve preeminence and the plural usage disappear almost entirely."[14] Myers's article is thus a mixed blessing for the familiar singularization thesis posited by historians. On the one hand, his data on plural usage seem to corroborate the notion that the Civil War's profound sociopolitical impact is observable in increased singular usage throughout the 1860s. On the other hand, one hardly has the impression while reading McPherson, Foote, or Wilentz that plural usage would have remained the primary grammatical treatment the United States in the next several decades after the war.

While Myers's work is a valuable examination of this largely understudied question, it represents only the initial foray into this topic. Despite the fact that the grammatical usage of Supreme Court justices affords a unique view of long-term trends in a relatively consistent institutional setting, the Supreme Court is hardly representative of nineteenth-century American society at large. To grasp the larger sociopolitical configuration of the United States, it would be more useful to examine a wider range of nineteenth-century cultural production. Myers's claims encompass almost 130 years of American history, but the basis

for these claims is a collection of fewer than three hundred Supreme Court opinions. From one perspective, Myers's data set is absolutely comprehensive; he examines every US Supreme Court opinion from 1790 to 1919 and draws valuable conclusions from them. But from another perspective, Myers's work begins with a rather small collection of texts from a single, elite governmental institution, from which we must then generalize to a fuller sociopolitical whole.

The methodological constraints of Myers's study throw into relief similar limitations vis-à-vis cultural studies and literary history. In these areas of the humanities, scholars generate large-scale cultural claims by reading closely a small percentage of texts in a given historical period. Close reading, a single critic's scrupulous attention to the meaningful intricacies of texts, yields uniquely productive insights into individual texts and the sociohistorical matrix within which they emerge. But this mode of reading solicits an imaginary relationship between the cultural object and the social field of reference, a relationship that is partly a creation of a media-specific orientation toward knowledge production. Similarly, large-scale data inquiries that usually fall under the name "distant reading" are modes of reading that call into being their own imaginary relationships between texts and contexts, thereby procuring their own media-specific terms for knowing. We wish to stress that all modes of reading, close and distant, manufacture useful, distinct kinds of knowledge. These modes make intelligible different types of information that can be profitably analyzed in light of each other, helping us critically consider and refine our operative categories and corresponding lexicons. The question we ask about the United States as a singular or plural entity pushes at the categorical boundaries of the prevailing "close reading" methodology in literary and cultural studies, and produces new scales and types of information for other scholars. For instance, examining grammatical usage in a handful of nineteenth-century fictional texts may reveal how a small group of canonical authors conceptualizes the unity or disunity of the United States, but such a study may or may not indicate the relationship between large-scale literary production and broader sociopolitical perceptions of the United States as a singular or plural entity.

With that said, we do not mean to imply that distant reading is somehow superior to traditional humanities methodologies or that it is more productive than close reading. Indeed, close reading remains the dominant way to approach literature because of its adaptability and its usefulness in dealing with the multiple semantic capacities inherent in imaginative—and indeed in all—prose. Nor do we want to claim that the digital and computational techniques used in our project are unequivocally better at producing new knowledge for literary

studies. Instead, our position emphasizes that there is, in the words of Mark McGurl, "no one proper scale of literary analysis."[15] Although McGurl does not use computational methods in *The Program Era: Postwar Fiction and the Rise of Creative Writing* (2009), his concern with the "proper scale" of literary analysis leads him to address indirectly similar debates in digital humanities. According to McGurl, "Not only do different perspectives yield different appearances of truth, but different scales of analysis can be differently insightful."[16] For our project, this assertion amounts to the relatively narrow claim that computational techniques are useful to address certain kinds of questions in literary and cultural studies, although we remain cognizant that our results rest on a proxy variable for a very large sociopolitical question.[17] We use these computational techniques to understand better one such original question in American literary and cultural history: the complex correlation between the rise of American national unity and the unity of grammatically singular usage in US fiction during the nineteenth century.

Questions of Scope

Our analysis hopes to provide another way to approach literary history and cultural studies by examining language usage in general, and grammatical usage in particular, at a large scale. This is not to provide a totalizing account of fiction's cultural role in questions of national identity. It is instead our intent to show how specific language use, assessed from a corpus-level perspective, can give scholars new ways to approach the question of literature and print media's place in nineteenth-century cultural production in general, and in the debates about national identity politics in particular. Our study examines the rate at which the "United States" is used as a phrase to refer to an unambiguously plural or singular entity in American fiction between 1790 and 1875. In the latter stages of this analysis, we compare our findings in American fiction to those from a Confederate newspaper, the Richmond *Dispatch*, during the Civil War years, 1861–65; to a larger collection of nineteenth-century newspapers; and to digitized books originally published in the nineteenth and twentieth centuries.

This focus necessarily emphasizes grammatical and syntactical cues rather than a reliance on specific context or detailed analysis of any given author. We are less concerned with an author's political allegiances, philosophy of government, or linguistic-ideological intentions than we are with the ways in which many authors' singular and plural uses of "United States" vary over time. The implied or imagined plurality or singularity of the United States in a given

text is important, to be sure, but this study more immediately concerns how patterns of everyday language use are enmeshed with ways of thinking about national identity. In other words, we are interested in how a complicated series of imaginary relationships are construed in—and called into being by—calculable grammatical constructions as they are employed in language by a large number of writers over more than a full century.

Though our method necessarily does not help us consider closely the historical contingencies or rhetorical nuances that inform an individual author's usage of "United States," it allows us to look closely at a foundational aspect of language and linguistic form and, by extension, its ideological correlatives, in an expansive context. It is no surprise, perhaps, that this project's central question has been of recent interest to linguists as well as historians. After all, grammar imposes order on language and thereby patterns how signs in a language system relate to one another. Richard Slotkin suggests that ideology and, more precisely, cultural memory or mythology are construed in and through a "mythological grammar," a routinized semiotic praxis.[18] At least as early as Fredric Jameson's *Political Unconscious*, literary and cultural scholars have been attentive to the underlying social and economic prerogatives that presuppose and shape language. And after Foucault, and then New Historicism, it can be argued that texts themselves play a substantial role in bringing social and cultural forms into being or in reproducing or reinforcing extant social imaginaries.[19] On some level, our investigation hopes to shed new light on the nature of the linkage between literary texts and cultural and intellectual production at large by examining how the language of American fiction patterns, at the most basic level in grammar, a way of thinking about what the United States is. Or are. Furthermore, this project extends the work of scholars like Minor Meyers while examining the specific cultural work that fiction performs with regard to other sites of cultural transmission in American print culture.

What have scholars made of the cultural work of American fiction with respect to the "is/are" question? The general consensus has been that the rise of the American novel from the mid-1800s forward was concomitant with the extension of a literate reading public, one that was increasingly nationalistic rather than regional in its outlook and sensibility. This has as much to do with the content of literary works as with the material developments and economic practices of nineteenth-century society, which were thought to have slowly consolidated a diverse series of provincial reading networks across the republic and therefore reconstituted what reading entailed as a civic activity. Infrastructural and technological conditions that facilitated mass printing and the dissemina-

tion of print materials from the Jacksonian era onward were thought to have provided concrete linkages and systems of exchange, a cultural architecture conducive to the conception of a national print culture.

But the narrative about fiction's cultural role with respect to the expression of national identity has been fairly consistent. Decades of American literary histories have argued for literature's impact in the United States as a unifying force where it is bound up with the formation of a coherent national consciousness and the articulation of a democratic subjectivity. F. O. Matthiessen and Sacvan Bercovitch in particular popularized for a time an etiologic model of writing literary history;[20] their famous analyses assume a contemporary view of historical developments and posit the origins of said developments in literature. Remarking on this generation of scholars in his 1997 introduction to Leslie Fiedler's *Love and Death in the American Novel* (1966), Charles B. Harris concludes,

> The distinctive feature of these studies is their shared attempt to conceptualize American literature as an integrated whole by subsuming it under a single overriding theme, such as the frontier (Smith), or devotion to the possibilities of democracy (Matthiessen); conflict, such as the collision between innocence and experience (Adams) or industrialism and pastoralism (Marx); or formal characteristic, such as symbolism (Fiedelson) or a foregrounded style (Poirier).[21]

The influence of this body of work persists. More recent scholarship has been guarded about making totalizing claims concerning literature's role in history, but submitting literature's cultural work in a given historical era to a single organizing theme has remained common. A pervasive republican historiography in this kind of literary scholarship posits and in the first instance assumes a certain cause-and-effect calculus. The results in individual pieces of scholarship can be compelling, especially when a critic skillfully demonstrates how a text's complex interworkings correlate with broader social and historical currents. But this model of scholarship often assumes that language (via literature) more or less reliably works *on* and *in* culture in a particular, nationalizing kind of way. It assumes, in other words, the mechanism of cultural production. This tendency has only been bolstered by the theoretical works of thinkers like Benedict Anderson, whose notion of an "imagined community" has provided a useful model for thinking about the rise of print culture and the growing popularity of novels as developments commensurate with centralized nationalism. Slotkin and Tompkins, among others, have authored compelling accounts of literary and cultural history in this vein. Cathy Davidson argues in her history of novels in the early republic, *Revolution and the Word* (1986),

that the novel and the print culture in which it developed were "a major locus of republican education" throughout the mid- to late nineteenth-century.[22] Along these lines, the novel in particular has often been theorized as a salient pedagogical site for its role in cultivating a shared civic sensibility among a diverse range of readers.

Against this critical backdrop, we have sought to interrogate not only the content of this scholarly narrative but the methodological parameters underlying its construction. In recent years, historians have reconsidered the substance of this unification narrative. Trish Loughran's argument in *The Republic in Print* is such an attempt at reconceiving the relationship between print culture and nationalization. Loughran argues against the historical models proposed and employed by Anderson and Davidson, giving an account of US print culture as a fractious and diverse series of networks rather than a singularized body. In her assessment, the realization of something like a coherent, truly "national" print culture threw into relief the real cleavages in—and the divergent character of—the country's language-based identity. Loughran proceeds in her account to describe the predominantly heterogeneous, region-specific print ecology that abounds in the early republic.[23]

We are interested in the ways Loughran contests the straightforward relationship between print culture and national sensibility, and we are compelled by her account of fractious regional print cultures. We find that her account productively critiques Anderson's long-standing theory and provides a plausible alternative account, but her argument tacitly reproduces Anderson's categories. Moreover, the extant methods used by Loughran and others are not sufficient to produce the kind of large-scale evidence necessary to examine systemic dynamics. Stated simply, we are unsure of the cause-and-effect relationship between print and power, text and time. We believe that our method helps us to reconsider these categories.

Method

The computational portion of our project is straightforward. We have used four corpora, two of which are very large (containing tens or hundreds of billions of words) and two of which are smaller (each containing tens or hundreds of millions of words) but of higher quality. The larger corpora include word- and phrase-frequency data (also called *n*-gram data) from the Google Books and Chronicling America projects, each of which is made up of machine-digitized texts not subject to extensive human review.[24] The Google corpus contains

books and periodicals scanned from institutional libraries; Chronicling America encompasses scans of select American newspapers published between 1836 and 1922.

We have also used two focused, nineteenth-century corpora, one literary and one journalistic. The literary corpus is of our own devising and comprises 1,540 volumes of American fiction first published between 1789 and 1875; together, these volumes contain about 117 million words. The corpus is based on the texts cataloged by Lyle Wright in the first two volumes of his *American Fiction: A Contribution toward a Bibliography*. Wright's work attempts to list "the fiction . . . written for adults by Americans and printed in the United States" between 1774 and 1900; he specifically excludes reprints, religious tracts, children's literature, genres other than narrative fiction, serials, and books by non-American writers published in the United States.[25] Wright consulted both physical copies held in libraries and lists of published titles from contemporary sources in the compilation of his bibliography.

Many of the roughly 5,700 titles listed by Wright as having been published between 1774 and 1875 have been digitized, but only a subset of those have been thoroughly hand-corrected and contain firmly established dates of publication. The present work is based on these 1,540 high-quality, XML-encoded, datable volumes.[26] The literary corpus thus includes 27 percent of all known American book-form fiction produced during the first century after independence. Of these, 490 volumes were published between 1789 and 1850 (17 percent of the volumes catalogued by Wright in that period, containing 37 million words); 489 volumes (36 million words) were published between 1851 and 1860; and 561 volumes (44 million words) were published in 1861 or later. The latter two collections together represent 36 percent of the volumes identified by Wright in volume two of his study, covering 1851–75.[27] Women authored 420 of the included volumes (about 30 percent of gender-assignable entries), men 1,005; 115 volumes were written by authors of unknown gender. In cases where authors' ethnicity can be traced, nearly all the included writers are white, though about a third of the books are of unknown ethnic origin. Writers from nonslave states were responsible for about 70 percent of the volumes in the corpus for which it was possible to determine the primary residence or geographic affiliation of the author; about a third of the volumes lacked such data.

Coverage across the period is not perfectly uniform; there are no books included that were published prior to 1789, and the years between 1800 and 1820 are lightly represented in the corpus (reflecting in part the overall pau-

city of domestic literary production during those decades), but other periods consistently contain 30–40 percent of the volumes cataloged by Wright. There is a slight overrepresentation of canonical authors, who were sometimes given priority during digitization, but such writers and their texts are far outweighed by the large bulk of minor and popular volumes.

The corpus, in short, is broadly representative of formally published, book-form nineteenth-century American fiction as it has been preserved in research libraries. This is plainly not the same thing as all nineteenth-century American fiction—in particular, it would be useful to have access to serial fiction and unpublished archives—but it is the best corpus currently available for large-scale literary work in the American nineteenth century, and its lacunae are significantly fewer than those of the conventional literary canon.

Our second primary data set is an archive of the full daily output of the Richmond *Dispatch* newspaper published between November 1, 1860, and December 30, 1865.[28] The *Dispatch* was not published from the evacuation of Richmond (on April 2, 1865) to early December 1865. The corpus includes about 25 million words across 1,384 daily editions of the paper. The content of each edition was converted from the supplied TEI-XML to plain text, excluding in the process embedded file metadata but including paratext (such as selected advertising and subscription information) that appeared in the original source.

After querying these two full-text nineteenth-century corpora for all occurrences of the phrase *United States*, we examined each occurrence in context by hand and tabulated our results. Our designation protocol was conceived in a way that emphasizes grammatical rather contextual meaning. We did not count entries that used the *United States* as a part of a title or used the phrase in an attributive capacity. For instance, such entries as "the constitution of the United States" and "United States Ship" (USS) were discounted. In these cases, the United States is not the proper subject or object of the sentence, and therefore grammatical structure does not unambiguously indicate whether the nation itself is plural or singular. Where *United States* appeared in either the subject or object position in the sentence, but was used in an unclear way with respect to singularity or plurality, it was designated as "ambiguous" and excluded from consideration. For example, "The United States went to war" was marked as ambiguous because, although singularity can perhaps be inferred in this particular case, the verb *went* does not unambiguously indicate plurality or singularity.

These parameters left us with a relatively small percentage of cases in which "United States" was used in a way such that singularity or plurality was expressly

indicated by grammatical usage. The following examples would be marked as singular based on subject–verb agreement:

> "The United States was responsible . . ."
> "The United States is destined . . ."

In contradistinction, phrases of the form "the United States were responsible" or "these United States allied against Britain" would be marked plural on the basis of subject–verb agreement as well as their use of plural pronouns.

This method has the consequence of glossing over cases where *United States* was used in a way that may have suggested plurality or singularity in the context of a sentence despite the grammatical markers employed in a given instance. We accepted this in light of the fact that this problem occurred infrequently and because, when it did occur, it happened in sentences that implied both singularity and plurality (such as "The United States can never be united again").

What We Found

The majority of recent literary history has emphasized fiction's unifying, centralizing cultural role in the context of the nation's development in the nineteenth century. As we stated at the onset, our aim in this project is not to overturn this body of work. We instead want to investigate, at a different scale and with new critical tools, the precise nature of the dynamic relationship between literary production and cultural change as it concerns national sensibility. In many cases, our initial expectations on the is/are question were guided by extant scholarship that we found both compelling and, on some level, intuitive. Fiction's cultural and pedagogical function, and the establishment of a coherent national print culture, was thought to be coextensive with the forces of post–Civil War industrial capitalism, where they together fostered an increasingly unified national identity. In accordance with this line of thinking, we expected to find that the United States would be referred to in the singular more frequently than the plural in American fiction between the late eighteenth century and 1875. Moreover, we expected to find that the rate of singularity would be higher in our corpus of American fiction than in the Richmond *Dispatch* from 1860 to 1865, a newspaper based in the capital of the Confederacy. This latter hypothesis turns on two primary assumptions supported by most literary scholarship. First, that the disposition of fiction tends, for reasons of market function, to be more nationalistic and less regionalized than that of nineteenth-century newspapers. Second, that the *Dispatch*

in particular, a primary news source for the Confederate States of America, would be more inclined than a typical newspaper to conceive the United States as an inherently plural arrangement of states, in keeping with the ideological convictions of the federalist South.

Before we processed our data from the nineteenth-century fiction and *Dispatch* corpora, we performed a provisional analysis of the Google Books American English corpus using three unambiguously singular and plural phrase pairs: "The United States is/are," "The United States has/have," and "The United States was/were." Figure 2 shows the results, which include instances of these phrases in all texts (regardless of genre) available via Google, including at least some non-American texts published in the United States and a great number of nonliterary texts written mostly or entirely in English (nonfiction books, in fact, make up the large majority of Google's holdings).[29]

At first glance, the graph may seem to provide evidence for a decisive linguistic shift in English usage from the plural to the singular around the early 1880s. This macro-view appears to more or less reiterate the consensus that, although a steady shift in usage was occurring before the Civil War, the conflict catalyzed and accelerated this semiotic shift so that the singular eventually overtook the plural for reasons closely connected to the war and its aftermath. However, we need to stress that this graph offers broadly suggestive intimations, but few details. For example, while outside the purview of this article, the shift as it is presented here correlates with the 1877 Hayes–Tilden compromise and the de facto end of Radical Reconstruction. Most importantly, the Google results do not reflect the careful grammatical distinctions that we used in our tabulation of American fiction. That is, it does not track the use of plural pronouns but depends exclusively on subject–verb agreement. It also excludes verb forms other than "to be" and "to have" and misses all instances of phrases beginning with a lowercase "the" (or any other determiner, for that matter), as well as including irregularly capitalized outliers such as "President of *The* United States is." Figure 2 thus represents a synoptic but approximate view of the shift from plural to singular in one aspect of the Anglophone landscape over the last two hundred years.

Turning to the hand-curated literary corpus, we observe a relatively clear trend toward singularization over the period 1789–1875 as summarized in tables 1 and 2 and depicted in figure 3.[30] Note, in tables 1 and 2, the large preponderance of cases labeled "N/A," in which occurrences of "United States" are used in ways that do not indicate the singularity of plurality of the nation itself.
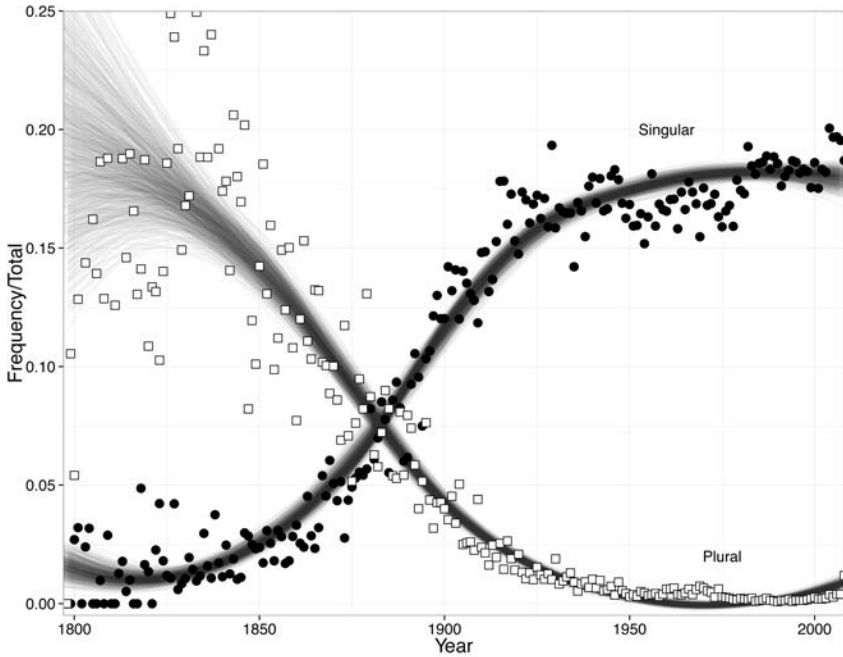
Figure 2.
Frequency of plural (white squares) and singular (black circles) forms of "The United States . . ." as a fraction of all occurrences of "The United States" in the American English corpus of Google Books for volumes published between 1797 and 2008. Fit lines (gray) represent best LOESS fits for 1,000 randomly sampled iterations of each data type.

We explored the possibility of subsetting our literary corpus further by year or decade, by gender, and by geographic origin in order to understand better the effects of those categories on singular and plural usage. We found, however, that the data were too sparse to support significant conclusions about differing rates of usage or change, and so do not present those data here. Gender, geographic, and temporal information is, however, included in the online data that accompany this article. Analogous demographic data are not available for the Google and Chronicling America corpora.

For the *Dispatch*, the transition is more rapid and less ambiguous, though still neither immediate nor entirely uniform. At the turn of 1861, the plural form dominated usage by around three to one; by the end of 1862, the proportions were reversed (see fig. 4).[31] Overall, across the period leading up to and through the war, the singular is the marginally more prevalent form, as shown in table 3.

Table 1.

Singular, plural, and indeterminate uses of "United States" in American fiction published before 1850.

	American Fiction, 1789–1850			
	Singular	**Plural**	**Ambiguous**	**N/A**
Count	13	53	2	1,157
%	1.1%	4.3%	0.2%	94.4%
S/P Ratio	19.7%	80.3%		

Table 2.

Singular, plural, and indeterminate uses of "United States" in American fiction published between 1851 and 1875 (percentage does not sum to 100 because of rounding).

	American Fiction, 1851-1875			
	Singular	**Plural**	**Ambiguous**	**N/A**
Count	58	90	27	2,269
%	2.4%	3.7%	1.1%	93.1%
S/P Ratio	39.2%	60.8%		

Table 3.

Singular, plural, and indeterminate uses of "United States" in the Richmond *Dispatch* published between November 1860 and December 1865 (the total percentage does not sum to 100 because of rounding).

	Richmond Dispatch, 1861-1865			
	Singular	**Plural**	**Ambiguous**	**N/A**
Count	367	310	580	11,179
%	3.0%	2.5%	4.7%	89.9%
S/P Ratio	54.2%	45.8%		

Finally, we calculated the relative usage of the simple forms "The United States is" and "The United States are" across the nine million pages of the Chronicling America newspaper corpus. Unlike the *Dispatch* corpus, the texts held by Chronicling America have not been hand-corrected, and the counts reported here reflect only the presence or absence of each of the target phrases within a page of the original source. Like the Google data, the scope of the collection precludes individual review of the results, but they provide a useful background against which to evaluate the more carefully curated data series.

What It Means

The picture painted by these data is in some ways a murky one. That the grammatical treatment of the United States in fiction, journalism, and library-
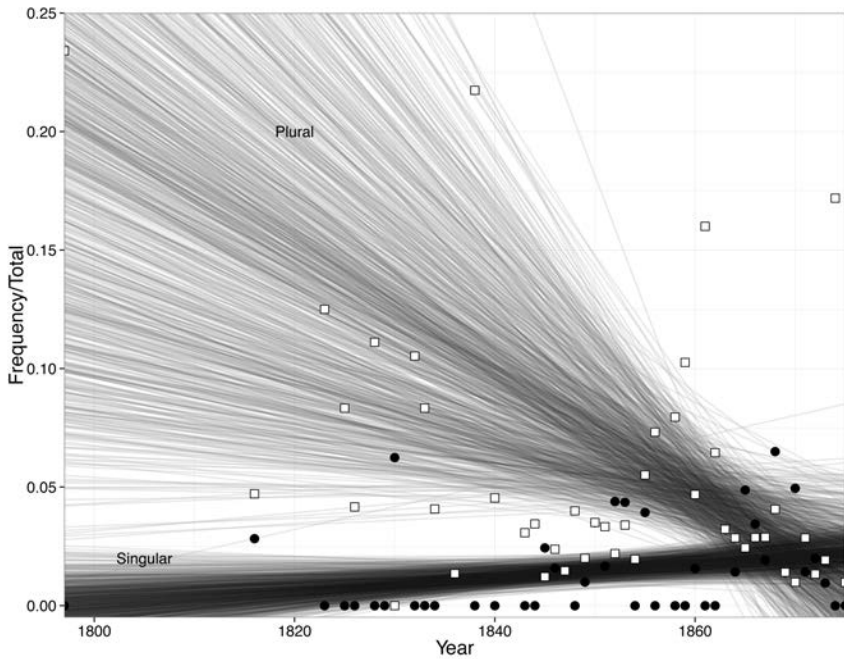
Figure 3.
Frequency of plural (white squares) and singular (black circles) forms of "United States . . ." as a fraction of all occurrences of "United States" in the Wright-based literary corpus for volumes published between 1797 and 1875. Fit lines (gray) represent best linear fits for 1,000 randomly sampled iterations of each data type.

collected books as a whole changed during the nineteenth century is plain: where once the British-inflected plural form—a form enshrined in the US Constitution—dominated American usage, the singular eventually took precedence. Complete standardization on the singular would not occur until the early decades of the twentieth century. But how this happened, how the process differed across media and geographic space, and why it proceeded as it did remain open questions.

Congruent with Minor Meyers's data on Supreme Court cases, our findings indicate that if and when a definitive linguistic shift was achieved in fiction, it occurred decades after the 1865 date at which Civil War historians have typically posited the larger change in American usage. It also appears that fiction used the unambiguous plural at an elevated rate relative to other print media such as newspapers. Indeed, the rate of plurality is higher in fiction written between 1851 and 1875 than in the primary Confederate newspaper published during the Civil War.
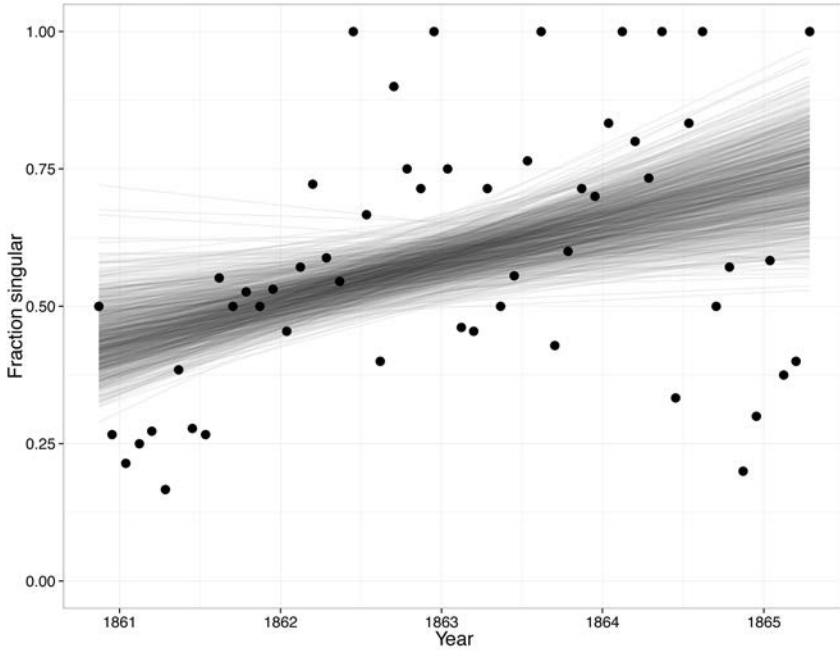
Figure 4.
Fraction of singular uses of "United States" in unambiguous cases in the Richmond *Dispatch*, November 1860 through April 1865. Fit lines (gray) represent best linear fits for 1,000 randomly sampled iterations of the data.

All three of our long-term data sets (Google Books, Chronicling America newspapers, and American fiction) show a transitional period beginning no later than the 1840s and ending no earlier than the beginning of the twentieth century (or, in the case of the literary corpus, not ending at all before the closing of the corpus itself) during which singular and plural usage coexisted at nontrivial levels.[32] In none of these cases is there a notably sharp change during the years of the Civil War.

These facts do not rule out a causal role for the war in driving changes in grammatical treatment of the nation. But it is clear that singularization was already underway before the war began and required perhaps two generations to complete once the war was over. What has long been posited as an event-driven, punctual shift in usage reflecting the outcome of a single political event appears in this light to have been a nearly century-long process without any single point of notably rapid evolution.

It may be the case that this is simply how large-scale cultural change in general took place in the nineteenth-century United States. If so, the emphasis
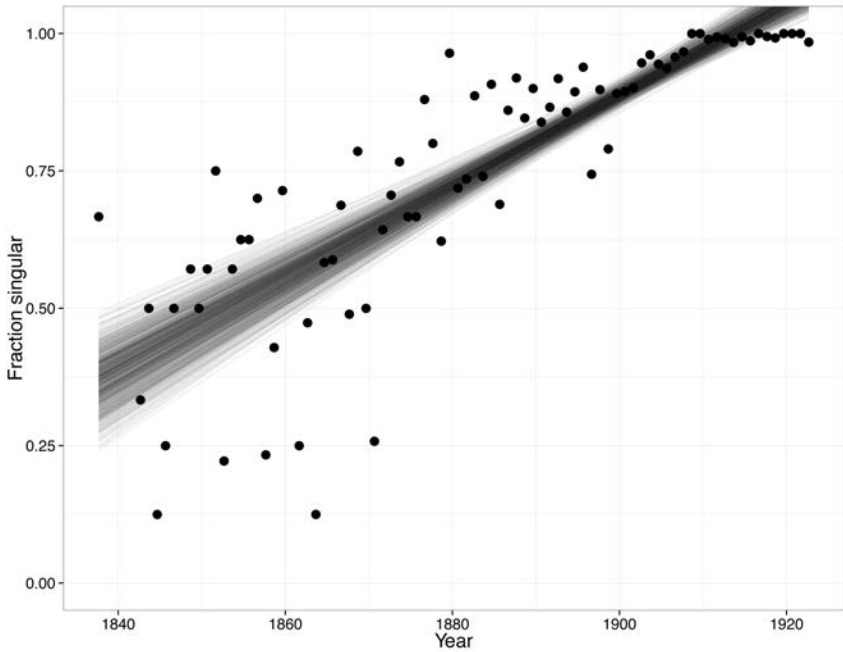
Figure 5.
Fraction of occurrences of "The United States is/are" using the singular form in Chronicling America newspapers, 1836 to 1922. Fit lines (gray) represent best linear fits for 1,000 randomly sampled iterations of the data.

often placed on specific events or turning points is, if not exactly wrong, then perhaps better understood as either a metaphor (it is *as if* the end of the Civil War put to rest the question of the nation's singularity) or an event in the more specialized Badiouian sense of inaugurating a much longer process of achieved transformation.[33] This is true in the rapidly circulating sphere of print journalism as much as in monographs and novelistic fiction, suggesting in turn that what we are seeing has less to do with medium dependence than it does with the mechanics of social transformation.[34] We can only speculate on this point, but the lack of media specificity in the present case in turn suggests the continued relevance of these nineteenth-century findings to contemporary questions. Although communication is now much more rapid, if our social and political concepts evolve not in response to simple mediated exposure to new ideas but as functions of larger social formation, then we ought not to expect fundamental changes to conceptual evolution on the score of increased exposure alone.

The exception to this rule in our data is then less of an outlier than it at first appears: the Richmond *Dispatch* is the one case in which we observe a

comparatively rapid shift to the singular form over a single year (mid-1861 through mid-1862, that is, during the first year of the Civil War).[35] As we noted at the outset, Confederate ideology before and after the war strongly favored plural usage as an index of the federalist understanding of state power. So why did the *Dispatch* move so rapidly in the opposite direction once the war began? Two factors seem to be at play. First, the United States became, after April 1861, the Confederacy's enemy other, no longer a federation of which Virginia and the other Southern states were a part. It thus became much easier—rhetorically valuable, even—to treat the United States as a monolithic entity after that date. This would also account for any reversion to plural usage after the war, when the United States was once again a nation of which the Southern states were part. Second and more prosaically, the *Dispatch* corpus is the only one among our data sources that comes close to being under unified editorial control, the others being collections of very diversely authored, published, and distributed writings. The *Dispatch* is therefore the source that we would most expect to respond rapidly to single editorial decisions rather than evolving as an amalgam of broader social usage. This does not make the *Dispatch* a bad source for our purposes, but it is meaningfully different from the others, despite resembling the literary corpus in size and the Chronicling America corpus in media type.

On the whole, our data complicate two common frameworks in literary studies concerning fiction and the postbellum project of nationalization. The first rests on the strong claim that fiction is a catalyst for nationalist sensibility and a force for unification. The second framework relies on the more modest claim that postbellum nationalization manifests itself in fiction and therefore reads literature as symptomatically expressive of national unification. Stated simply, our findings reveal a small but significant fissure in the assumptions underlying this discourse. In accord with Loughran's argument, we find compelling reasons to rethink the tacit linkages between nineteenth-century literary expression and narratives of national unity. A corpus-level perspective makes clear that what has been called "American fiction" might be reconsidered and extended to include an account of cultural production that is more contingent and politically dissonant than has previously been imagined. Again, multiple *generations* were required for the singular to become an undisputed norm, a result that holds as much for the fast-moving medium of print journalism as it does for book-form writing and the tradition-bound usage of the Supreme Court.

One takeaway is that US fiction does not appear to have played a substantial leading role in propelling the singularization of the United States as a gram-

matical subject, moving, as it does, roughly in step with changes observed in newspapers and in the opinions of the Supreme Court. This fact is significant and surprising for two reasons. It suggests that despite marked differences in the rates of circulation across the media in question—newspapers being published daily or weekly and consumed as disposables; novels gestating over months or years and remaining in circulation for similar spans; court opinions falling somewhere between the two—at least this particular aspect of their content evolved together. Given that the shift to singular usage appears to have been complete only after some three generations, we might posit an evolutionary mechanism in which the changing role of the federal government interacted gradually and imperfectly with what we could call the conservatism of personal formation. If, as we have begun to see elsewhere, writers depict the world less as it is than as it was during their youth, it would matter less, on a question of such basic outlook as the singularity or plurality of the nation, that newspapers respond to current events and novels have a historically broader remit than it would that the authors and consumers of both see the world in terms of their own past.[36] This is less a question of intentionality than it is of institutional and cultural inertia, such that even the sharp shock of the Civil War, colliding with the early stages of a shift in usage that was already underway, was absorbed and dissipated over one or two generational cycles of cultural formation.

Second, the relative synchronicity of these media suggests that fiction lacks any special predictive privilege with respect to future cultural development. The ramifications for humanities scholars interested in understanding the relationship between cultural production and material-political realities are potentially significant; our results weigh against the predictive interpretation, at least in the present case. A more imperative, though less visible, conclusion is that "American fiction" as a category embeds its own historiography. As a heuristic for understanding social reality, "fiction" is already commensurate with a form of nationalist thinking. For someone like Matthiessen, working at the dawn of postwar American studies, to invoke "American fiction" is really to talk about capital-L "Literature" as a category of inquiry that has a set of reading practices and a historiography of nationalization built into it. But even today, after literary scholars have jettisoned the notion that a few canonical texts provide unparalleled insights into social questions, our continued reliance on similar reading practices reiterates certain structural homologies. When our method depends on isolating a select number of salient literary texts to consider the nation, we presume a relationship between subject and object—between text and an imagined field of reference—that reproduces a proxy category very much like "Literature" that makes us susceptible to overestimating the active

influence of literary texts on the nation as a coherent imagined community. In other words, as is always the case, the category produces the landscape to which it refers.

Conversely, a corpus-level data set begins to reveal the epistemological limits of what has hitherto been signified by the term *American fiction*. The sense of national identity that appears in our results is more discordant than extant literary histories and, by extension, definitions of literature have allowed. We suggest that these findings reveal a need for critical discourse to accommodate the emerging dialectic between close and distant reading. Taking up this challenge may lead to productive reexaminations of parts of the discipline, its objects of study, and the ways in which we observe and classify those objects. As we hope is clear in our own modest inquiry, we have tried to modify how we define and know what "literature" is with respect to its cultural work. Like the nation it implicitly signals, aspects of American fiction may need to be reconsidered in the light of large-scale data.

Notes

1. A fuller treatment is coming below, but for now, think in particular of James McPherson, Shelby Foote, and Sean Wilentz.
2. Trish Loughran, *The Republic in Print: Print Culture in the Age of U.S. Nation Building, 1770–1870* (New York: Columbia University Press, 2007), xix.
3. Blight and Baker belong to an important group of cultural historians and sociologists interested in American identity formation in and after the Civil War. Chief among their works are Blight's *Race and Reunion: The Civil War in American Memory* (Cambridge, MA: Harvard University Press, 2001) and a collection of essays, *Beyond the Battlefield: Race, Memory, and the American Civil War* (Amherst: University of Massachusetts Press, 2002), as well as Baker's series of articles on Southern sectionalism and his contribution (Blight also contributes to the collection) in *Where These Memories Grow: History, Memory, and Southern Identity* (Durham, NC: University of North Carolina Press, 2000).
4. Drew Gilpin Faust, *This Republic of Suffering: Death and the American Civil War* (New York: Vintage Civil War Library, 2008), 269.
5. James M. McPherson, *The Battle Cry of Freedom: The Civil War Era* (Oxford: Oxford University Press, 1988), 859.
6. Ibid.
7. Shelby Foote, quoted in *The Civil War*, produced by Ken Burns (PBS Broadcasting, 1990).
8. Sean Wilentz, *The Rise of American Democracy: Jefferson to Lincoln* (New York: Norton, 2005), 790.
9. Basil Lanneau Gildersleeve, *Hellas and Hesperia; or The Vitality of Greek Studies in America* (Baltimore, MD: Lord Baltimore Press, 1909), 16.
10. On memorialization outside print culture, see especially Erika Doss, *Memorial Mania: Public Feeling in America* (Chicago: University of Chicago Press, 2010).

11. Jefferson Davis, *The Rise and Fall of the Confederate Government*, vol. 1 (Gloucester, UK: Dodo, 1881), 109.

12. Ibid. Davis supported his interpretation by quoting several passages of the US Constitution wherein the phrase *United States* is used as an unambiguously plural noun. These passages are found in Article I, Section 9; Article II, Section 1; Article III, Section 2; Article III, Section 3. Davis failed to note that the collective nouns "House of Representatives," "Senate," and "Congress" are referred to as plural in the same document. This more complete reading of the Constitution's grammatical treatment of collective nouns would, of course, undermine Davis's argument, as it could plausibly testify to the influence of standard contemporary British grammatical usage.

13. Quoted in Wilentz, *Rise of American Democracy*, 774.

14. Minor Myers, "Supreme Court Usage and the Making of an 'Is,'" *Greenbag 2D* 11.4 (2008): 458.

15. Mark McGurl, *The Program Era: Postwar Fiction and the Rise of Creative Writing* (Cambridge, MA: Harvard University Press, 2009), 401.

16. Ibid.

17. For one of the authors' views on the compromises involved in computational analysis, see Matthew Wilkens, "Canons, Close Reading, and the Evolution of Method," in *Debates in the Digital Humanities*, ed. Matthew Gold (Minneapolis: University of Minnesota Press, 2012), 249–58. On distant reading more generally, the single best reference remains Franco Moretti, *Maps, Graphs, Trees: Abstract Models for a Literary History* (New York: Verso, 2005).

18. To be clear, Slotkin is referring to grammar not in a standard linguistic sense but in a more figurative sense to signify an assumed syntactical operation in ideology that resembles the governing logic of grammatical usage. See Richard Slotkin, *The Fatal Environment: The Myth of the Frontier in the Age of Industrialization, 1800–1890* (New York: Atheneum, 1985).

19. Although not entirely germane to our immediate set of questions, it is worth noting that more recently neoformalism has extended this line of thinking in literary studies. Vis-à-vis the likes of Jane Tompkins, Caroline Levine argues persuasively that all texts, and the domain accorded to literature in particular, can be a heterotopic cultural field that reproduces, modifies, or resists the dominant formulations it encounters in a given social paradigm. See Levine, "Strategic Formalism," *Victorian Studies* 48.4 (2006); and Levine, *Forms: Whole, Rhythm, Hierarchy, Network* (Princeton, NJ: Princeton University Press, 2014).

20. For the seminal studies, see Sacvan Bercovitch, *The Puritan Origins of the American Self* (New Haven, CT: Yale University Press, 1975); and F. O. Matthiessen, *American Renaissance: Art and Expression in the Age of Emerson and Whitman* (London: Oxford University Press, 1941).

21. Charles B. Harris, introduction to *Love and Death in the American Novel*, by Leslie A. Fiedler (1966; rprt. New York: Stein and Day, 1997), ix.

22. Cathy Davidson, *Revolution and the Word: The Rise of the Novel in America* (New York: Oxford University Press, 1986), 70.

23. As Loughran points out, myriad localized print cultures and attendant political constellations took hold in the period because of uneven levels of infrastructural development. The fact of this regional granularity is our point of emphasis rather than something we will, at present, elucidate and interpret. While interested in the regional aspect Loughran highlights, we are not primarily concerned with explaining the complex historical fact of regional diversity within a bounded nation. We consider instead whether literature, commonly understood as a "nationalizing" force, reflects a corresponding political sensibility as indicated by language usage.

24. We have made particular use of the Google *n*-gram data (books.google.com/ngrams). For details, see Jean-Baptiste Michel and Erez Liberman Aiden, "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331.6014 (2011): 176–82; *For Chronicling America*, which collects a wide range of digitized newspapers published between 1836 and 1922; bookworm.culturomics.org; and the Library of Congress Chronicling America site, chroniclingamerica.loc.gov/newspapers/. We are grateful to Benjamin Schmidt for assistance with the Chronicling America data. Both corpora are very large, totaling millions of volumes (for Google, the 1800–2000 American English corpus used in the present work comprises 252 billion words across 2.3 million volumes) or newsprint pages (for Chronicling America, more than 10 billion words in 9.1 million pages).

25. For a full statement of Lyle Henry Wright's selection principles, see the preface to the third volume of his study *American Fiction, 1774–1900: A Contribution toward a Bibliography*, 2nd ed., 3 vols. (San Marino, CA: Huntington Library, 1966).

26. Digital files were supplied by the University of Virginia Scholar's Lab (for Wright, vol. 1, 1774–1850) and by the Indiana University Digital Library Program (for vol. 2, 1851–75). The support of both institutions is acknowledged with gratitude. A full list of included titles is available at mattwilkens. com/data, as are all other redistributable data used in this study.

27. For additional details on this portion of the literary corpus, see Matthew Wilkens, "The Geographic Imagination of Civil War–Era American Fiction," *American Literary History* 25.4 (2013): 803–40.

28. Digital files were supplied by the University of Richmond, whose support is acknowledged with gratitude. Original digitization was made possible by a National Leadership Grant from the Institute of Museum and Library Services in collaboration with the University of Richmond, Tufts University's Perseus Project, and the Virginia Center for Digital History. For full details of the *Dispatch* corpus, see Robert C. Kenzer et al., *Richmond Daily Dispatch, 1860–1865*, University of Richmond, dlxs. richmond.edu/d/ddr/index.html.

29. The American English data set includes "books predominantly in the English language that were published in the United States," but, unlike Wright's meticulous bibliography of American fiction, does not attempt to exclude "non-American" volumes on any other basis. Querying Google's "English-All" and "English-Fiction" data sets produces broadly similar, if sometimes statistically noisier, results. Because the Google data are case sensitive, preceding "United States" with "The" (rather than "the," or nothing) helps restrict these queries, which cannot be reviewed by hand, to instances in which the United States itself is the subject of a sentence, eliminating occurrences such as "President of *the United States is*."

30. Like the LOESS smoothing applied to the Google data, the fit lines here are best understood as guides to the eye. This is also why we have chosen to visualize uncertainty, somewhat unconventionally, as an overlapping set of fits drawn from repeated random subsamples of the data. In the linear case, there is a minimal embedded model of the underlying historical process, namely, that the shift in usage is linear in time over the period in question. We know that that is not true for the range of dates present in the Google data, making a linear fit inappropriate. But in any case, we use the model for visualization only.

31. Note that the larger number of underlying instances in the *Dispatch* data (see table 3) has allowed us to switch in figure 4 from the raw frequencies shown in figure 3 to the fraction of singular occurrences.

32. In support of the claim that singular usage remained unsettled into the twentieth century, we note—in addition to the sources cited in the introduction—the presence of an editorial, "Are or Is?," by John W. Foster in the May 4, 1901, issue of the *New York Times* arguing in favor of the apparently controversial singular usage. Similar editorial pieces are not difficult to find through the first decade of the century, often arguing against the claim that the Constitution's use of the plural implied that a formal amendment would be required if the singular were to be adopted.

33. On the event as an inauguration of a transformative process, see Alain Badiou, *Being and Event*, trans. Oliver Feltham (New York: Bloomsbury Academic, 2013).

34. On the shared circulation and reuse of nineteenth-century newspapers, see David A. Smith, Ryan Cordell, and Elizabeth Maddock Dillon, "Infectious Texts: Modeling Text Reuse in Nineteenth-Century Newspapers," *Proceedings of the Workshop on Big Humanities* (IEEE Computer Society Press, 2014), 86–94.

35. Anecdotally, we see similar shifts in a handful of other Confederate newspapers, but they have not been digitized to a standard comparable to the *Dispatch*, making a direct, systematic comparison impossible.

36. For more on the temporal dynamics of literary attention, see Matthew Wilkens, "Literary Attention Lag," *Work Product*.