Maryland Population Research Center

WORKING PAPER

“Neglected, Ignored, and Abandoned”? The Working Class in Popular U.S. Culture

PWP-MPRC-2019-009
November 2019

Author:

Reeve Vanneman
University of Maryland

“Neglected, Ignored, and Abandoned”?
The Working Class in Popular U.S. Culture

Reeve Vanneman
University of Maryland
reeve@umd.edu

Abstract

This paper develops new text-mining methods to measure the recognition of American workers in the U.S. press and in American movies. The text-mining program searches 167,193 newspaper articles and 18,056 movie plots for over 35,000 job titles and codes them into standard U.S. Census occupational categories. These occupations are then recoded into common definitions of the working class and tracked over time. For *The New York Times* since 1980, recognition of working-class jobs has not declined, but it was always low. For regional American papers like the *St. Louis Post Gazette*, the *Detroit News*, or the *Tampa Bay Times,* working-class occupations had once enjoyed higher levels of recognition, but the rates have declined recently to levels similar to the *New York Times*. U.S. produced movies show a similar decline since 1930 in working-class inclusion.

KEYWORDS: Working-class, Blue-collar, U.S. Census, Popular culture, Job-titles, Text-mining

“Neglected, Ignored, and Abandoned”? The Working Class in Popular U.S. Culture

> “…the people I have met all across this nation that have been neglected, ignored, and abandoned. I have visited the laid-off factory workers, and the communities crushed by our horrible and unfair trade deals. These are the forgotten men and women of our country.” Donald Trump at the Republican National Convention 2016.

The Trump campaign message resonated with a broad section of American voters. But was there any truth in the claim? Had the American cultural elite neglected, ignored, and abandoned the working class? Working-class incomes had stagnated for decades. Small towns and mid-size cities throughout mid-America were suffering. Was anybody paying much attention?

This paper develops new text-mining methods to evaluate the presence of and changes in the recognition of American workers in the popular press and in American movies. The text-mining program searches 167,193 newspaper articles and 18,056 movie plots for over 35,000 job titles and codes them into standard U.S. Census occupational categories. These occupations are then recoded into common definitions of the working class and tracked over time.

Two aspects of the neglect hypothesis are investigated. First, has the presence of the working class declined in recent years compared to what was standard in past decades? Second, were coastal elites especially neglectful of workers compared to popular culture in mid-America?

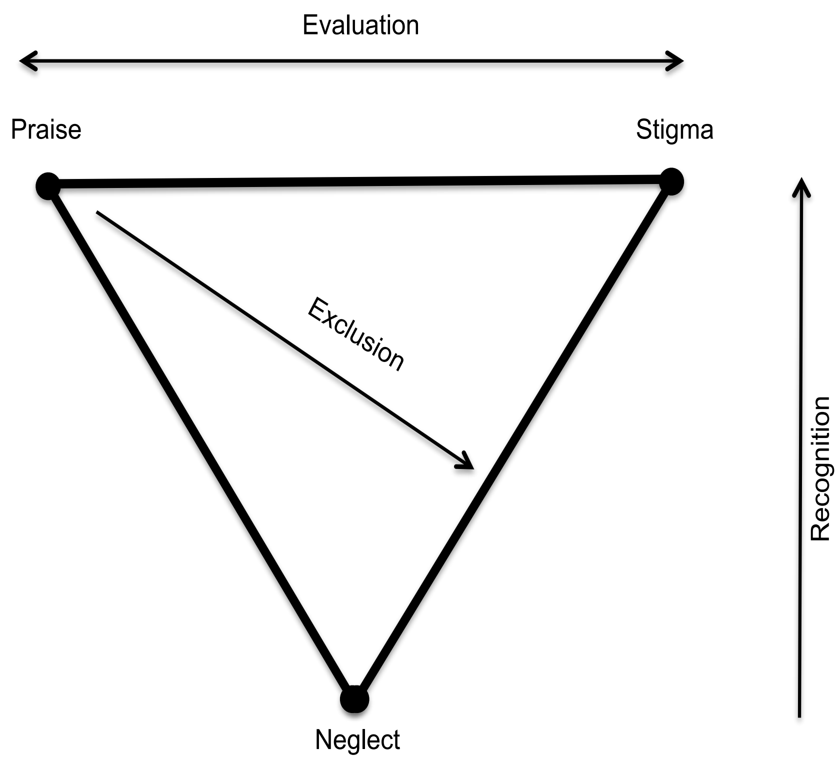
The results show that for a coastal elite newspaper like *The New York Times,* recognition of working-class jobs has not declined, but it was always low. For regional American papers like the *St. Louis Post Gazette* or the *Detroit News*, working-class occupations had enjoyed higher levels of recognition, but the rates have declined recently to levels similar to the *Times*. A second analysis of U.S. produced movies shows a similar decline since 1930 in working-class inclusion. Some measures show a brief resurgence in the 1970s but declining consistently in the decades since.

Conceptual Approach

In her 2017 ASA presidential address, Michéle Lamont paints a two dimensional space of cultural inclusion and exclusion: recognition versus neglect and praise versus stigma (Lamont 2018). Both dimensions are important, but it’s often praise versus stigma that captures our attention while recognition versus neglect is overlooked as a means of social exclusion. Trump appealed to working-class Americans along both dimensions, but it was the “neglected, ignored, and abandoned” line that formed a distinctive appeal. It both resonated with workers’ beliefs in their “rightful place” in American society and their contempt for progressive elites who had ignored their struggles while benefiting from the globalization that had wrecked their working-class communities. Lamont, Park, and Ayala-Hurtado (2017) counted 217 times in 73 election campaign speeches that Trump referred to workers, “which makes this group one of the most frequently mentioned categories.”

The separate dimensions of recognition and valuation are related in interesting ways. Together, they do not create a fully saturated two-dimensional space. Recognition may be either positive or negative; stigmatizing a group is not ignoring them. But neglect, almost by definition, can be neither stigmatizing nor honorific; it is exclusionary but without any specific content that is stigmatizing. So, we need to consider a two dimensional space that is not fully saturated. A good representation would be more triangular (see Figure 1).

Figure 1. Recognition vs. Neglect and Stigma vs. Praise.



A full consideration of cultural inclusion and exclusion must address both dimensions. Trump not only highlighted workers' distress, he also praised the traditional working-class values of hard work (Lamont, Park, and Ayala-Hurtado 2017). But because neglect can be neither explicitly positive nor negative, it requires our first research attention. Once recognition is established, we can ask whether it asserts honor or stigma.

Past Research

Although there is a rich research literature on *how* the working class is portrayed in popular culture (e.g., Ross 1998, Bodnar 2003, Kendall 2005), there is less on how much, where, and when it is represented at all. Recognition or neglect have been more prominent issues for other excluded groups: the sheer extent to which women, ethnic and racial minorities, the poor, or LBGTQ groups are represented in movies, television, or the media is a well-accepted part of research on their cultural exclusion. But the role of neglect is less often incorporated into cultural studies of the working class.

There is an interesting study of working class images in magazine advertisements (Paulson and O'Guinn 2012) – the hardworking Kelly Springfield tire man, the neighborly Culligan water man, and the collaborative team of Amtrak workers. The sample is small, but the frequency of these working-class images has declined since 1970 although they were less common before 1970 also.

There is a substantial, more qualitative. literature on the content of that recognition when it does occur, recognition of other groups in the news and in movies, and a rich ethnographic tradition of recognition of working-class life among individual Americans. Content analyses of class divisions have more often focused on the poor (e.g., Misra, Moller, and Karides 2003, Rose and Baumgartner 2013) or elites (Van de Rigt et al. 2013). The growing development of text-mining methods together with the availability of historical digitized records opens up new opportunities for tracking changes in American culture not only for recognition of the working class but for a wide variety of groups and social relationships.

Methods

Coding occupations from texts

The text-mining program (Vanneman 2019) uses a lexicon that matches texts to a list of over 35,000 natural language job titles each of which is paired to a numeric code based on the 2010 U.S. Census occupation codes. The job titles are one- to three-word phrases commonly found in English texts. The program matches first on the three-word job titles, then the two-word, then single word job titles. This order ensures that “deputy attorney general” is coded as a government official, not a police officer (“deputy”), lawyer (“attorney”), and a military officer (“general”).

The origins of the job title list were two files of occupational coding instructions, a 2016 Census list of over 31,000 job titles (https://www2.census.gov/programs-surveys/demo/guidance/industry-occupation/occupation-index-september-2016.xlsx) and a Bureau of Labor Statistics (BLS) file of 10,000 job titles. Both lists required substantial revision before they could be used with natural language text files. The Census list, while quite comprehensive, often had multiple occupation codes for a single job title depending on the industry location, only one of which could be used in a lexicon. Many job titles, although readily interpreted by human coders, were not phrased as they would appear in natural language (e.g., elevator mechanics are listed as “Mechanic, elevator”). The BLS list, while having a unique occupation code for each job title, excludes job titles that have more than a single Census occupation code.

The Census and BLS lists have been expanded to include military titles (e.g., General, troop) and several illegal activities (e.g., thief, sex worker) not usually included in occupational codes. Several occupational categories have been divided into subcategories because of their frequency in popular culture (e.g., chief executives have been divided to identify the President, governor, and other government chief executives). Several gendered job titles have been divided to maintain that distinction (e.g., waiter is 4110 and waitress is 4111; police officer is 3850 and policeman is 3851). New occupational categories have been created for ambiguous job titles that fit within a broad range of occupational codes (e.g., “senior partner” is coded as “Professional or Manager, not specified” = 3280).

Some of the difficulties in applying a list of job titles to natural language text mining are inherent in the complexities of English. These ambiguities create both errors of omission and errors of commission. For example, many job titles are also proper names (e.g., “Potter”) or names of sports teams (e.g., “Packers”, “Boilermakers”). To exclude these, a second list of over 80 titles were coded as jobs only if not capitalized, eliminating the proper names but thus missing some cases of actual jobs which were capitalized because of sentence or text context (e.g., in newspaper headlines or at the start of a sentence). Other phrases are job titles *only* when capitalized (e.g., “general”, “justice”) and so were coded as occupations only if capitalized.

Some job titles have other common English meanings that cannot be disambiguated by capitalization. “Cast”, for instance, is often the cast of a movie or play (2700, actors) but is also used “to cast a vote”, “to cast doubt” or an “orthopedic cast” for a broken bone. When the alternative meanings could themselves be identified (e.g., cast a vote), the program codes them as non-jobs (9999). Having distinguished these other non-occupational meanings, “cast”, by itself, then most often means the cast of a show and is coded as 2700. Other job titles have meanings that are only occasionally actual job titles in natural language (e.g., “driver”, “guard”, “page”) and so are not given an occupation code, even though they do sometimes identify a job.

While many job titles are associated with multiple occupations, often those occupations are quite similar so coding errors would be minor. However, some job titles can be used for very different occupations. A painter could be an artist or a construction worker. A scout could be a military observer or a talent scout for a sports team. Without more consideration of the context than is now practical, these job titles cannot be disambiguated. The text mining program often uses the code for the job more commonly found in these texts (e.g., painters are 2600: artists; scouts are 9812: military). When the text is more detailed, alternate codes can be assigned (e.g., “house painter” is 6420: painters – construction). Nevertheless, some job titles are so ambiguous that they have been given a separate occupation code, 9997, for example, “crew”, “intern”, or “officer” and are not included in the analyses.

Some job titles refer to quite different occupations, but can be grouped into a broad category encompassing the main meanings. For instance, “director” can be either a director of a movie or play (= 2710, producers and directors) or a director of a corporation or nonprofit (= 430, managers, other). More detailed job titles can be better classified (e.g., “executive director” is most often a manager not a dramatic director). However, when unmodified, “director” is coded as a “professional and manager, not specified” (= 3280). Other job titles that usually, but not always, refer to managerial and professional positions (e.g., “aide”, “associate”, and “staff”) are given the code of 3288, general, likely professional or managerial.

The job title list, the expanded occupational categories, and the python program using the lists are publicly available (https://github.com/ReeveVanneman/occupations). All are continually being revised based on experience, and suggestions or corrections are welcome.

Working class code

A long research literature has debated which occupations and economic roles should be classified as working class. The classification problem is further complicated in natural language texts by ambiguous terms that often, but not always, connote working-class positions. The main definition used in this paper derives from a tradition that emphasizes class as a social relationship between positions with or without power over other workers (Wright and Perrone 1977). Middle-class positions, in this definition, are those who, like capital, control other workers but who, unlike capital, do not own the means of production that employ them. Although initially this definition cross-cut occupational classifications (e.g., carpenters can be workers, middle class, petty bourgeois, or even employers), middle-class positions can be approximated by occupations that the Census defines as managerial and professional (e.g., Vanneman 1977). Most of these occupations have either direct supervisory control over other workers (e.g., managers) or control over the organization of work (e.g., engineers) and its ideological supports (Poulantzas 1974, Ehrenreich and Ehrenreich 1979). Coding professionals and managers as middle class leaves most other occupations as working class. These include not only blue-collar manual occupations (which encompass service work like food, recreational, cleaning, and personal services), but also white-collar occupations such as clerical and retail sales work as well as all technicians.

The overall working-class category also includes some job titles that could not be coded into a single occupational category but encompass a broad range of work and occupations that are clearly working class (e.g., "laborer" or "unskilled worker" and are coded into a broad occupation category, 9760= laborer, nec). Such broad working-class occupational categories are created for unskilled labor, skilled manual labor, and unspecified clerical and retail sales work. They do not fit a standard Census code but are given their own code that is included with other working-class occupations into the broad working-class category.

More problematic are job titles that could refer to almost any occupational category (e.g., “employee”, “crew member”, or an unmodified “worker”). In context, these ambiguous titles most often implicitly refer to working-class rather than professional-managerial positions. They too are given an “occupation” code separate from the Census codes, and can be combined with other working-class occupations in the broadest working-class classification.

For robustness, this research has used three separate definitions of working-class occupations ranging from the narrowest encompassing only manual blue-collar occupations, an intermediate definition that also includes white collar occupations in clerical and most sales jobs as well as most technicians, and the broadest which adds the ambiguous titles that usually but not always signal working-class positions (e.g., “employee”). The frequencies of an article with any working class title vary as expected across the three definitions. For instance, 30 percent of articles in the *Times* include a mention of at least one working-class job title according to the broad definition; 24 percent according to the intermediate definition; and only 11 percent when restricted to blue-collar occupations. However, the regression results are generally similar across all three definitions, so the reported results focus on the intermediate definition which includes white collar job titles but excludes the somewhat ambiguous titles such as “employee” or “worker”. The occasional differences in results across the three definitions are noted when appropriate.

Other occupational classes

Four other categories of occupations are kept separate from a working-class middle-class dichotomy because of their prominence in institutional culture like the news and Hollywood movies and their alternative positions in a simple middle-class vs. working-class clasification. First, an “upper-class” category is reserved for capitalists (e.g., “financiers”, “industrialists”, “tycoons’), inherited aristocratic positions (“queen”, “duke”, “sultan”), and inherited wealth (e.g., “heiress”, “landed gentry”).

Second, both the news and movies include a disproportionate inclusion of police so they are separated into their own classification (including “cop”, “detective”, “FBI agent”, “police chief” and other related titles). Similarly, the military, both officers and enlisted, are a separate category excluded from the middle class – working class division. Finally, although usually a smaller category than any of the above, farmers are separated into their own category. Together, these other categories account for 31% of all occupational mentions in the *Times* and occur at least once in 55% of the articles.

In addition to these occupation-based class codes, the program also separately counts direct mentions of class and income-based categories: “blue collar”, “working class”, “middle class”, “rich”, “poor”, and similar references. While it might be expected that these class categories would be associated with the occupation-based categories, the correlations are actually quite low (between .04 to .09) and generally lower than the correlations across class within either the occupation-based or class-based codes. The class-based codes are also less common. Especially for the newspaper articles, it would appear that the two types of codes reflect different styles of recognizing class: one more specific and concrete, the other more general and conceptual. The regression results are also quite different.

Measures of working-class recognition

The basic measure of working-class recognition is the simple dichotomy of whether the text (a newspaper article or a movie plot) includes any mention of a working class job title (as defined in the three alternatives described above). A second, more intensive, measure requires at least three mentions of a working-class job title. Finally, a third measure calculates the proportion of all occupational job titles coded in the text that are working class. The three measures generally show similar comparisons across newspapers and trends over time. The two alternative measures are reported only when substantive differences are found with the simple presence vs. absence measure.

Text Sources

Newspaper articles were downloaded from Lexis/Nexis fully digitized records. Lexis/Nexis has a limited sample of digitized newspapers and limited but varying years for each source. Table 1 lists the years and number of articles for six newspapers that have sufficient annual records to test the declining recognition hypothesis. Seven days were sampled in each year (e.g., the second Monday in January, the last Thursday in October). All articles from the sample day in each year were downloaded into a text corpus for each newspaper. Very short articles (under 25 words) or very long articles (over 2500 words) are dropped from the analyses, generally between 2 and 5 percent of all articles in each paper.

Table 1: Newspaper samples.

	Initial year	# news-papers	total # articles	sample # articles*	# words	# job titles
New York Times	1980	275	57,785	56,705	37,990,500	712,107
St. Louis Post-Dispatch	1989	214	29,201	28,002	15,119,121	296,779
Detroit News	1999	123	6,403	6,142	3,502,709	60.499
Pittsburgh Post-Gazette	1993	185	27,123	26,269	14,701,801	245,926
St. Paul Pioneer Press	1995	165	13,978	13,529	7,318,845	123,809
Tampa Bay Times	1987	229	32,703	31,209	17,736,345	330,284

*Very long articles (> 2500 words) and very short articles (< 25 words) were dropped from the analysis of each newspaper.

As in many analyses of the press, *The New York Ti*mes is the principal representative of the Eastern elite news media. Digitized texts are available from 1980 until the present. Unfortunately, the major California newspapers are not available in Lexis/Nexis for comparison of elite news sources from both coasts. In mid-America, the *St. Louis Post-Dispatch* has the longest series of digitized articles dating from 1989. Comparisons are also made to three other mid-American newspapers, the *Detroit News,* the *Pittsburgh Post-Gazette,* and the *St. Paul Pioneer Press*. Finally, Lexis/Nexis has a reasonably long series from the *Tampa Bay Times* (formerly the *St. Petersburg Times*) for analysis of a non-elite but Southern news source.

Movie plots were downloaded from Wikipedia entries for U.S. produced (or co-produced) movies from 1930 to the present. While other internet summaries are available, no other source of plot summaries is as complete or as well standardized. Wikipedia editing guidelines suggest 500-700 word plot summaries, and non-conforming summaries are regularly noted and revised. Wikipedia also maintains lists of all U.S. produced movies for each year totaling 23,721 movies from 1930 through 2018. Of these, 98% have their own web pages on Wikipedia, and of these 79% have a plot summary on the web page. Coverage is better for more recent decades. Prewar decades may be especially incomplete, but results do not suggest any discontinuities so the full annual results are reported here.

Analyses

Trends were analyzed with logistic regressions of working-class presence on the year of newspaper publication or the year of the movie release. To compare working-class recognition in the *Times* with regional newspapers, a simple dummy variable for the regional newspaper and its interaction term with year are included. Year is centered at 2000 in these models so that the coefficient for the regional newspaper estimates the difference in the two newspapers for that year. For these comparisons, the longer *Times* data were limited to the years when digitized records were available from the regional newspaper. More detailed analyses are presented graphing estimated probabilities for each year from a regression that included dummy variables for each year.

All regressions include controls for the size of the text, measured as a cubic function of the number of words in the text. Those three size coefficients are almost always statistically significant: the longer the text the more likely a working-class job title is included but to a declining degree. For newspapers, a control is also included for whether the article appeared in the sports section. Athletes, referees, and coaches are, by Census definition, professionals, so sports articles include on average more middle-class and fewer working-class job titles. Sports articles are also a larger proportion of articles in regional papers so this control corrects for that difference. Dummy variables are also included for the day of the week, but these differences are far smaller than the controls for the length of the article or the section of the paper. All predicted probabilities reported below include controls for article length (estimated at 600 words for newspaper articles or 400 words for movie plots), sports section, and the day of the week. Full results are available in the online tables.

Results

New York Times

The fifty most common job titles in the *Times* are listed in Table 2. Except for the general and somewhat ambiguous titles of "employee" (#21) and "worker" (#25), none are from the working class. The highest ranking job title coded as working class is "chef" (#137), and many of those titles are references to head chefs running a restaurant who are more properly middle class. Next most common is #195 "secretary", but a significant minority of those are misclassified (lower case) references to Cabinet and corporate secretaries (e.g., "Secretary of State" is #110). Even further down the list are "salesman" (#223), "maid" (#246), and "mechanic" (#253).

Table 2. Fifty most common job titles in *The New York Times*, 1980-2018.

rank	jobtitle §	Census occupation	# articles	# mentions	% plural
1	official	Managers, other 430	9629	20680	69%
2	president	Chief executives 10	8904	16803	3%
3	lawyer	Lawyers 2100	5091	11514	43%
4	spokesman	Public relations specialists 2825	4902	6568	4%
5	player	Athletes, coaches, umpires, nec 2720	4855	13418	66%
6	director of	Professional & managerial, ns 3280	4806	5934	0%
7	Dr.	Professional & managerial, ns 3280	4772	18823	0%
8	President	President of the country 12	4653	9919	3%
9	executive	Managers, other 430	4571	8037	52%
10	police	Police & sheriff's patrol 3850	4038	10419	0%
11	owner	General, likely prof/mgr 3288	4023	7101	46%
12	chairman of	Chief executives 10	3966	5167	0%
13	director	Professional & managerial, ns 3280	3889	6160	24%
14	writer	Authors 2850	3664	5892	33%
15	management	Managers, other 430	3607	5571	1%

16	author	Authors 2850	3575	5882	23%
17	expert	Advisers & experts 2005	3549	5220	71%
18	judge	Judges 2110	3438	10357	16%
19	governor	Chief executive, government (exc.Pres) 15	3356	9079	7%
20	reporter	Journalists 2810	3330	4992	62%
21	employee	Employee nec 9770	3323	6709	77%
22	artist	Artists 2600	3261	8393	52%
23	critic	Advisers & experts 2005	3254	4472	74%
24	staff	General, likely prof/mgr 3288	3252	4423	3%
25	worker	Employee nec 9770	3222	6859	87%
26	senator	Legislators 30	3213	8901	18%
27	manager	Managers, other 430	3203	5347	28%
28	coach	Athletes, coaches, umpires, nec 2720	3058	7150	17%
29	chief	Managers, other 430	2998	4136	8%
30	analyst	Social & political analysts 1863	2989	5810	67%
31	head of	Managers, other 430	2988	3540	0%
32	chief executive	Chief executives 10	2914	4484	4%
33	mayor	Chief executive, government (exc.Pres) 15	2852	7511	4%
34	candidate	Politicians, candidates 32	2756	6157	50%
35	chairman	Chief executives 10	2576	3742	3%
36	doctor	Physicians 3060	2574	5853	60%
37	investor	Investors 125	2446	6073	81%
38	cast	Actors 2700	2415	3221	7%
39	founder	Owner manager, 25	2381	2967	16%
40	king	Government official, inherited 33	2354	4208	13%
41	politician	Politicians, candidates 32	2185	3098	73%
42	producer	Producers & directors 2710	2132	3447	44%
43	actor	Actors 2700	2130	4279	50%
44	teacher	Other teachers 2340	2075	4707	59%
45	prime minister	Chief executive, government (exc.Pres) 15	1980	3623	2%
46	adviser	Advisers & experts 2005	1940	2907	40%
47	executive director	Managers, other 430	1877	2153	0%
48	editor	Editors 2830	1874	2858	23%
49	prosecutor	Lawyers 2100	1849	4388	65%
50	Representative	Legislators 30	1841	3268	6%

§ Capitalized job titles are counted only when capitalized in the text.

On average since 1980, 24.1% of *Times* articles have mentioned a working-class job title (29.4% if we include the more general, employee, code; but only 10.8% articles include a blue-collar job title). These working-class job titles account for only 5.1% of all job titles included in the *Times*. These frequencies are, as expected, far lower than for middle-class occupations (90.0% of articles) or even upper-class occupations (40.1%).

The annual trend of working-class job titles in the *Times*, controlling for article length, is negligible (β= +0.001, which could easily have occurred by chance). More intensive measures of working-class recognition do sometimes show negative trends over time, but they are still quite weak. For instance, articles with three or more mentions of a working-class job title show a weak decline over time (β= −0.003, $p<.10$). Only the mention of specifically manual, blue-collar occupations has declined noticeably over time (β= −0.005). This translates to a predicted 1.6 percentage point drop (from 9.6% to 8.0%) between 1980 and 2019 for a standard length article of 600 words.

Similarly, the recognition of the working class as a specific class category is low and unchanging. This code is a composite of several phrases but mostly "working class" (61%) and "blue collar" (31%). Only 1.4% of *Times* articles use one of these working-class labels. This is less than the use of labels for the "middle class" (1.8% of articles) or for the "upper class" (including "rich", 7.1%) or for the "poor" (9.8%). Like working-class occupations, the use of a working class label has been quite steady over time (β= +0.004 which is not statistically significant).

In short, there is only weak evidence of any major decline in the *Times* recognition of working-class occupations or a working-class label itself with a possible exception of the subcategory of manual, blue-collar occupations. The *Times'* recognition of the working class is below that of middle-class or even upper-class occupations, but that has always been true. There is little to suggest any recent changes.

St. Louis Post-Dispatch comparison

The *St. Louis Post-Dispatch* has the longest digitized record of any mid-America paper (from 1989) and provides an appropriate comparison for coverage of the working class with that of the *Times.* The *Post-Dispatch* was the paper of the legendary Joseph Pulitzer and remained in the Pulitzer family until 2005. It is now the only paper publishing in the St. Louis metropolitan area.

The *Post-Dispatch* has had a generally higher inclusion of working-class job titles than the *Times* (29.8% of articles annually vs. 23.8% for *Times* articles of similar lengths in the same years), but that inclusion has declined so the year interaction comparing trends with the *Times* (β= -0.014) is statistically significant. Other measures of inclusion of working-class job titles (e.g., limiting the measure to manual, blue-collar job titles, counting more intensive article with three or more mentions of working-class occupations) show a similar pattern of higher recognition in the *Post-Dispatch* but declining over time.

Table 3. Logistic regressions of *St. Louis Post Dispatch - New York Times* differences in working-class inclusion using three definitions of the working class and three measure of inclusion.

	paper		year (-2000)		paper X year	
Any mention in the article:						
1. all working-class titles	**0.2845**	***	-0.0011		**-0.0140**	***
	(0.0199)		(0.0013)		(0.0022)	
2. occupations only (i.e., no employee")	**0.2574**	***	-0.0006		**-0.0140**	***
	(0.0212)		(0.0013)		(0.0023)	
3. blue collar titles only	**0.2756**		**-0.0049**	**	**-0.0133**	***
	(0.0281)		(0.0018)		(0.0031)	

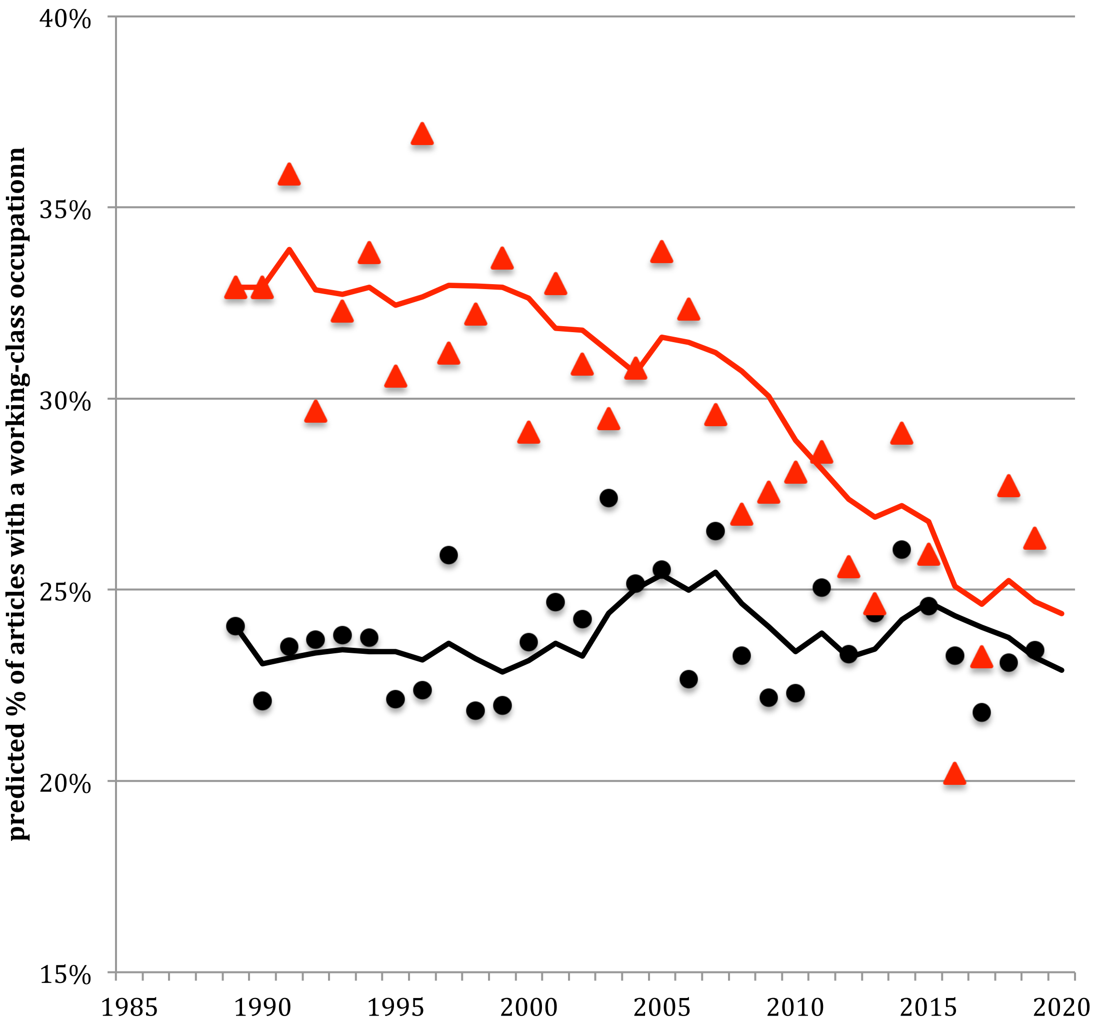
Occupations only (i.e., no "employee")			
4. At least three mentions in the article:	**0.4546** ***	**-0.0064** **	**-0.0113** **
	(0.0396)	(0.0024)	(0.0043)
5. Proportion working class titles	**0.3135** ***	-0.0008	**-0.0090** ***
	(0.0200)	(0.0013)	(0.0022)

Sample sizes: *New York Times*= 43,174; *St .Louis Post-Dispatch*= 28,002. Each logistic regression includes controls for the day of the week, a cubic function of article length, and a dummy variable for sports section. Year is centered at 2000, so, because of the interaction term, the coefficient for "paper" (the *St. Louis Post-Dispatch*) represents the estimated difference from *The New York Times* at that year.

Most other measures of working-class recognition show this same pattern of higher but decreasing levels in the *Post-Dispatch*. The exception to this pattern is a named "working class" in an article. While neither paper uses "working class" or its close affiliates like "blue collar" very often, in the *Post-Dispatch* it is even less common (0.4% of articles) than in the *Times* (0.8%). Like working-class job titles, a named working class is found slightly less often in recent years in the *Post-Dispatch* ($\beta = -0.020$, $p=.05$), but not in the *Times* ($\beta = +0.005$, n.s.). So, although for neither newspaper is there much use of working class labels, it appears that the *Times* is more willing to discuss the working class in the abstract – as a general category – but less often actual specific working-class jobs that working-class readers might recognize. And while the *Post-Dispatch* began the period with more recognition of working-class jobs, in recent years that recognition has declined to the same low levels of the *Times*.

Figure 2 plots the decline in working-class job titles more precisely by estimating each year separately. The decline is fairly steady after the turn of the century, and by the end of the second decade of this century, the predicted probabilities are close to the low level of the *Times*.

Figure 2. Predicted annual probabilities of working-class job titles from *The New York Times* and the *St. Louis Post-Dispatch*, 1988-2019.



Note: Predicted probabilities are computed after controls for article length, day of the week, and sports section. Solid lines represent moving averages from the previous 5 years.

Comparisons with other regional newspapers

Two of the other three mid-America newspapers show much the same pattern as the *Post-Dispatch* although the data cover fewer years (Table 4). The digitized record for the *Detroit News* begins only in 1999, but the regression results are quite similar to the *St.Louis Post-Dispatch.* In 1999, the predicted probability of including a working-class job title was 31.8%, compared to 25.2% for the *Times*. It soon declined, however, and for the latest year the *Detroit News* predicted probability (24.0%) was barely above the *Times* (23.3%). The rate of decline ($\beta = -0.020$) is even steeper than for the *Post-Dispatch*, perhaps because of the later initial year.

Table 4. Logistic regressions of differences between five regional newspapers and *The New York Times* in levels and trends of inclusion of a working-class job title .

	# articles	paper		year (-2000)		paper X year	
1. St. Louis Post-Dispatch	71,176	**0.2574**	***	-0.0006		**-0.0140**	***
		(0.0212)		(0.0013)		(0.0023)	
2. Detroit News	29,042	**0.3803**	***	**-0.0052**	*	**-0.0143**	*
		(0.0626)		(0.0026)		(0.0066)	
3. Pittsburgh Post-Gazette	64,141	**0.2148**	***	-0.0017		**-0.0092**	***
		(0.0254)		(0.0016)		(0.0028)	
4. St. Paul Pioneer Press	48,775	**0.0850**	*	-0.0011		0.0002	
		(0.0362)		(0.0018)		(0.0042)	
5. Tampa Bay Times	77,274	**0.3876**	***	0.0006		**-0.0063**	*
		(0.0191)		(0.0012)		(0.0020)	

Note: See Table 1 for years and sample sizes of individual newspapers. All five logistic regressions include controls for the day of the week, a cubic function of article length, and a dummy variable for sports section. Year is centered at 2000 in each of the repressions, so, because of the interaction term, the coefficient for the regional newspaper represents the estimated difference from *The New York Times* at that year.

The *Pittsburgh Post-Gazette* also shows a significantly higher proportion of articles with a working-class job title in 2000 than does the *Times* , and the decline since 1993 ($\beta = -0.011$) is steeper than for the *Times*. The *St.Paul Pioneer Press* is a partial exception to these patterns. In 2000, the estimated proportion of articles with a working-class job title

(31.9%) is only slightly above (β= +0.085, $p<.05$) that for the *Times* (31.8%) and neither the *Pioneer Press* nor the *Times* shows any consistent decline from 1995 to 2019.

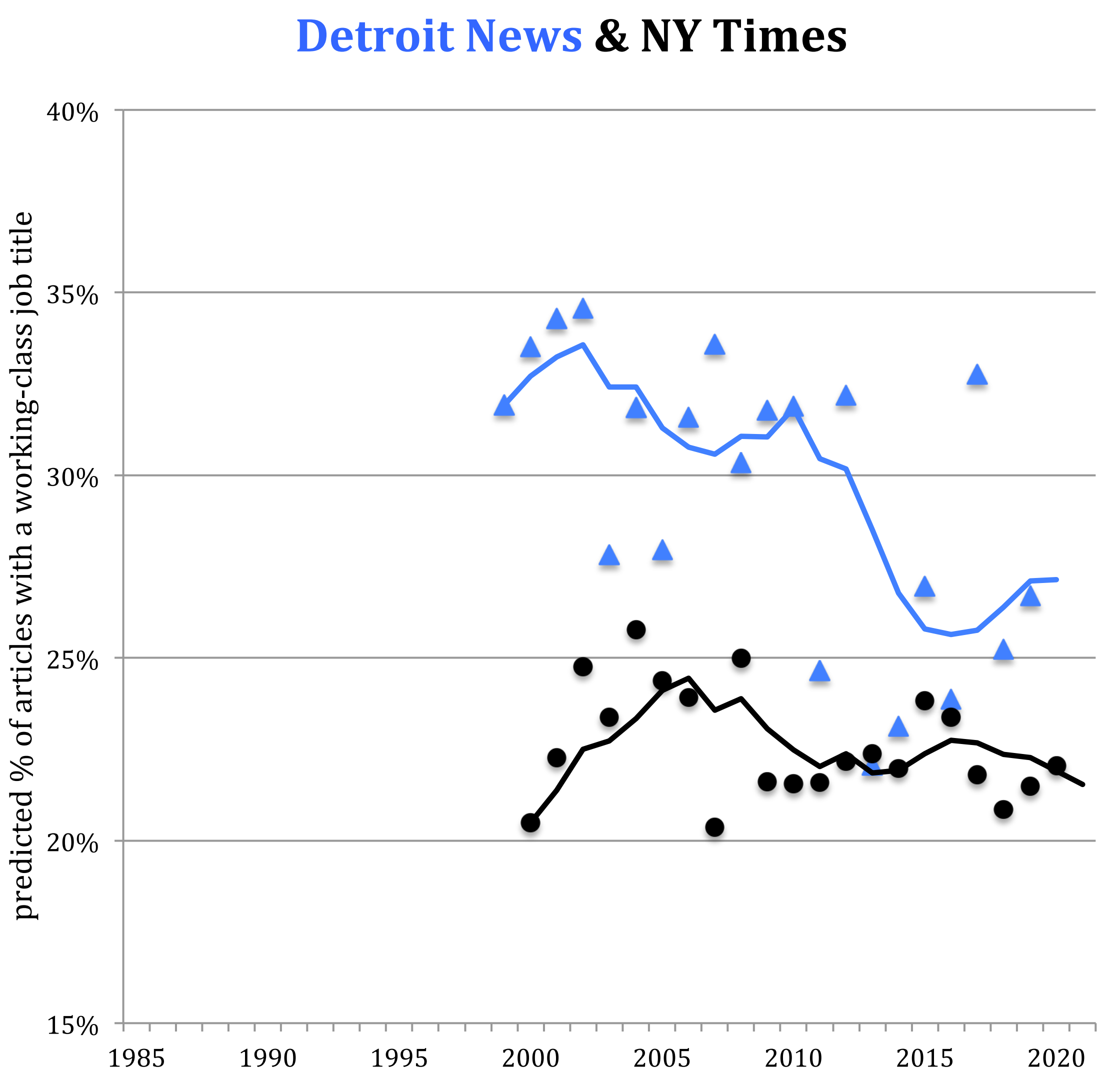
The *Tampa Bay Times* shows a similar pattern to the *St.Louis Post-Dispatch* and the other mid-America papers. The digitized texts date from 1987 (when it was the *St. Petersburg Times*), and for the rest of that century the proportion of articles with a working-class job title was consistently higher than for *The New York Times*. In 2000, an estimated 31.9% of *Tampa Bay Times* articles included a working-class job title, significantly more (β= +0.388) than the estimate for *The New York Times* (24.1%). But working-class recognition declined over the period (β= −0.011).

The regression results for the regional newspapers confirm the "neglected" hypotheses: the elite New York Times has paid less attention to working-class positions than have the regional papers, but the regional papers have recently declined to levels not far above the *Times*.

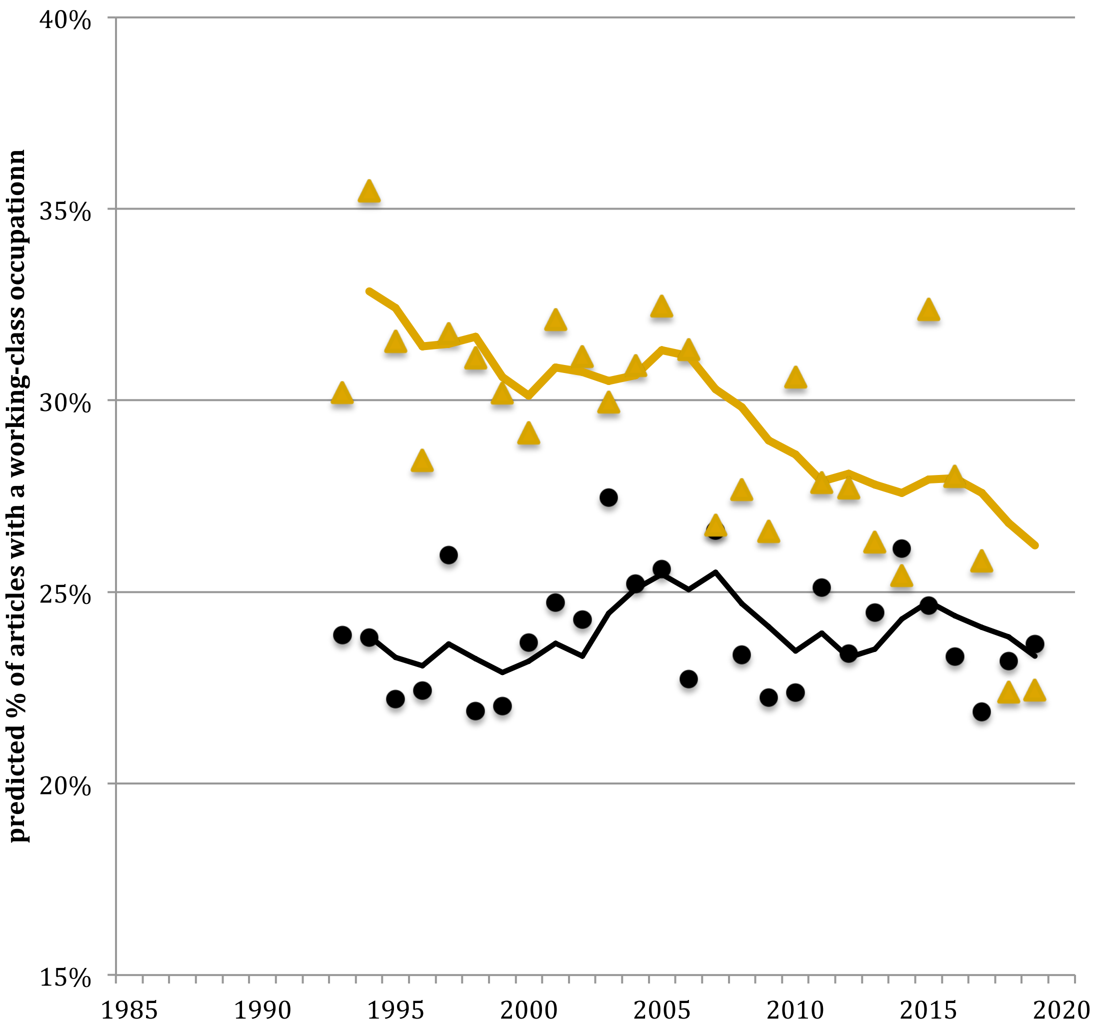
The linear year coefficients in the regressions of Tables 3 and 4 actually understate the declining recognition of the working class. Figure 2 showed that the decline for the *St.Louis Post-Gazette* was steeper in more recent years. That acceleration of the decline is even more apparent for the other regional papers. Figure 3 plots the predicted annual percentages of articles with a working-class job title for each of the other four regional papers. For each paper, the decline accelerates or is only observed in the 21st century. The decline for the *Tampa Bay Times* is especially steep. Even the *St.Paul Pioneer Press*, for which the linear regression line was not statistically significant, shows a decline in working-class recognition after the first decade of this century.

Figure 3. Predicted annual probabilities of working-class job titles from *The New York Times* and four regional newspapers.

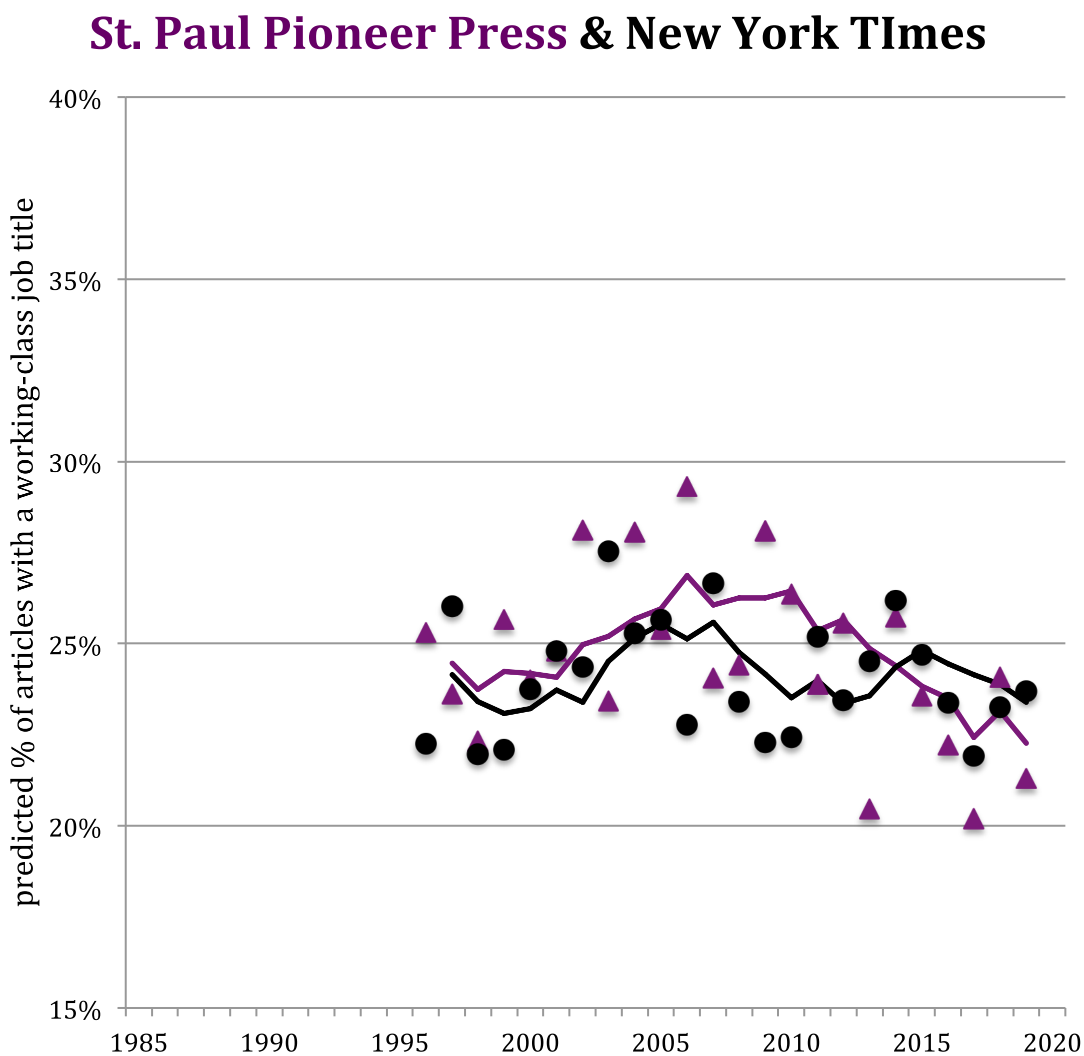
The Detroit News:



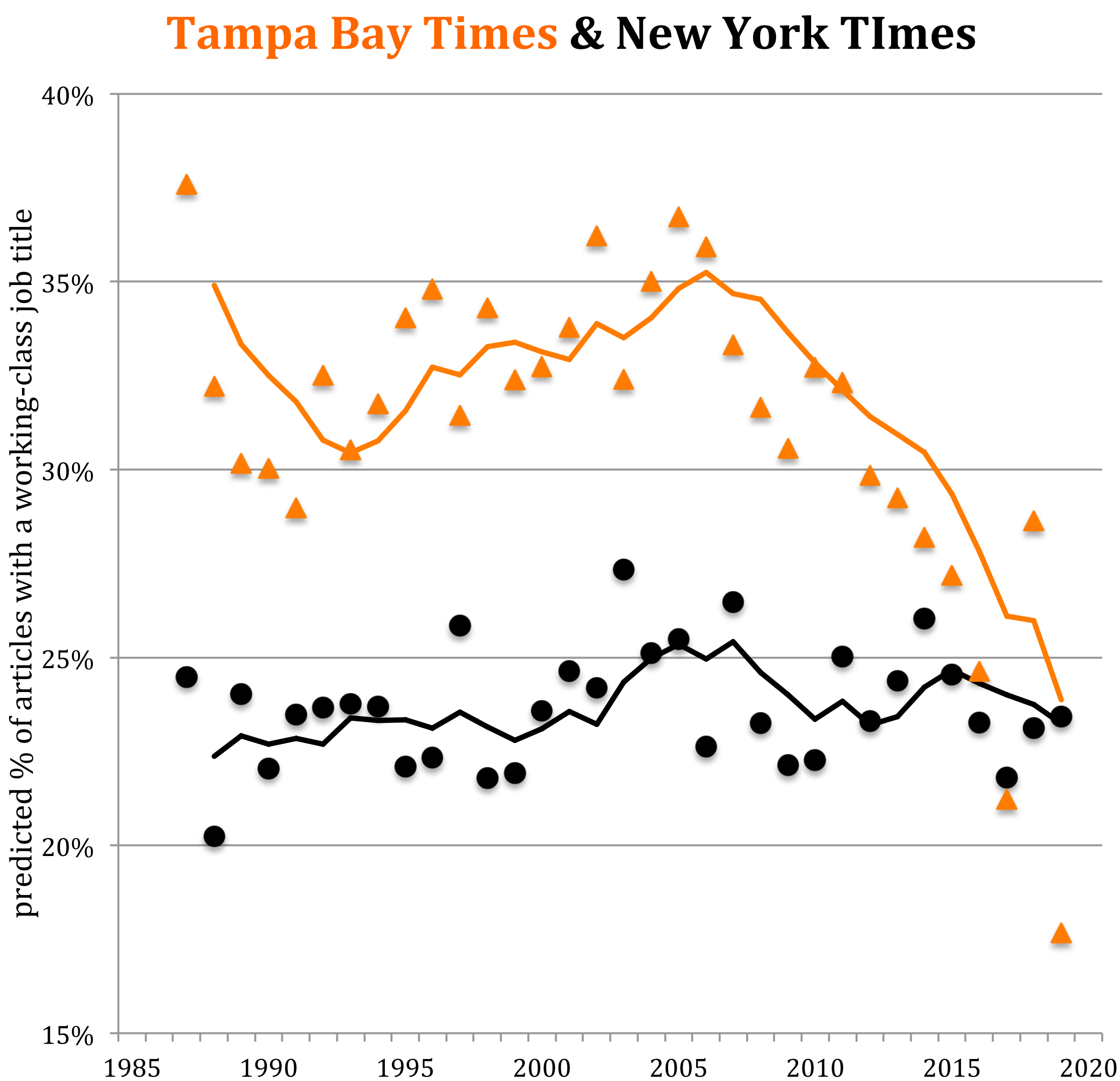
The Pittsburgh Post-Gazette:



The St. Paul Pioneer Express:



The Tampa Bay Times:



Note: Predicted probabilities are computed after controls for article length, day of the week, and sports section. Solid lines represent moving averages from the previous 5 years.

Movies

Wikipedia plot summaries are available for 18,056 U.S. produced movies from 1930 to 2018. The 50 most common job titles (Table 5) show many similarities and some interesting differences with the list from the *Times* in Table 2. Not surprisingly, the police are more prominent, including detective (#12), sheriff (#16), cop (#22), and police officer (#24). The military is also more prominent, as are criminals and performers.

Table 5. Fifty most common job titles in U.S. produced movies, 1930-1918.

rank	job title §	Census occupation	# movies	# mentions	% plurals
1	police	Police & sheriff's patrol, 3850	3143	5841	0%
2	Dr.	Professional, managerial, ns, 3280	1722	4203	0%
3	doctor	Physicians, 3060	1417	2218	12%
4	owner	General, likely prof/mgr, 3288	1287	1589	13%
5	gang	Criminal /ns, 9850	1200	2503	3%
6	captain	Military officer, 9800	1167	2286	1%
7	boss	Managers, other, 430	1148	1623	4%
8	soldier	Military, rank ns, 9812	1095	2112	66%
9	officer	Ambiguous: which occ., 9997	1041	1706	39%
10	criminal	Criminal /ns, 9850	931	1196	35%
11	killer	Murderer, 9856	911	1638	10%
12	detective	Detectives, 3820	883	1374	23%
13	crew	Ambiguous: which occ, 9997	878	2072	4%
14	king	Government official, inherited, 33	783	2195	2%
15	lawyer	Lawyers, 2100	783	973	7%
16	sheriff	Supervisor: police, 3710	742	1538	1%
17	assistant	General, likely prof/mgr, 3288	728	851	6%
18	agent	Detectives, 3820	719	1345	39%
19	singer	Musicians, singers, 2750	693	869	5%
20	prisoner	Prisoner (or ex-Prisoner), 9858	663	1038	49%
21	reporter	Journalists, 2810	639	859	24%
22	cop	Police & sheriff's patrol, 3850	630	954	42%
23	pilot	Pilots & flight engineers, 9030	616	1068	27%
24	police officer	Police & sheriff's patrol, 3850	609	720	34%
25	judge	Judges, 2110	607	1167	8%
26	spy	Intelligence officers, 3825	582	873	31%
27	gangster	Mafia, organized crime, 9857	578	911	37%
28	henchman	Criminal /ns, 9850	577	852	70%
29	head of	Managers, other, 430	568	610	0%
30	nurse	Registered nurses, 3255	566	766	18%
31	sergeant	Enlisted military supervisors, 9810	561	891	2%
32	manager	Managers, other, 430	561	715	4%
33	scientist	Physical scientists, other, 1760	543	888	42%
34	colonel	Military officer, 9800	529	985	1%
35	employee	Employee nec, 9770	520	645	44%
36	thief	Thief, 9854	511	767	45%
37	lieutenant	Military officer, 9800	504	700	4%
38	teacher	Other teachers, 2340	494	683	17%
39	chief	Managers, other, 430	491	744	2%
40	professor	Postsecondary teachers, 2200	487	943	4%

41	murderer	Murderer, 9856	479	561	12%
42	secretary	Secretaries, 5700	468	535	2%
43	actor	Actors, 2700	467	665	32%
44	prostitute	Sex worker, 9855	462	611	25%
45	General	Military officer, 9800	459	799	1%
46	policeman	Police & sheriff's patrol, 3850	452	568	33%
47	commander	Military officer, 9800	442	662	4%
48	actress	Actresses, 2701	442	563	5%
49	band	Musicians, singers, 2750	438	1050	5%
50	staff	General, likely prof/mgr, 3288	428	579	0%

§ Capitalized job titles are counted only when capitalized in the text.

For our interest in working-class neglect, the general pattern is quite similar. Again, the most common references are to the general, broad mentions of “employee” (#35) and “worker” (#61). The movie list does include “secretary” (#42), but all the remaining job titles are middle and upper class or police and criminals. Further down on the list (but higher than for the *Times*) are “servant” (#59), “maid” (#67), “waitress” (#87), “bodyguard” (#95), and “sailor” (#98),

Less than half (40.2%) of Wikipedia movie plot summaries include even a single working-class job title, well below the percentage with a middle-class job title (81.0%). Movies with a manual, blue-collar job titles are even less common (34.0%). The low recognition of working class jobs in movies has declined since 1930 ($\beta = -0.011$), a predicted fall from 54.2% of 1930 movies to 31.1% in 2018. Other definitions of working-class occupations and other measures of incidence show similar negative trends (Table 6).

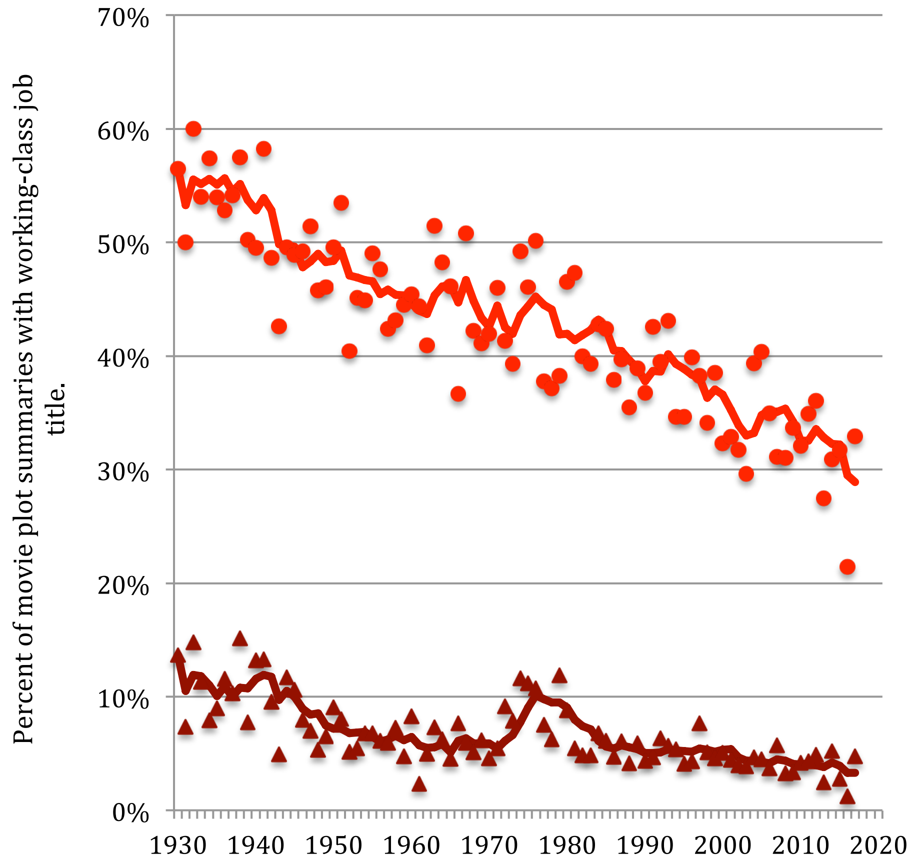
Table 6. Annual trends of working-class job titles in U.S. movie plot summaries, 1930-2018.

	Measure of inclusion		
	at least one mention	three + mentions	percent of all job tittles
working-class + employee job titles	-0.0092	-0.0101	-0.0086
	(0.0007)	(0.0010)	(0.0007)
working-class job titles	-0.0110	-0.0125	-0.0105
	(0.0007)	(0.0011)	(0.0007)
blue-collar job titles	-0.0092	-0.0118	-0.0087
	(0.0007)	(0.0013)	(0.0007)

N = 18,056
Note: logistic regressions control for a cubic function of plot summary length. Standard errors are reported in parentheses below the linear year coefficients.

The 1930-2018 decline appears quite steady throughout this period for at least one mention of a job title for the main working-class definition (see the top line in Figure 4). Most other definitions of working-class occupations and other measures of recognition show similar steady declines throughout this period. An interesting exception to the steady declines is the decline for the more intensive working-class recognition (three mentions or more in the plot summary). While also a generally negative trend ($\beta = -0.013$), there is a noticeable resurgence at the end of the 1970s when movies such as *Taxi Driver* (1976), *Deer Hunter* (1978), *Norma Rae* (1979), and *9 to 5* (1980) appeared (see the lower line in Figure 4).

Figure 4. Predicted annual values of working-class job titles in U.S. movies, 1930-2018.



Note: predicted values of standard length plot summary of 400 words. Solid lines are five-year moving averages. Top line is for one or more job titles, bottom for three or more.

Discussion

The results for both newspapers and movies show a surprisingly clear confirmation of the declining recognition of the working class in American popular culture. While three of the four regional newspapers examined had once included substantially more working-class content than did the elite press (as represented by the *Times*), that recognition is no longer true for any of them. The decline in working-class recognition since the turn of the century is striking in all four regional newspapers. And U.S. produced movies have shown a steady decline of inclusion of working-class characters since 1930. The U.S. working class may have always had reason to feel ignored by the elite *New York Times*, but by the time of the 2016 election, that feeling of abandonment could well have extended to their local papers as well. Hollywood, too, has steadily reduced their inclusion of the working class, with a possible exception of a short period in the late 1970s.

How working-class Americans perceived this decline or whether this neglect played a role in the Trump election is beyond the scope of this paper. The first step is to recognize that cultural change has occurred and that in the news media at least it has historically varied across the country. Cultural exclusion is, by itself, a cause for concern because it contributes to inequality (Lamont 2018). Next, we might investigate possible consequences; for example, were the regional declines in working-class recognition correlated with the Trump vote in unexpected places?

If investigating consequences is a next step for these content analyses, we do not have to look far for candidates that are possible causes for the cultural neglect. The decline of working-class positions in the labor force is well documented. Using Current Population Survey data from IPUMS with harmonized 2010 Census occupation codes (Flood et al. 2018) and with the same working-class definition used in the text mining analyses above (but adjusting for the changing Census codes for occupations), the negative slope of working-class occupations across time is as strong ($\beta = -0.017$) as for most of the cultural indicators shown above. Working-class recognition may have declined in American popular culture because the proportion of the working class in the American labor force also declined.

Given the political prominence of the claim about working-class neglect, it might seem surprising that there has been little attention paid to studies of this dimension of cultural exclusion. Racial and gender exclusion from popular culture are more often subjects of both academic and popular concern. And there is no equivalent dearth of interpretive studies of *how* working-class characters are portrayed in movies, television, and the popular culture. But the dimension of recognition versus neglect of the working class has received little systematic study.

One answer for this failure may be a lack of adequate tools to evaluate class presence and absence. Even simple content analyses of texts often depend on a search for a particular phrase or a handful of equivalent phrases. But the presence or absence of “working class” or “blue collar” exhibits the opposite pattern as the presence or absence of actual working-class job titles. The more abstract descriptive phrases are *more* common in the *Times* than in three of the four regional papers. And even in Wikipedia movie plot summaries where those phrases are rare (0.7%), its use is slightly increasing over time (β= +0.012), in contrast to the decline in actual working-class job titles (β= −0.011).

Having available a computer program that can code occupations from natural language texts may be as important a contribution of this research as evaluating any particular claim about popular culture and the working class in recent times. Occupations are basic data for a wide variety of sociological research, well beyond issues of class divisions. This initial program is a start, but more extensive use could result in improvements to the program and, especially, to the lexicon of job titles and their occupation codes.

The texts analyzed are, of course, only a small sample of U.S. popular culture. While the availability of digitized texts has grown rapidly in recent years, and the growth of these sources has contributed to the enthusiasm over text mining, easy access to long samples of newspapers or other cultural produces remains a problem. The digitized *Times* corpus only starts in 1980. While earlier editions are available from other sources, those records are now stored as .pdf files. Translating those files to the raw text files needed for text mining remains a major obstacle. The lack of texts from before 1980 is especially frustrating because the movie texts show some intriguing evidence of a resurgence of interest in the working class during the1970s that matches the earlier results for magazine advertisements (Paulson and O'Guinn 2012).

The availability of regional newspapers is even more daunting. Most regional newspapers have no digitized text archives and those that do date back a limited number of years. Even compiling a representative sample of raw text articles for any of these papers that do have digitized archives is a laborious task given the limitations imposed by the existing archives. It is possible that similar analyses of other coastal and regional newspapers would not support the patterns found here, but that is a task for future work once more accessible archives become available.

Access to digitized movie plot summaries is far easier because of the work of Wikipedia in collecting and producing them in a generally consistent format. Even here substantial text cleaning is needed before analysis can begin. The main limitation for movie plots is the dependence on Wikipedia which relies on volunteers for writing and editing plot summaries. No other source is as comprehensive as Wikipedia, but a comparison with even a limited number of movies plots from other sources – or, better, with a substantial library of movie scripts – would be instructive. However, no comparison text archive is readily accessible.

For now, another major limitation of these analyses of working-class neglect is the restriction of the text analysis to the single dimension of recognition versus neglect. If a newspaper or a movie includes a working-class character, there needs to be an analysis of how that character is portrayed. The analyses above show whether a working-class character is included or not, but we know nothing about what the newspaper or the movie is saying about that character. The next step for text analyses of working-class exclusion should be an exploration of the praise versus stigma dimension of cultural inclusion.

Beyond simple sentiment analyses, we need to inquire whether texts with working-class characters include themes typically associated with working-class virtues: solidarity (e.g., Deer Hunter), resistance (e.g., Norma Rae), and strong families (e.g., Marty) or with typically working-class stigmas like violence (e.g., Taxi Driver) and crime (e.g., On the Waterfront). While topic modeling might produce an inductive list of themes that could be tracked over time (DiMaggio, Nag, and Blei 2013), the literature on working-class cultural representation is sufficiently developed (e.g., Lamont 2000) that a more deductive set of themes can be identified and coded using traditional content analysis procedures. These coded texts should then provide the needed training set for developing a more automated analysis of the much larger number of texts needed to track annual changes.

An era of sociological analyses of large text databases is surely coming (Bail 2014), and it will change the discipline just as the arrival of large numerical databases did. The two are not in opposition but can complement each other. If we are fortunate, text mining could provide an intermediate line of research that bridges the gap between quantitative social science and more interpretive studies of cultural change.

References

Bail, Christopher A. 2014. "The Cultural Environment: Measuring Culture with Big Data." *Theory and Society* 43(3): 465 - 482.

Bodnar, John. 2003. *Blue-Collar Hollywood: Liberalism, Democracy, and Working People in American Film*. Baltimore, MD: Johns Hopkins University Press.

Bureau of Labor Statistics. 2010. "Direct Match Title File." https://www.bls.gov/soc/soc_2010_direct_match_title_file.xls

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41(6): 570-606.

Ehrenreich, Barbara and John Ehrenreich. 1979. "The professional-managerial class." Pp. 5-45 in Pat Walker (ed.), *Between Labor and Capital*. Boston: South End Press.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles and J. Robert Warren . 2018. "Integrated Public Use Microdata Series, Current Population Survey: Version 6.0." IPUMS. https://doi.org/10.18128/D030.V6.0

Kendall, Diana. 2005. *Framing Class: Media Representations of Wealth and Poverty in America*. Lanham, MD: Rowman & Littlefield.

Lamont, Michèle. 2000. *The Dignity of Working Men: Morality and the Boundaries of Race, Class, and Immigration*. Cambridge, MA: Harvard University Press.

--- 2018. "Addressing Recognition Gaps: Destigmatization and the Reduction of Inequality." *American Sociological Review* 83(3): 419-444.

Lamont, Michèle, Bo Yun Park, and Elena Ayala-Hurtado. 2017. "Trump’s Electoral Speeches and His Appeal to the American White Working Class." *The British Journal of Sociology* 68(S1): S153-S180.

Misra, Joya, Stephanie Moller, Marina Karides. 2003. "Envisioning Dependency: Changing Media Depictions of Welfare in the 20th Century." *Social Problems* 50(4): 482-504.

Paulson, Erika L. and Thomas C. O’Guinn. 2012. "Working-Class Cast: Images of the Working Class in Advertising, 1950-2010." *The Annals of the American Academy of Political and Social Science* 644: 50-69.

Poulantzas, Nicos. 1974. *Classes in Contemporary Capitalism*. London: New Left Books.

Rennels, Tasha R. 2015. "Here Comes Honey Boo Boo: A Cautionary Tale Starring White Working-Class People." *Communication and Critical/Cultural Studies* 12(3): 271-288.

Rose, Max and Frank R. Baumgartner. 2013. "Framing the Poor: Media Coverage and U.S. Poverty Policy, 1960-2008." *Policy Studies Journal* 41(1): 22-53.

Ross, Steven Joseph. 1998. *Working-Class Hollywood: Silent film and the Shaping of Class in America*. Princeton, NJ: Princeton University Press.

United States Bureau of the Census. 2016. "2016 Census Occupation Index.", published online September, 2016. https://www2.census.gov/programs-surveys/demo/guidance/industry-occupation/occupation-index-september-2016.xlsx

Van de Rijt, Arnout, Eran Shor, Charles Ward, and Steven Skiena. 2013. "Only 15 Minutes? The Social Stratification of Fame in Printed Media." *American Sociological Review* 78(2): 266-289.

Vanneman, Reeve. 1977. "The occupational composition of American classes." *American Journal of Sociology* 83: 783-807.

---. 2019. “jobs.py: A Python Program to Code Occupations in Text Files.” url: https://github.com/ReeveVanneman/occupations.

Wright, Erik Olin and Luca Perrone. 1977. "Marxist class categories and income inequality." *American Sociological Review* 42: 32-55.