APPENDIX

A Word Embedding Model Specifications

A.1 Algorithmic Details

Consider the following example, where the occupation word *engineer* and gender word *he* both appear close to local context words (*i.e.*, words that surround the central word), *work* and *company*, in the 5-gram case A and B. SGNS positions the two words in a closer vector space than with words such as *housewife*, which may rarely appear in the above context of workplace. By design, adjacent words *engineer* and *he* as in case C also share an overlapped context (*successful* and *studies*) and a closer vector space.

$$\ldots\ \textit{in}\ \overbrace{\underbrace{\textit{most}}_{\text{context}}\ \underbrace{\textit{cases}}_{\text{context}}\ \underbrace{\textit{engineers}}_{\textit{central}}\ \underbrace{\textit{work}}_{\text{context}}\ \underbrace{\textit{in}}_{\textit{stopword}}\ \underbrace{\textit{companies}}_{\text{context}}}^{\textit{5-gram}}\ \ldots \tag{A}$$

$$\ldots\ \textit{in}\ \textit{his}\ \overbrace{\underbrace{\textit{working}}_{\text{context}}\ \underbrace{\textit{age}}_{\text{context}}\ \underbrace{\textit{he}}_{\textit{central}}\ \underbrace{\textit{owns}}_{\text{context}}\ \underbrace{\textit{a}}_{\text{stopword}}\ \underbrace{\textit{company}}_{\text{context}}}^{\textit{5-gram}}\ \ldots \tag{B}$$

$$\ldots\ \textit{to}\ \overbrace{\textit{become}\ \textit{a}\ \underbrace{\textit{successful}\ \textit{engineer}\ \textit{he}\ \textit{studies}}}^{\textit{5-gram window I}}\!\!\underbrace{\ \textit{hard}}_{\textit{5-gram window II}}\ \textit{in}\ \ldots \tag{C}$$

One of the desired properties of the word embedding algorithm is that the model is intrinsically *relational*: as word meanings are represented by their relations to other context words, the words that never appear in the same text window (the first-order connection), but co-occur with a third word (the second-order connection) or with words that share close semantic contexts (third-order connection and beyond), would also share a close vector space, which enables a recovery of subtle and diffuse semantic relations in the whole corpus. For example, phrases such as *male nurse* appear more frequently than *female nurse* in publications, because common assumptions are often left unsaid (Gordon and Van Durme 2013). However, *nurse* is closer in the vector space with female-signaling phrases than with male ones, suggesting that second-order relations and beyond are captured by SGNS.

SGNS achieves the above goals by maximising the likelihood $L(\theta)$ that context words w_{t+j} appear within a m-sized window given the central word w_t for each position $t = 1, ..., T$ in the whole corpus:

$$L(\theta) = \prod_{t=1}^{T} \prod_{-m \leq j \leq m} P(w_{t+j}|w_t, \theta)$$

where θ are all parameters to be optimized. m is set to be 2 as per the 5-gram structure. In SGNS, each word w is represented by two vectors, v_w when w is a central word, and u_w when w is a context word; both v_w and u_w are set to be 300-dimension vectors, whose parameters ($i.e.$, dimension values) are to be optimized as θ.

The objective loss function $J(\theta)$ is the average negative log-likelihood across all possible central word positions $t \in \{1, ..., T\}$ when the m-sized window moves across the corpus:

$$Loss = J(\theta) = -\frac{1}{T} \sum_{t=1}^{T} \sum_{-m \leq j \leq m} \log P(w_{t+j}|w_t, \theta)$$

$$\text{where } P\left(w_{t+j}|w_t, \theta\right) = \left(exp\left(u_{w_{t+j}}^T v_{w_t}\right)\right) \left(\sum_{k \in V} exp\left(u_{k_{t+j}}^T v_{w_t}\right)\right)^{-1}$$

The softmax function for $P\left(w_{t+j}|w_t, \theta\right)$ models a multi-classification task, where the number of classes are the unique words V in the corpus. To minimize $J(\theta)$, the optimal θ is approximated by gradient descending with learning rate α, such that in a single pair of central word and one of its context words:

$$\theta^{new} = \theta^{old} - \alpha \nabla_\theta J(\theta)$$

In the above example B, the loss function for a single pair (central word *he* and context word lemma *work*) is:

$$
\begin{aligned}
J_{t,j}(\theta) &= -\log P(work|he) \\
&= -\log \frac{exp(u_{work}^T v_{he})}{\sum_{k \in V} exp(u_k^T v_{he})} \\
&= -u_{work}^T v_{he} + \log \sum_{k \in V} exp(u_k^T v_{he})
\end{aligned}
$$

with parameters updated as:

$$
\begin{aligned}
v_{he} &:= v_{he} - \alpha \frac{\partial J_{t,j}(\theta)}{\partial v_{he}} \\
u_k &:= u_k - \alpha \frac{\partial J_{t,j}(\theta)}{\partial u_k}, \forall k \in V
\end{aligned}
$$

Accordingly, the algorithm increases the similarity (represented by dot product) of v_{he} and u_{work}, and at the same time, decrease the similarity between v_{he} and all other words k in the corpus. To save time in computation, instead of updating all k words, only a small set of random words are selected to be "negatively" sampled. When training the data, I negatively sample 5 words in each update.

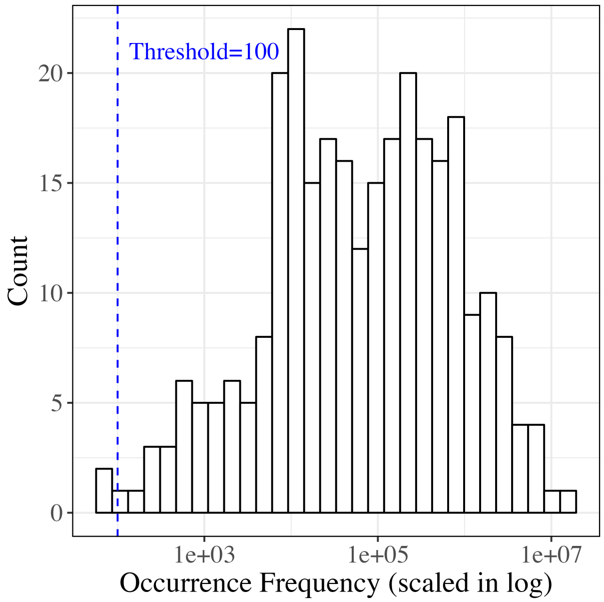
Hyperparameters are tuned according to Mikolov et al. (2013), where words that appear at least 100 times in 5-gram and 20 times for COCHA in each decade are vectorized in a 300-dimension space; frequent words, including common stop-words such as "the" and "or", are down sampled and randomly excluded; 5 epoches are used to train and update word vectors. Co-occurrence is defined in a 5-word window as per the 5-gram structure. Words are lower-cased, but not lemmatized nor stemmized, and punctuations that may include semantic and emotional information are included in the training stage. Each embedding vector is normalized to have length 1. I use the *word2vec* module of *gensim* (*ver.* 4.3.0), the mainstream Python package for training the model.

A.2 Analyzed Corpora and Occupation Embeddings

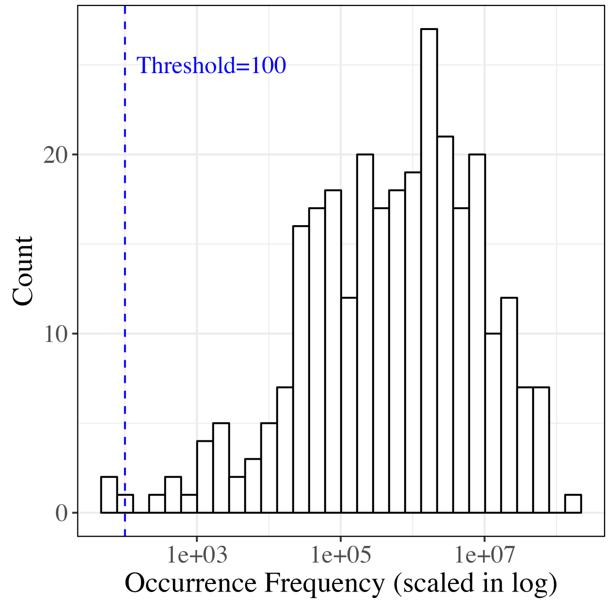
The total number of words that appear in the three corpora by decade are listed in Table S1. Google Ngram include over 126 billion words over a century, while COHA and COCA combined consist of around another 1 billion. Figure S1 shows the distribution of title occurrences in Ngram in 1930-1939 (the decade with the least total words) and 2000-2009. In both decades, the vast majority of single-word titles exceed the minimum frequency threshold to be included in the training process. The median frequency is 72,563 and 642,767 for the two decades, suggesting a non-trivial presence of occupation titles in the corpus. With all unique single words that appear in the two corpora being vectorized in a 300-dimensional semantic space, Figure S2 maps each vector (Ngram, 2000-2009) onto a 2-dimensional space through t-Distributed Stochastic Neighbor Embedding—a common technique that reduces vector dimensions while preserve the clustering structure in the higher dimension.

Table S1: Number of words in corpus

	number of words		
Decade	5-gram	COHA	COCA
1900	11.1B	20.3M	
1910	10.1B	20.7M	
1920	7.1B	24.0M	
1930	5.8B	23.0M	
1940	6.2B	22.8M	
1950	8.1B	23.0M	
1960	13.2B	22.4M	
1970	14.0B	22.1M	
1980	15.5B	23.4M	
1990	19.8B	25.9M	227.1M
2000	26.9B	27.2M	228.9M
2010			227.5M
total	126.4B	234.5M	683.5M

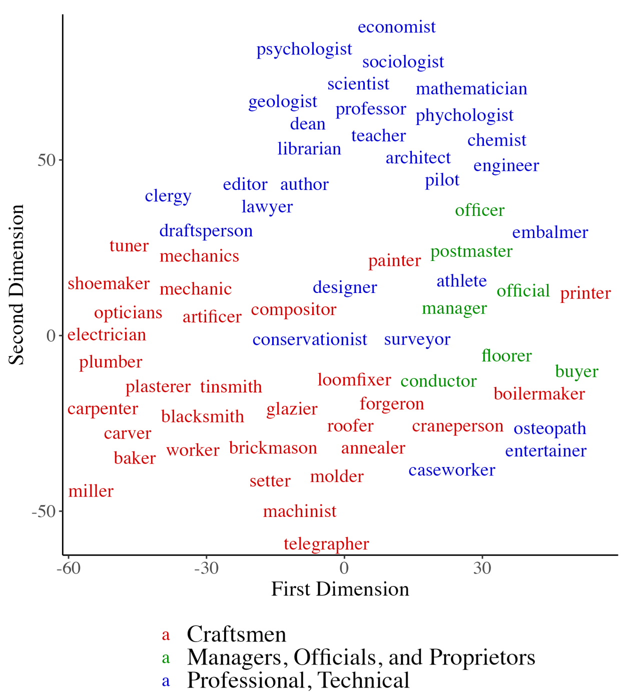


Panel A: 1930-1939, Ngram

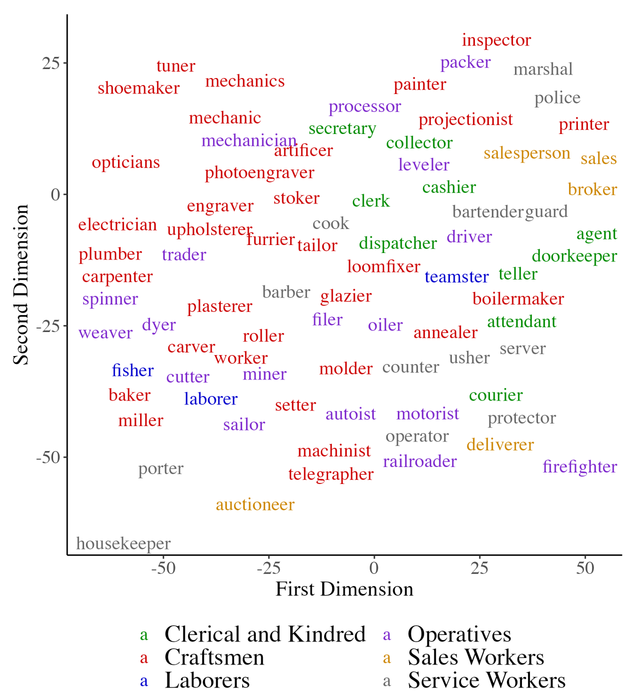


Panel B: 2000-2009, Ngram

Figure S1: Frequency of Occurrences of Occupation Titles in Ngram



Panel A: Occupation Group 1



Panel B: Occupation Group 2

Figure S2: t-Distributed Stochastic Neighbor Embedding dimension reduction of occupation vectors, Google Ngram, 2000-2009

Notes: perplexities in t-SNE are set to be 10, meaning that up to 10 neighbors are allowed in the 2-dimensional space.

B Semantic Dimensions Essential

B.1 List of Antonym Words

Table S2 to S4 present the word pairs that are used to construct the various semantic dimensions of occupations' cultural properties. When compiling the list, I mainly refer to the work of Kozlowski et al. (2019) and Van Loon and Freese (2023), who relied on multiple thesauri, including three contemporary thesauri: Bartlett's Roget's Thesaurus, Oxford Thesaurus, and Webster's Collegiate Thesaurus; and two historical thesauri (Roget 1911; Smith 1910). I also expand the list of words that may encode cultural association with women or men from the work of Garg et al. (2018).

Table S2: Word pairs that construct semantic dimensions

Gender		**Prestige**		**Cultivation**	
he	she	prestigious	unprestigious	educated	uneducated
daughter	son	honorable	dishonorable	learned	unlearned
hers	his	esteemed	lowly	knowledgeable	ignorant
her	him	reputable	disreputable	trained	untrained
mother	father	distinguished	commonplace	taught	untaught
woman	man	eminent	mundane	literate	illiterate
girl	boy	illustrious	humble	schooled	unschooled
herself	himself	renowned	humble	tutored	untutored
female	male	acclaimed	prosaic	lettered	unlettered
sister	brother	dignitary	modest	-	-
aunt	uncle	venerable	commoner	-	-
niece	nephew	exalted	unpretentious	-	-
masculine	feminine	estimable	ordinary	-	-
women	men	prominent	common	-	-

Table S3: Word pairs that construct semantic dimensions, continued 1

Affluence		**Affluence (cont.)**		**Evaluation**	
rich	poor	luxurious	threadbare	good	evil
affluence	poverty	posh	plain	moral	immoral
affluent	destitute	moneyed	unmonied	good	bad
wealthy	impoverished	exorbitant	impecunious	honest	dishonest
costly	economical	advantaged	needy	virtuous	sinful
expensive	inexpensive	opulent	indigent	virtue	vice
exquisite	ruined	plush	worthless	righteous	wicked
extravagant	necessitous	privileged	underprivileged	chaste	transgressive
flush	cheap	propertied	bankrupt	ethical	unethical
lavish	economical	prosperous	unprosperous	unquestionable	questionable
luxuriant	penurious	developed	underdeveloped	uncorrupt	corrupt
solvency	insolvency	sumptuous	plain	scrupulous	unscrupulous
swanky	basic	thriving	disadvantaged	altruistic	selfish
upscale	squalid	valuable	valueless	chivalrous	knavish
classy	beggarly	ritzy	ramshackle	honest	crooked
opulence	indigence	solvent	insolvent	commendable	reprehensible

Table S4: Word pairs that construct semantic dimensions, continued 2

Potency		**Activity**		**Evaluation cont.**	
strong	weak	loud	quiet	pure	impure
powerful	powerless	fast	slow	holy	unholy
deep	shallow	hot	cold	valiant	fiendish
thick	thin	sharp	dull	upstanding	villainous
large	small	burning	freezing	guiltless	guilty
complex	simple	active	inactive	unquestionable	questionable
difficult	easy	intense	calm	decent	indecent
many	few	young	old	chaste	unsavory
competent	incompetent	loose	firm	righteous	odious

B.2 Semantic Dimensions' Inter-Associations

Cultural dimensions, such as gender and general prestige, are generally not orthogonal in the vector space. Figure S3 shows the decade-specific cosine similarity between gender and other dimensions in the semantic space in the past century. Notably, gender (levels of femininity) is negatively linked with prestige words, while a higher level of femininity positively predicts affluence. The negative association between femininity and cultivation that appeared in the first several decades was reversed in the late twentieth century, which is consistent to the general trend where the gender education gap converged and was reversed in recent years.

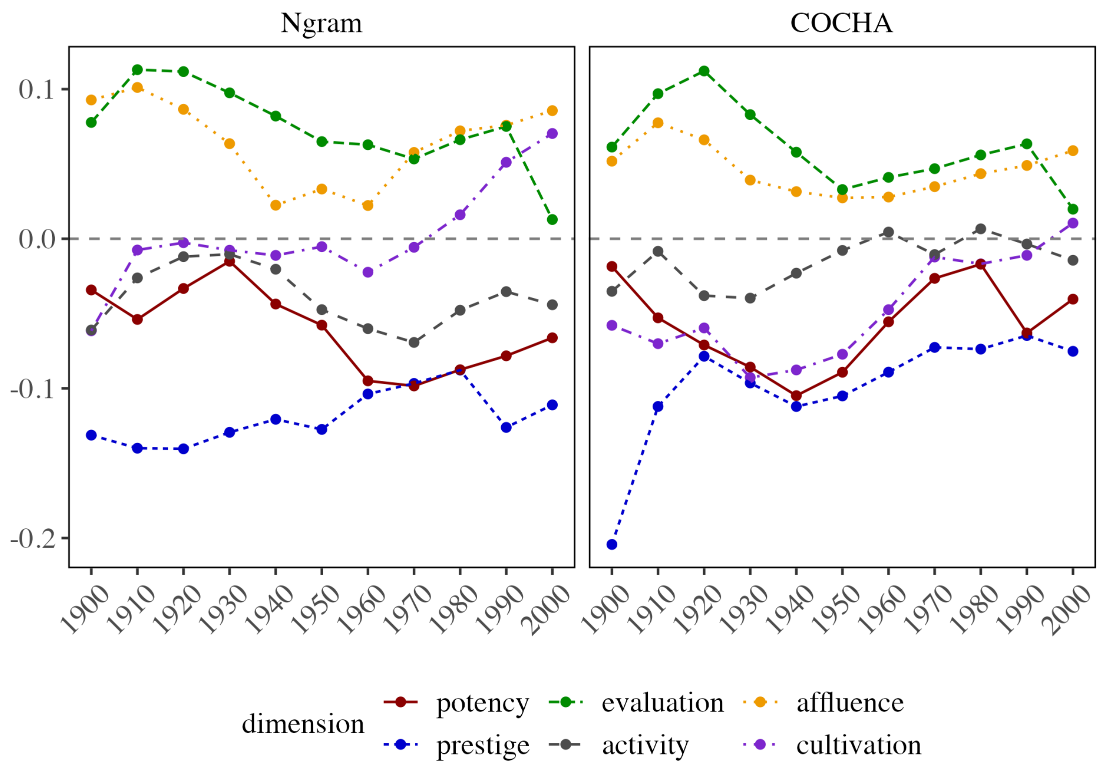


Figure S3: Cosine Similarity between Gender and Other Dimensions, Ngram and COCHA, 1900-2009

C Comparison with the Occupation Embeddings used by Garg et al. (2018)

Using word embedding models to understand occupational gender typing is not a completely new idea. Garg et al. (2018), for example, identified a number of occupation titles (in 1950 COC code) in historical publications, and studied their cultural association with women using the same word embedding model as in this paper. However, our embeddings differ in three important ways. First, Garg et al. (2018)'s analysis focused the COHA corpus,[1] which was trained by Hamilton et al. (2016) in a parsimonious way, where only words that appear more than 400 times are used in the model. This procedure produces a much smaller set of words with their embeddings than my approach does, which includes any words that appear over 20 times. As COHA (or COCHA that combines COHA and COCA) is a relatively small corpus, many occupations were *not* represented in Garg et al. (2018) due to the strict criterion, while they *did appear* in my embeddings. This may be seen from Figure 1 in Garg et al. (2018), where fewer than 70 unique occupations appeared in their validation process, while mine includes over 200 occupations. My approach ensures that the text-based measures of occupational gender typing and prestige adequately reflect the general occupation profiles in the labor market.

Second, when mapping occupation titles into single words, Garg et al. (2018)'s approach misrepresents a number of occupations. According to their replication file, many occupations that have unique working contexts, including dyers, glaziers, furriers, bookbinders, electricians, and spinners (textile), were all summarized by the single word "tradsperson" without explicit reasons.[2] Surprisingly, the word "tradesperson" itself never appeared in COHA. These occupations, therefore, were simply omitted in their analysis by construction. In my embeddings, I managed to keep these occupations' original names, most of which were found in the two corpora (Ngram and COCHA). I also corrected several problematic mappings: for example, waiters and waitresses were summarized as waitstaff in Garg et al. (2018), while the word never appeared in the COHA embedding they used. I revised it into server, which appears more than 20 times in COCHA. For a detailed mapping crosswalk used in this paper, see `linktobeupdated`.

A final difference with Garg et al. (2018) and the embeddings they used is that they

[1] Although the authors claimed in the main text that they used both COHA and Google Ngram, in the replication site, `https://github.com/nikhgarg/EmbeddingDynamicStereotypes`, they corrected that they only used COHA.

[2] See the replication file from `https://github.com/nikhgarg/EmbeddingDynamicStereotypes/blob/master/data/occupation_map.csv`

never trained bi-gram occupation titles. For example, one may train *chemical engineer* as a "single" word with its own embedding rather than approximate it using a combination of *chemist* and *engineer*. To test whether results are sensitive to the way words are embedded, I trained bi-gram embeddings from COCHA and present the TWFE and the generalized robust estimates of the effect of female typing on prestige in Figure S4. These bi-gram embeddings arguably produce a more accurate representation of multi-word occupations in the vector space. Similar to the main findings, I find a robust negative effect of occupation's stereotypical association with women on its general prestige and potency, but not moral standing or liveliness. Results suggest that the main findings are not mainly driven by the way some occupation titles are split into single words.

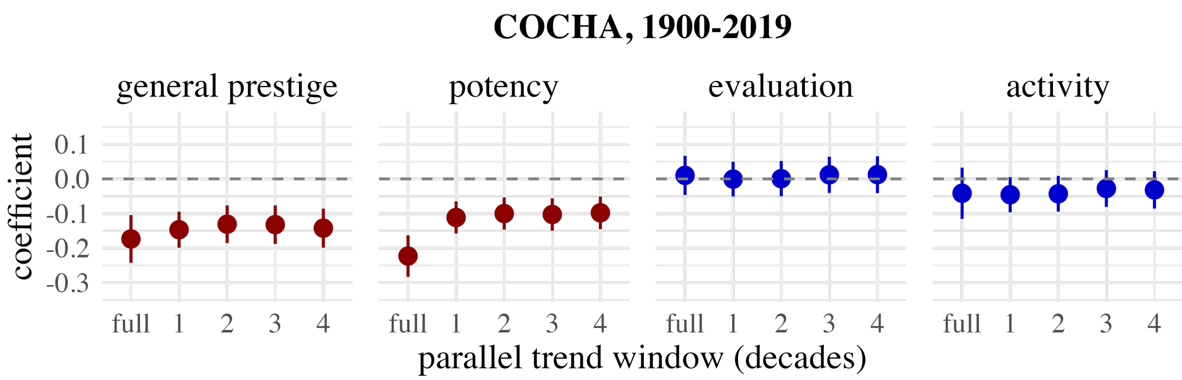


Figure S4: Cross-sectional association between female typing and prestige, bi-gram, COCHA
Notes: Error bar in this and in the following figures represents the 95% confidence interval (two-tailed tests).

Another approach to check the robustness of the results against how multiple single-word occupation titles are used to represent multi-word occupations is to weight single-word titles by their occurrence frequency in a specific decade. For example, the gender typing of *chemical engineer* in decade t would become the weighted mean of cosine similarity between *chemist* and the gender dimension and between *engineer* and the gender dimension, with weights proportion to the number of occurrences of words *chemist* and *engineer* in decade t, respectively. Figure S5 presents the main TWFE and generalized robust estimates of the effect of female typing on the four dimensions of prestige. The penalties of culturally feminized occupations in general prestige and potency remain statistically significant.

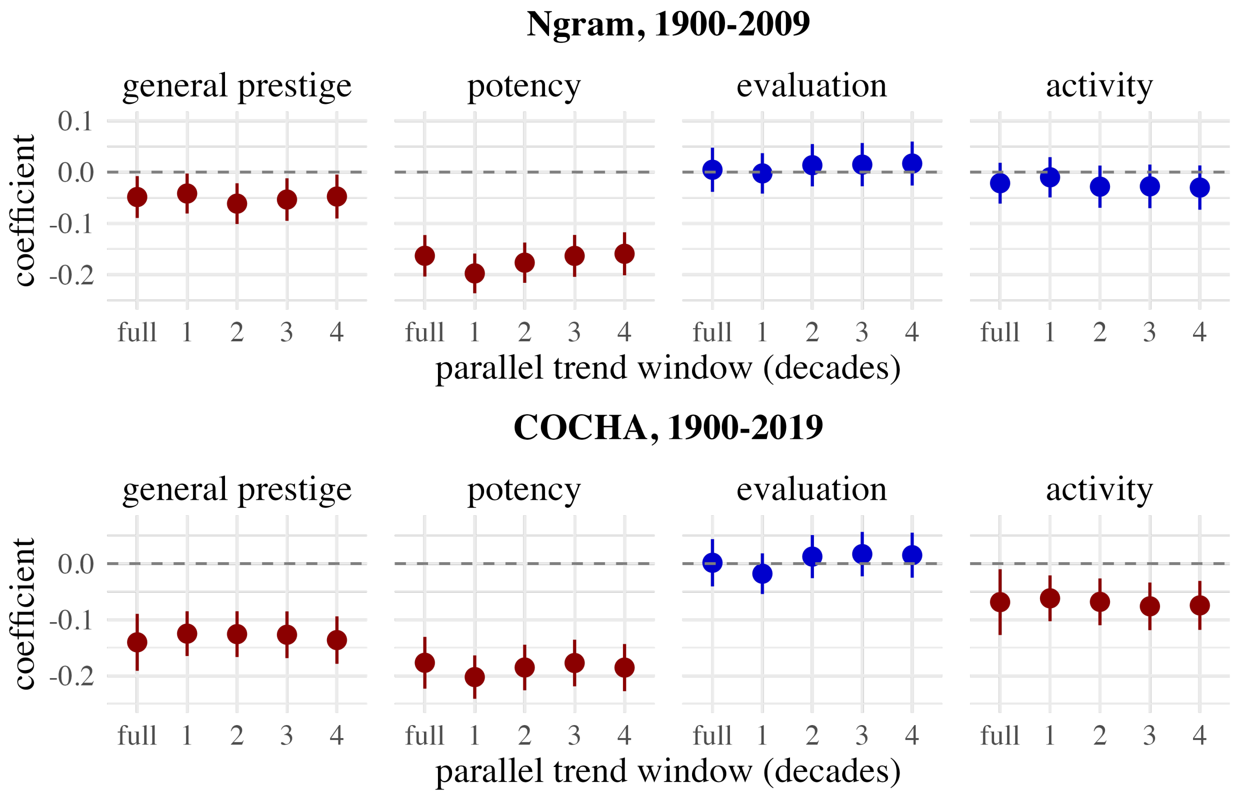


Figure S5: Generalized Robust TWFE estimates using Ngram and COCHA, 1900-2019. Cosine similarities between occupation titles and cultural dimensions are weighted by occupation titles' occurrence

Notes: Error bars represent the 95% confidence interval; the 95% confidence intervals that do not cover zero are colored in red.

D Additional Validation of Occupation Embeddings

I supplement four additional sets of validations of embedding-based measures of gender typing and prestige. First, following Garg et al. (2018), Figure S6 compare the overall trend of embedding-based mean (weighted) female typing and the proportion of women in the labor market. Results suggest that, despite several exceptions in some decades, female typing embedded in text generally "growed" over the century (*i.e.*, gender typing moved from the masculine side to neutrality), reaching the highest level in the 1950s and 60s (Ngram) when the Civil Right Movement was at its peak, and in the most recent decade (COCHA and

Ngram). Notably, Ngram-based measures of gender typing were greater than 0 in the 1960s, 1970s, and 2000s, meaning that the overall gender perception of the workforce was feminine rather than masculine. By contrast, even though the perception gradually approached 0 in COCHA, gender neutrality was never reached. The difference may stem from the different sources of language in the two corpora: Ngram may over-represent scientific works (Pechenick et al. 2015) that encode fewer social stereotypes than the non-scientific ones and the general public's conception as represented by COCHA (Shi et al. 2021).

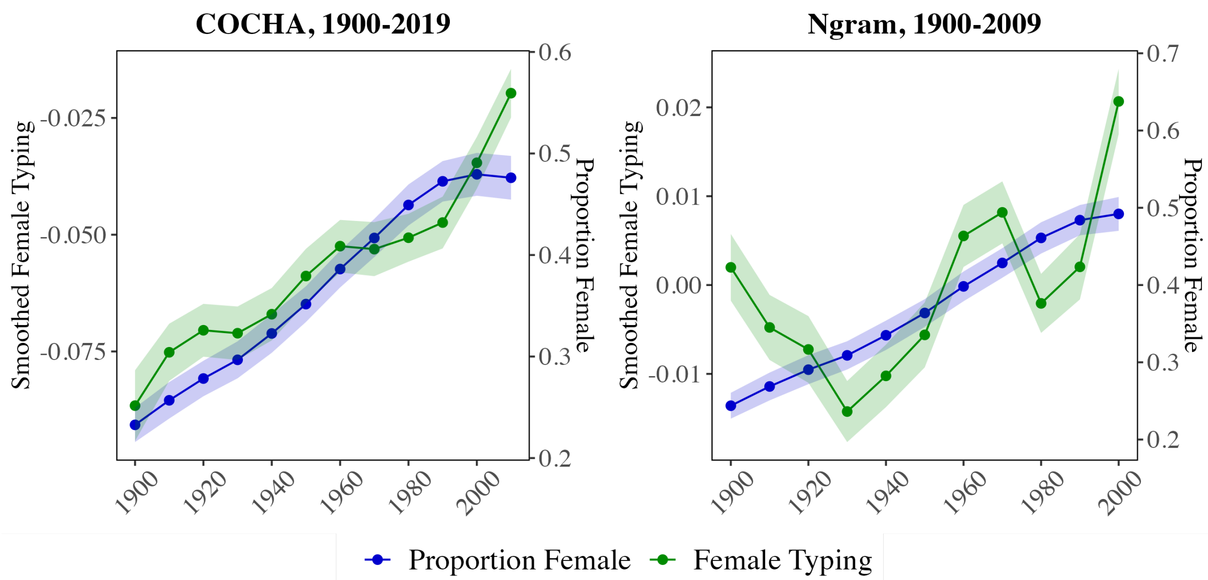
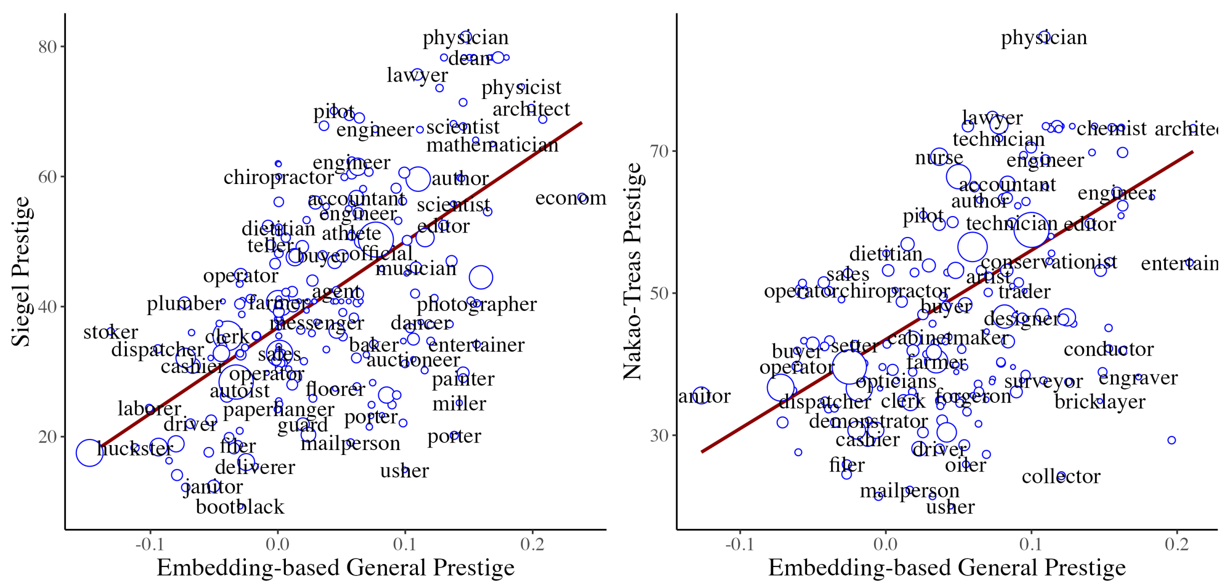


Figure S6: The mean female typing in text and actual female proportion in the labor market, 1900-2019
Notes: Standard errors (95%) are derived from bootstrapping samples for 2000 times

Second, I compare the text-based (COCHA) and survey-based measures of general prestige of each occupations and visualize their relationship in Figure S7. In both cases I find a statistically significant correlation between the two measures in their corresponding years; deans, professors, physicists, architects, for example, are highly revered in both text and survey responses, while laborers, attendants, and janitors are less valued than most other occupations. It is worth noting that some discrepancies between the embedding- and survey-based measures may stem from the limitations of the word embedding model itself. For example, in both decades, COCHA document a lower prestige of *physicians* and *lawyers* than the respondents of GSS reported. As physicians and lawyers often appear in a semantic context where suffering, poverty and/or crime co-exist, their prestige may be downward biased towards the objects they serve and interact with (Kmetty et al. 2021). By contrast, the prestige of *ushers* who often attend the upper-class and appear along with high-brow culture may be over-estimated (see Panel B). It is important to note, however, that a precise alignment between the two measures are not required: as the main analyses rely entirely on the text data, it is the *within-occupation change* rather than *level* of the measure that lies at the core of the interest. Any systematic biases of the embedding that are stable over time are modeled by the occupation-specific fixed effects term in the main analyses, which ensures that the estimates are based on the semantic shift of the occupation, rather than any potentially biased measures at a single point.

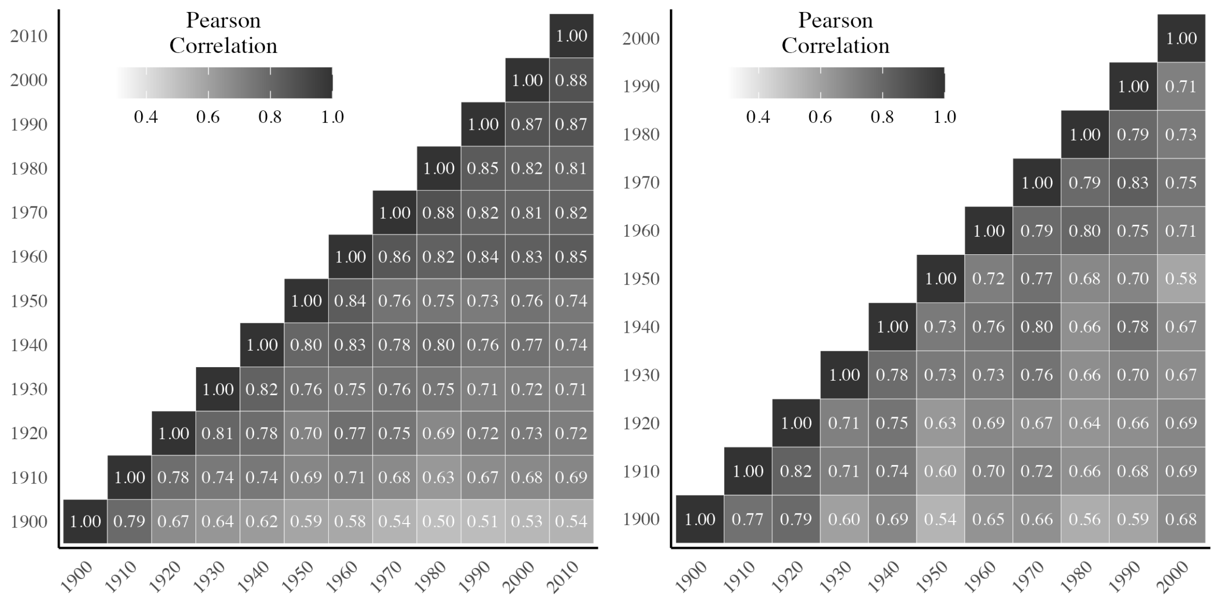


Panel A: baseline Siegel (1971)

Panel B: baseline Nakao and Treas (1994)

Figure S7: Comparison between embedding- and survey-based general prestige measures
Note: Circle sizes are scaled to reflect the relative employment size of each occupation in the decade shown.

Third, I test whether occupational *general* prestige embedded in text, as Treiman (1977) found using survey data, remain largely stable over time. Figure S8 presents the pearson correlation between all possible decades. Two substantive patterns emerge. First, occupation prestige between adjacent decades are highly correlated: in both COCHA and Ngram-based measures, prestige of an earlier decade strongly correlates with the current decade ($\rho \approx 0.8$). The inertia persists even over a century, where the prestige measured in 1900-1909 is still strongly correlated with the prestige in 2000-2009 ($\rho \approx 0.6$). Second, as expected, strongest correlations are typically found between decades that are temporally close; the association becomes weaker as the time span increases. Results remain similar when the first principal component of the four prestige dimensions are used (omitted).



Panel A: COCHA

Panel B: Ngram

Figure S8: Pearson correlation of occupation prestige over decades, 1900-2009

Finally, I conduct the *3CosAdd* test (Mikolov et al. 2013) common for word embedding, that is, I search for the word whose vector is closest to $\overrightarrow{x} - \overrightarrow{man} + \overrightarrow{woman}$; when $x = king$, the target solution should be $\overrightarrow{queen}$. Although there are no standard correct answers, the validity of the embedding is typically judged by common sense. Table S5 shows the closest occupation to 10 example occupations x in 2000-2009 across the three corpora. Two general patterns appear. First, all the three embeddings are correct in the basic grammatical understanding of occupation. The female titles of actors and waiters, not surprisingly, are actress and waitress. Second, women in many cases are still the "second sex" being subordinate to men in the semantic space. For example, the female counterparts of drivers, soldiers and commanders are passengers, citizens, and lieutenants, respectively, who are degraded to be inferior in status and power. Overall, the test results are consistent with common reasoning and gender theories.

Table S5: 3CosAdd test, 2000-2009

occupation x	Ngram	COHA	COCA
actor	actress	actress	actress
waiter	waitress	waitress	waitress
doctor	therapist	gynecologist	nurse
dentist	psychiatrist	gynecologist	nurse
sheriff	receptionist	coroner	coroner
professor	teacher	psychologist	sociologist
soldier	citizen	medic	servicewomen
commander	lieutenant	lieutenant	lieutenant
lawyer	therapist	schoolteacher	attorney
driver	passenger	passenger	passenger

E Supplements for Main Results

E.1 Temporal Heterogeneity of TWFE Results

In the main text, I presented the results when the temporal window for the parallel trend to be plausible is restricted within four decades. In other words, in the generalized robust FE estimate, I explored the sensitivity of the results with $\ell \in \{1, 2, 3, 4\}$ from Equation 1.

$$\hat{\beta}_{FE}^{robust} = \left(\sum_{k=1}^{\ell} \hat{w}_k^D \hat{\beta}_{FD,k}^D\right) \left(\sum_{k=1}^{\ell} \hat{w}_k^D\right)^{-1} + \hat{B} \tag{1}$$

To show the effects in the longer term ($\ell > 4$), Figure S9 presents $\hat{\beta}_{FD,k}^D$ and $\hat{w}_k^D$ for each possible k in Ngram and COCHA, and the term $\hat{B}$ when general prestige and potency is used as the outcome variable, respectively. I find two large patterns that corroborate the original findings. First, in virtually all FD periods, the FD estimates of the effect of female typing on general prestige and potency are significantly negative at p=0.05; the negative effects are generally larger in the long run, and the estimates as presented in the main text are likely to be the lower bound. Given the dramatic cultural and demographic changes in the twentieth century, the stability of the cultural bias against occupations stereotypically associated with women, both in the short- and in the long-run is probably striking. Second, the bias term $\hat{B}$ is trivial compared to the magnitude of the main FD estimates. Therefore, as I mentioned in the text, whether $\hat{B}$ is included in Equation 1 or not, or whether placed in the numerator or outside of the fraction does not change the results.

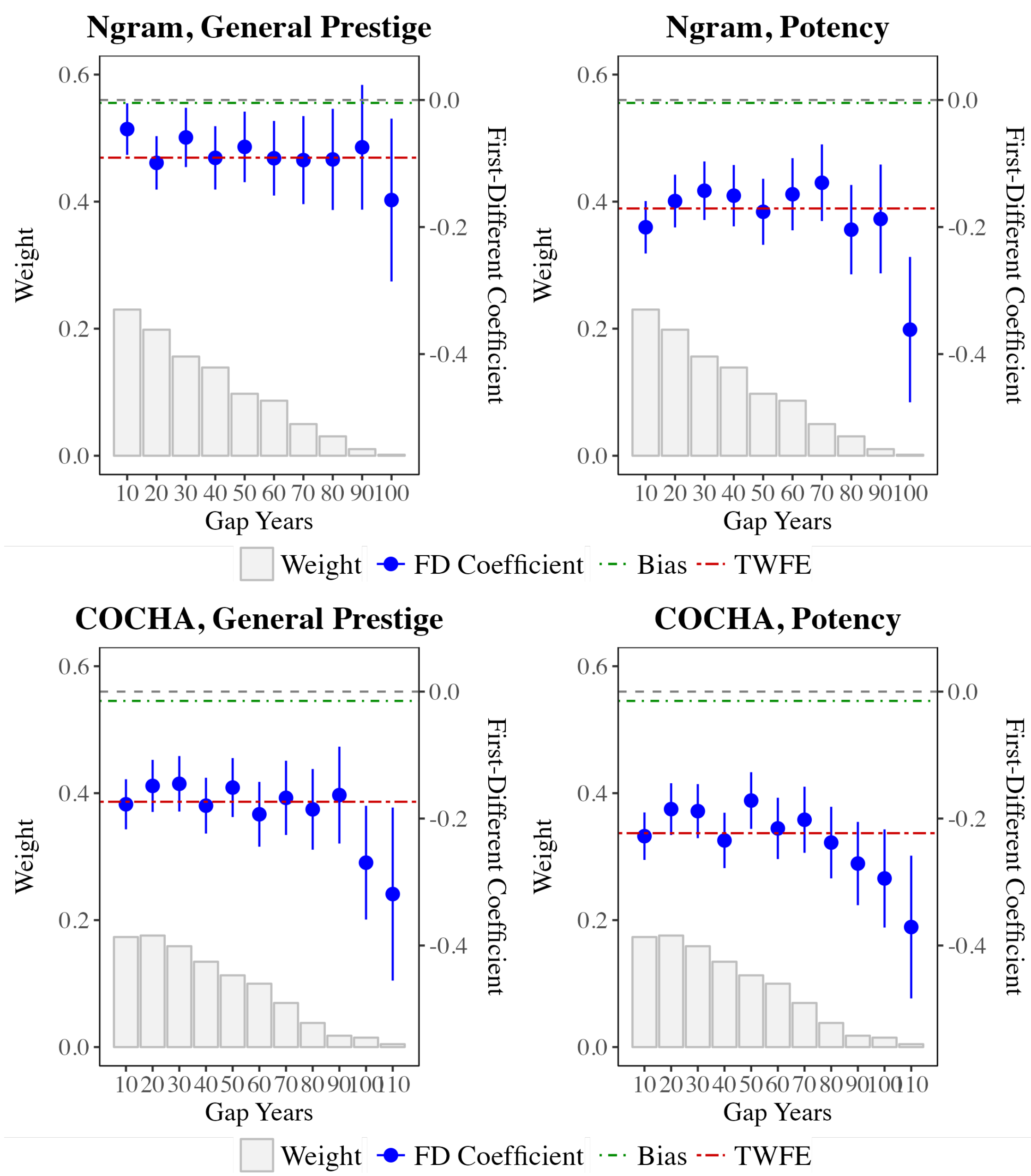


Figure S9: The FD estimates for each possible k and bias $\hat{B}$ using Ngram and COCHA, 1900-2019

Notes: Error bars represent the 95% confidence interval.

It is important to note that the k-year FD estimates as presented in Figure S9 do not speak directly to the effect in a specific decade, but mix in the average effect between all possible k-decade gaps over the century. I test the temporal hetereogeneity of the effect alternatively by interacting the occupation's cultural association with women with year fixed-

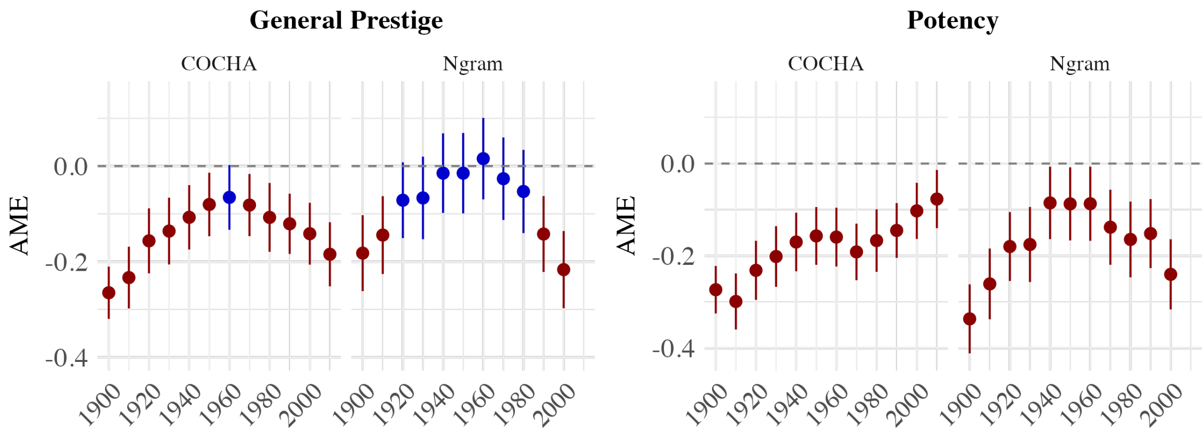


Figure S10: Average Marginal Effect (AME) of occupations' female typing on their general prestige and potency by decade, COCHA and Ngram

effects, and summarize their average marginal effects (AME) over time in Figure S10. I find a consistent under-valuation of women's work in the potency dimension throughout the study period for the two corpora, with no clear evidence for the decline (Crawley 2014) nor reinforcement (Mandel 2018) of devaluation in recent decades. The penalties on general prestige similarly appear in COCHA consistently across time, but present a U-shaped trend in Ngram, where significant devaluation in the general prestige dimension appears in the first and last few decades of the analyses. The explicit reason for the heterogeneity is not clear: it is possible that the over-representation of scientific publications down-weighted most sexist views in the general public in the mid-twentieth century, while popular magazines and movie scripts as represented in COCHA still encoded cultural biases against female labor.

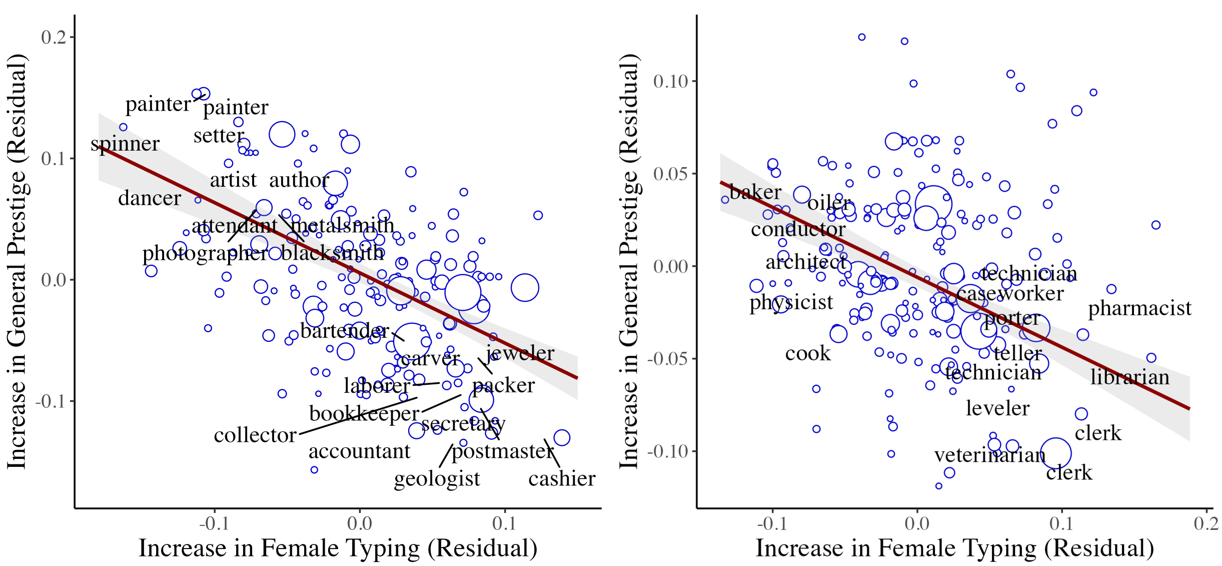
E.2 Occupation-Specific Changes in Cultural Feminization and Prestige

Figure S11 applies the Frisch-Waugh-Lovell Theorem (FWL) to control for the two cultural dimensions (affluence and cultivation) and shows the FD estimate (the slope) about the changes of the stereotypical association with women on the general prestige of occupations in COCHA given by:

$$\hat{\beta}_k = \mathrm{Cov}\left(\Delta k D_{it} - \Delta k W_{it}' \hat{\delta}^W, \Delta k Y_{it} - \Delta k W_{it}' \hat{\mu}^W\right) \left[\mathrm{Var}\left(\Delta k D_{it} - \Delta k W_{it}' \hat{\delta}^W\right)\right]^{-1} \qquad (2)$$

where $\hat{\delta}^W$ and $\hat{\mu}^W$ are the coefficients of $\Delta k D_{it}$ and $\Delta k Y_{it}$ on $\Delta k W_{it}$, respectively. Results on the dimension of potency are similar and omitted.

Embedding-based measures suggest that typical lower- to middle-income occupations, including cashiers, accountants, secretaries, bookkeepers, and some manual work in vegetable and fruit grading and packing (packers) in the first half of the twentieth century, and clerks, librarians, pharmacists, and (bank) tellers in the second half experienced most increases in female typing; their symbolic values in general prestige, consistent with the predictions of the devaluation hypothesis, decline. Virtually no curvilinear relations are found.



Panel A: FD Model, 1900-1949 Panel B: FD Model, 1950-2009

Figure S11: Change in female typing and general prestige, residuals from FWL with controls of affluence and cultivation, COCHA

Note: Circle sizes are scaled to reflect the relative employment size of each occupation averaged over the study period.

E.3 Cross-Sectional Association between Gender Typing and Each Dimension of Prestige

While the paper's main analyses focus on the longitudinal association between gender typing and prestige, some prior studies also examined the cross-sectional relationship between occupations' gender composition and survey-based prestige (England 1979; Treiman and Terrell 1975). As I discussed in the main paper, the results have been mixed, with some studies finding a negative association between female share and prestige (Bose and Rossi 1983), while others finding a curvilinear relation, with gender-segregated occupations (*i.e.*, predominantly male or female occupations) having higher prestige (Valentino 2020).

To examine the cross-sectional association between female typing and multiple dimensions of prestige embedded in text, I proceed with two models. First, Figure S12 shows the linear coefficient of female typing on prestige in each decade when only the first-order term is included (COCHA and Ngram pooled with corpus fixed-effects; results are similar when each corpus is evaluated separately). I highlight two findings. First, perceived potency and general prestige—the two dimensions that explain the most survey responses to prestige score—is negatively associated with female typing at a statistically significant level ($p < 0.05$) for all decades. The largest negative association between female typing and both potency and general prestige is observed in the earliest decade of analysis (1900), with a generally declining potency and prestige penalty over time (but still statistically significant). Second, the first-order association between female typing and evaluation or activity/liveliness is more ambiguous. In only 4 of the 11 decades, I find four statistically significant association between an occupation's semantic association with women and levels of perceived liveliness; yet the direct of such association differs. The same pattern applies to the association of evaluation as well: the negative association between evaluation and female typing appears in 1900 but never in the following decades. Only half of the following decades show a statistically significant positive association between (moral) evaluation and female typing.

To examine the cross-sectional association between female typing and multiple dimensions of text-embedded occupational prestige, I estimate two models. First, Figure S12 presents the coefficient on female typing from a simple OLS model that includes only the first-order term, estimated separately by decade (using COCHA and Ngram pooled data with corpus fixed effects; results are similar when each corpus is analyzed separately). Two key findings emerge. First, female typing is negatively associated with both perceived potency and general prestige—the two dimensions most strongly aligned with survey-based prestige scores—with statistically significant coefficients ($p < 0.05$) in every decade. The magnitude of this negative association is largest in the earliest decade (1900), with a generally declining but still significant penalty in subsequent decades. Second, the first-order associations between female typing and either evaluation or activity/liveliness are more mixed. For liveliness, only 4 out of the 11 decades show statistically significant coefficients, and the direction of association varies across time. A similar pattern applies to moral evaluation: a negative association is observed in 1900, but this pattern does not persist. In the decades that follow, roughly half show a statistically significant *positive* association between evaluation and female typing.

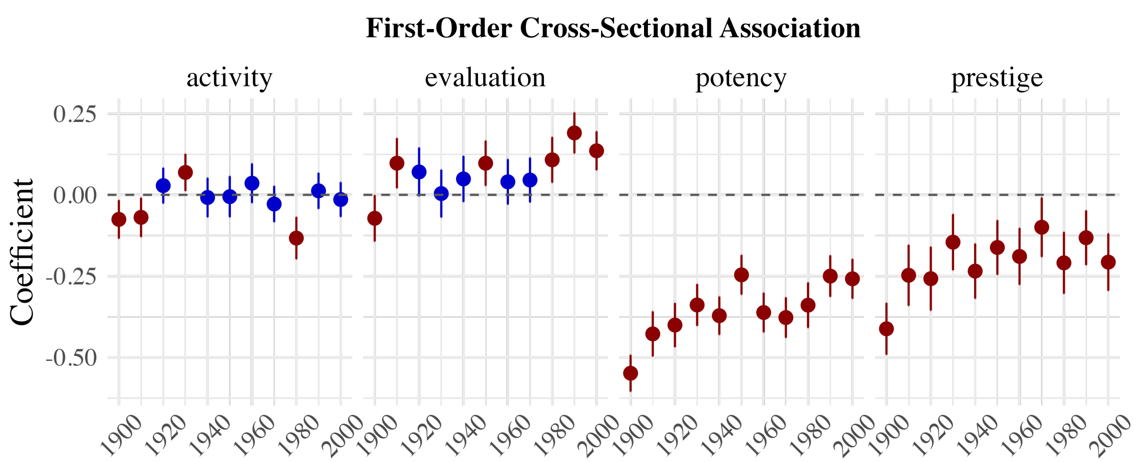


Figure S12: The cross-sectional association between female typing and dimensions of prestige with a first-order term of female typing, Ngram and COCHA pooled, 1900-2009

The second model, summarized in Figure S13, examines both the first-order and quadratic coefficients of female typing on occupational prestige in each decade, using an OLS specification that includes both terms (COCHA and Ngram pooled, with corpus fixed effects). I focus primarily on the coefficient and statistical significance of the quadratic term, using the first-order term to assess whether any curvilinear pattern emerges *within* the observed support of female typing. I highlight two substantive findings. First, there is virtually no evidence of curvilinear relationships for potency or general prestige. Combined with the results from the simpler model above, this suggests that these two dimensions are negatively and monotonically associated with female typing across the full period examined.

Second, turning to moral evaluation, the results indicate a consistent U-shaped relationship across most decades: occupations seen as highly female-typed or highly male-typed are evaluated more favorably than gender-neutral ones, with the curvature typically turning around at a point very close to $x = 0$ (*i.e.*, occupational gender neutrality). This pattern aligns with theories of benevolent sexism, in which women are symbolically rewarded or venerated for engaging in stereotypically feminine roles—such as care work—or embodying culturally feminine traits like beauty and fragility (Glick and Fiske 1996). Depending on how much the evaluation dimension contributes to overall occupational prestige (Freeland and Hoey 2018; MacKinnon and Langford 1994), this finding may help explain why Valentino (2020) and Krueger et al. (2022) report curvilinear relationships between female share and survey-based prestige. By contrast, the activity (or liveliness) dimension shows no consistent curvilinear pattern. The estimated effects are close to zero for both the first-order and quadratic terms, and an F-test of joint significance supports this null finding: only 4 out of the 11 models yield a statistically significant F-statistic.

The second model as summarized in Figure S13 examines both the first-order and quadratic coefficient of female typing on prestige in each decade when both terms are included in the same OLS model (COCHA and Ngram similarly pooled with corpus fixed-effects). I focus on the coefficients and their statistical significance on the quadratic term, while using the first-order term to determine whether any curvilinearity appears *within* the support of female typing. I highlight two substantive findings. First, virtually no curvilinear relations are found on potency and general prestige; combined with the patterns above, results suggest that the two dimensions are significantly associated with female typing *monotonically*. Second, focusing on the quadratic term, results suggest that the extent to which an occupation is perceived as morally worthy and justified has a U-shaped relationship with female typing in most decades, with curvature turning around at about $x = 0$ (*i.e.*, gender typing being neutral); in other words, the more segregated the occupation is understood as, the higher it is morally evaluated in a given decade. This is consistent with the psychology of benevolent sexism, where women are venerated and symbolically rewarded for performing duties that are stereotypically feminine, such as household and care work, or reflecting feminine traits, such as beauty and fragility (Glick and Fiske 1996). Depending on how much the evaluation dimension contributes to the overall prestige of occupations (Freeland and Hoey 2018; MacKinnon and Langford 1994), the finding may partly explain why Valentino (2020) and Krueger et al. (2022) found a curvilinear cross-sectional relation between occupational female share and survey-based prestige score. The activity dimension, on the other hand, show no consistent results of its curvilinearity; overall, the effect is close to null for both terms (first-order and quadratic) when an F-test is conducted; only 4 of the 11 models show a statistically significant F score (results omitted).

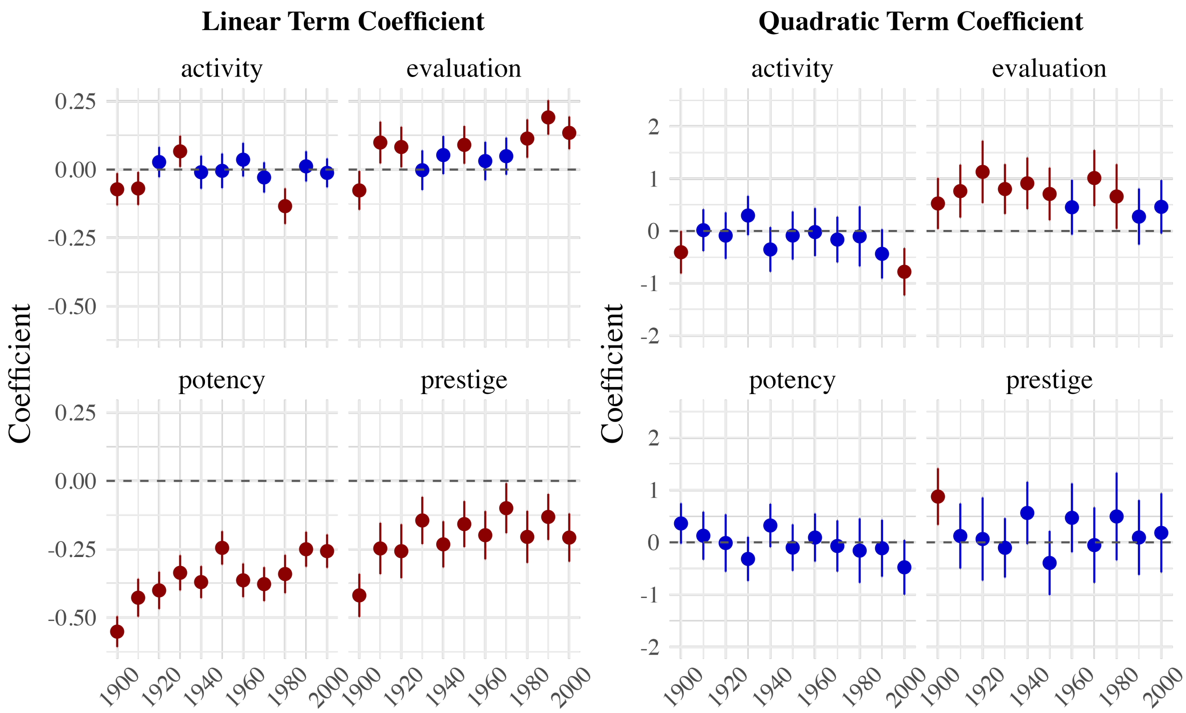


Figure S13: The cross-sectional association between female typing and dimensions of prestige with both first-order and quadratic term of female typing, Ngram and COCHA pooled, 1900-2009

E.4 A Test of Curvilinearity in the Fixed-Effects Panel Model

To examine whether a curvilinear (U-shaped) relationship exists between female typing and any dimension of symbolic value over time, I use the "double-demeaned" estimator recommended by Giesselmann and Schmidt-Catran (2022), which provides an unbiased estimate of the quadratic term within a TWFE panel data framework, assuming the data-generating process is correctly specified. Tables S6 and S7 report the estimated coefficients on the quadratic term of female typing using the COCHA and Ngram corpora, respectively. Across both datasets, I find no evidence of curvilinearity: none of the quadratic terms are statistically significant in predicting any dimension of occupational symbolic value.

Two explanations may account for this null finding. First, the curvilinear relationships observed in cross-sectional analyses may be spurious, driven by unobserved, time-invariant characteristics that are differenced out in the fixed-effects framework. Second, and perhaps more plausibly, the fixed-effects model estimates curvilinearity only from within-occupation variation over time, which is substantially more limited than the full range of gender typing used in cross-sectional models. If most within-occupation changes in gender typing occur predominantly in one direction (either increasing or decreasing), the opposing sides of a U-shaped curve may not be jointly observed, and any curvilinear pattern could be obscured or canceled out in the longitudinal specification.

Taken together, one of the major concerns regarding curvilinearity—primarily driven by the cross-sectional association between female typing and moral evaluation, as discussed above, and potentially underlying some of the mixed findings in prior literature (Krueger et al. 2022; Valentino 2020)—does not appear to hold in longitudinal analysis using a fixed-effects estimator. Over time, increasing cultural association with women is not significantly associated with changes in an occupation's perceived moral evaluation, whether in a monotonic (as concluded in the main paper) or curvilinear form.

Table S6: TWFE Estimates of the Association Between Female Typing and Occupational Symbolic Value with Double-Demeaned Quadratic Terms of Female Typing (Giesselmann and Schmidt-Catran 2022), COCHA

	Dependent Variable			
Variables	Prestige	Potency	Evaluation	Activity
Female Typing (demeaned)	-0.175***	-0.226***	0.004	-0.041
	(0.027)	(0.023)	(0.022)	(0.029)
Female Typing Squared (double-demeaned)	0.481	-0.143	0.070	-0.091
	(0.415)	(0.340)	(0.305)	(0.393)
Original Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Occupation Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R^2 (Within)	0.199	0.208	0.224	0.106
Observations	2,496	2,496	2,496	2,496

Standard errors are clustered at occupation level.
*p<0.05; **p<0.01; ***p<0.001 (two-tailed test).

Table S7: TWFE Estimates of the Association Between Female Typing and Occupational Symbolic Value with Double-Demeaned Quadratic Terms of Female Typing (Giesselmann and Schmidt-Catran 2022), Ngram

	Dependent Variable			
Variables	Prestige	Potency	Evaluation	Activity
Female Typing (demeaned)	-0.091***	-0.171***	-0.011	-0.008
	(0.025)	(0.022)	(0.025)	(0.022)
Female Typing Squared (double-demeaned)	-0.720	-0.228	0.100	-0.985
	(0.545)	(0.495)	(0.459)	(0.498)
Original Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
Occupation Fixed Effects	Yes	Yes	Yes	Yes
Adjusted R^2 (Within)	0.136	0.193	0.114	0.183
Observations	2,545	2,545	2,545	2,545

Standard errors are clustered at occupation level.
$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$ (two-tailed test).

E.5 The Mediation Analysis between Gender Typing and Actual Hourly Wages

Figure S14 demonstrates the SEM model used for the mediation analysis. Specifically, I decompose the effect of occupational gender typing (G_{it}) on current or future actual hourly wages (W_{it}) into the direct effect, γ_{gw}, and indirect effect, $\beta_{gp}\gamma_{pw}$, that is mediated by gendered evaluation of occupational prestige (*i.e.*, the first principal component of the four prestige dimensions, P_{it}). Controls of skills (S_{1it} to S_{kit}) as explained in the main paper are allowed to covary with prestige and hourly wages. I report the proportion of the total effect, $\beta_{gp}\gamma_{pw} + \gamma_{gw}$ that is explained by the indirect effect, $\beta_{gp}\gamma_{pw}$ in Figure 7 in the main paper; that is, $\beta_{gp}\gamma_{pw}\left(\beta_{gp}\gamma_{pw} + \gamma_{gw}\right)^{-1}$.

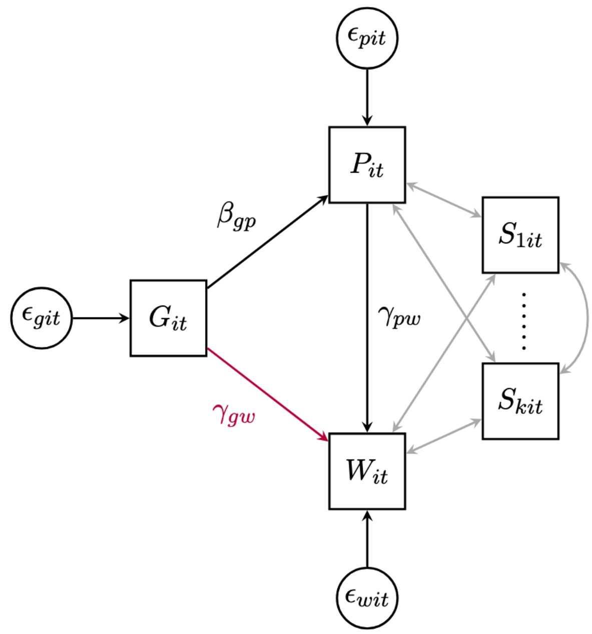


Figure S14: The Structural Equation Model (SEM) estimating the mediating effect of prestige change on the total association between gender typing and actual hourly wages

E.6 Robustness Checks

E.6.1 Potential Reversed Causality

One prominent threat to the cultural interpretation of the observed association between female typing and prestige is that the link is driven by a reversed process, where a drop in prestige (either in the dimension of general prestige or potency) is a precursor of increased female typing. Some prior works used a cross-lagged panel model with fixed-effects, which includes a lagged dependent variable to control for the aforementioned reversed causality in a Structural Equation Model (SEM) framework as represented by the Panel A of Figure S15 (Levanon et al. 2009). Specifically, when estimating the effect of gender typing on prestige, the current measure of prestige (*e.g.*, y_3) is a function of the latent fixed-effects term α, the current measure of female typing x_3, and the prestige in the last decade y_2, which may predict x_3 from the reciprocal effect of prestige on gender typing as well as y_3 due to prestige inertia (*i.e.*, y_2 is an important omitted variable). To accommodate the Nickell bias arising from controlling for both y_2 and fixed-effects (Nickell 1981), the disturbance term of y_2 (ϵ_2) is allowed to correlate with all future values of x (*i.e.*, x_3) (Moral-Benito et al. 2019). The SEM model would essentially generate the same estimate as the canonical Arellano-Bond model but with higher efficiency (Arellano and Bond 1991) if the data-generating process is specified correctly.

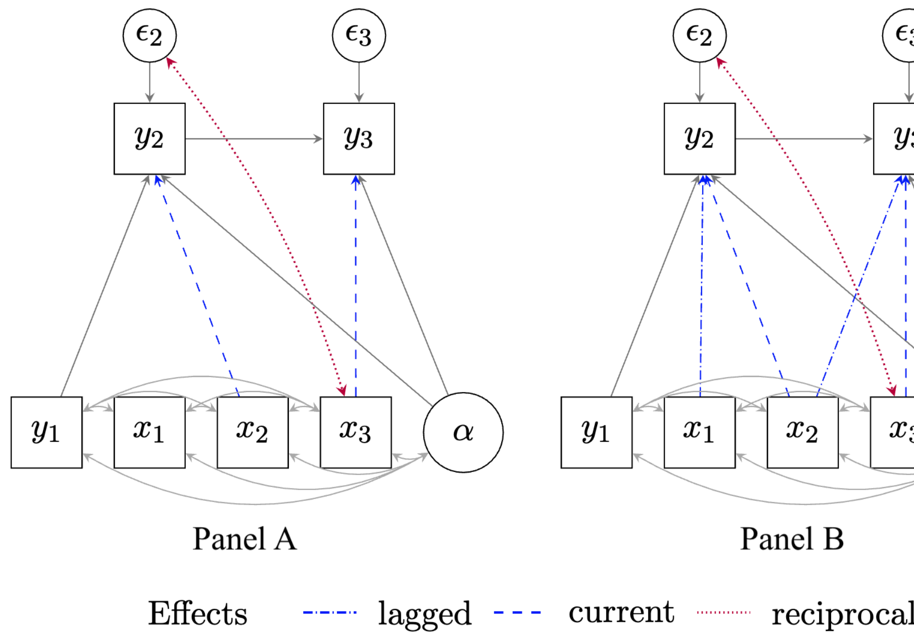


Figure S15: Effects of gender typing on occupational prestige with a reciprocal queuing process, 3-period panel data

Using the SEM model (Panel A of Figure S15), Figure S16 present the results showing the effect of "contemporaneous" female typing (or prestige) on prestige in each of its four dimensions (or female typing) in the devaluation (or queuing) model. I find a robust devaluation effect of female typing on general prestige and potency (p=0.05) even when the reversed reciprocal process is modeled; the queuing effect of general prestige on female typing, on the other hand, only appears in COCHA but not Ngram, and the magnitude of queuing effect as appeared in the potency dimension is generally smaller than the main devaluation effect.

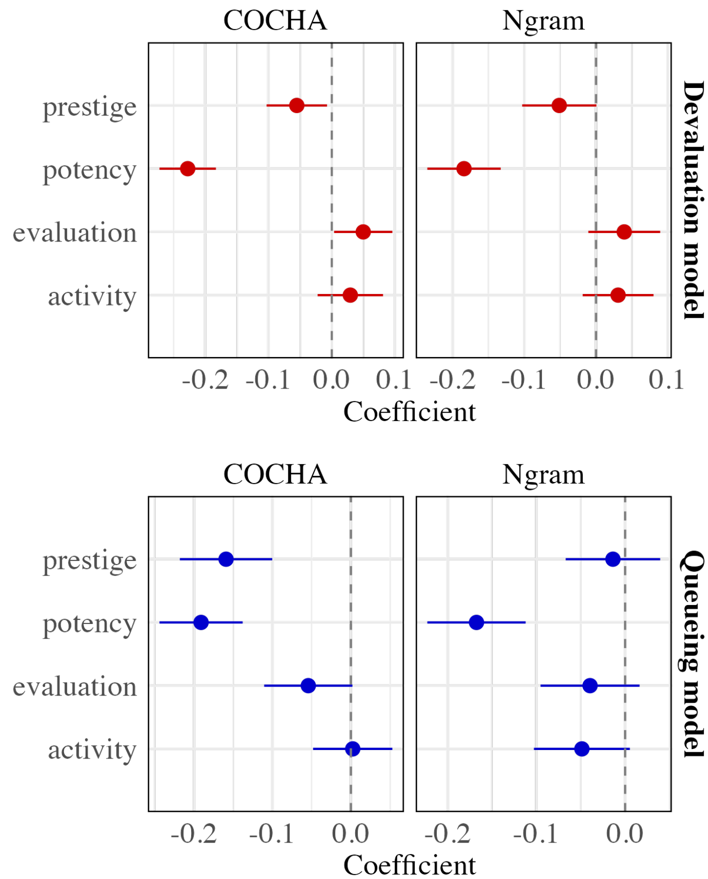


Figure S16: The devaluation and queueing effect of gender typing on different dimensions of prestige; no lagged independent variable is included

The cross-lagged panel model with fixed-effects as Moral-Benito et al. (2019) developed recently received a series of challenges in their potential estimate biases (and in many cases, with opposite signs to the actual data-generating process) when the key explanatory variable is not lagged correctly (Vaisey and Miles 2017). For example, the estimated queueing effect of prestige on gender typing above is likely to be biased, as the process of men leaving prestige-declining occupations and new gender typing being formed is never completely instantly (*i.e.*, I did not lag x in the above model). Notably, an unbiased estimate, as Vaisey and Miles (2017) pointed out, depends on a precise temporal lag during which men leave the occupation and new gender typing is formed. The actual time window for the process to be completed is, however, essentially unknown. To partly address the bias, Leszczensky and

Wolbring (2022) recommended an inclusion of both lagged and contemporaneous x when timing of the main effect is agnostic as in the queuing model (but not in the devaluation model as the cultural bias against female work appears as soon as gender typing is formed). Therefore, I estimate the queuing process alternatively using the model represented by the Panel B of Figure S15, where both lagged and contemporaneous terms of x are included. According to Figure S17 that presents the contemporaneous (upper panel) and lagged (lower panel) effects of the four dimensions of prestige on gender typing, I find some evidence for the queuing effect on the potency dimension (upper panel). These estimates, however, are likely to be biased, as the lagged terms suggest a *positive* effect of prestige increase on elevated female typing, which is not explained by any theories or previously found in any empirical studies. It arises largely because the explanatory variables are not lagged correctly (Vaisey and Miles 2017). It is because of such sensitivity of the results relative to the lagging window that I did not use these models in the main analysis.

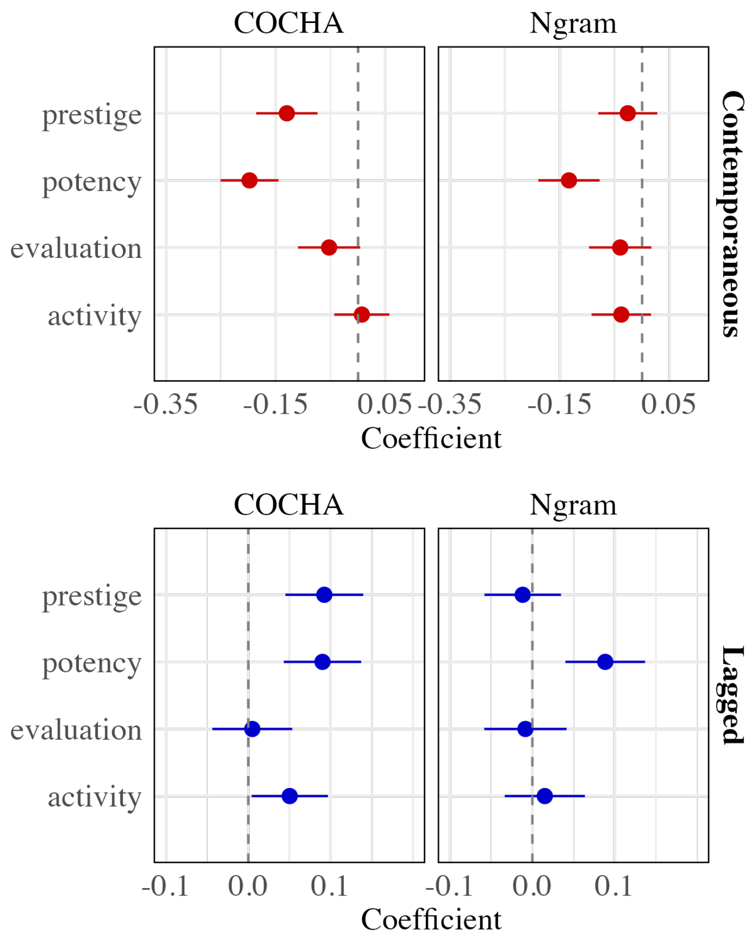


Figure S17: The queueing effect of gender typing on different dimensions of prestige; lagged independent variable is included

E.6.2 Additional Labor Market Controls

The observed link between female typing and occupational prestige may be driven by factors not being controlled in the main analysis. Some of the most important additional controls included in the robustness check are the measures of required skills and their changes over time for each occupation. To construct these measures, I use the Dictionary of Titles (DoT) in 1965 (the third edition), 1977 (fourth edition), and 1991 (revised fourth edition), where the required level of routine cognitive and human interactive skills, and other aptitude measures were reported *under the same scale* over time. The third edition published in 1965 has rarely been used in prior studies due to limited access to the high-quality digital 1965 DoT file, but is particularly important for this study that relies on temporal changes; I address this hurdle by using a newly developed transcription of the 1965 document (Althobaiti et al. 2022).

I include most skill measures that consistently appear in the three editions. They include all *aptitude* measures as appeared in the original 1965 documentation (see Figure S18), including required levels of intelligence, verbal, numerical, spatial, form perception, clerical perception, motor coordination, finger dexterity, manual dexterity, eye-hand-foot coordination, and color discrimination skills. I also include two *temperament* measures, *i.e.*, routine cognitive skills that measure the adaptability to situation requiring the precise attainment of set limits, tolerances or standards, and human interactive skills that measure the adaptability to dealing with people beyond giving and receiving instructions. Last, I include the mean value of General Educational Development (GED) measures of reasoning, mathematical, and language capacities. As skill requirements typically change slowly, I impute the 1980 measure for each occupation using linear interpolation to form a continuous skill measure for four decades.

II. APTITUDES

Specific capacities and abilities required of an individual in order to learn or perform adequately a task or job duty.

- *G* **INTELLIGENCE: General learning ability. The ability to "catch on" or understand instructions and underlying principles. Ability to reason and make judgments. Closely related to doing well in school.** ha!
- *V* **VERBAL: Ability to understand meanings of words and ideas associated with them, and to use them effectively. To comprehend language, to understand relationships between words, and to understand meanings of whole sentences and paragraphs. To present information or ideas clearly.**
- *N* **NUMERICAL: Ability to perform arithmetic operations quickly and accurately.**
- *S* **SPATIAL: Ability to comprehend forms in space and understand relationships of plane and solid objects. May be used in such tasks as blueprint reading and in solving geometry problems. Frequently described as the ability to "visualize" objects of two or three dimensions, or to think visually of geometric forms.**
- *P* **FORM PERCEPTION: Ability to perceive pertinent detail in objects or in pictorial or graphic material; To make visual comparisons and discriminations and see slight differences in shapes and shadings of figures and widths and lengths of lines.**
- *Q* **CLERICAL PERCEPTION: Ability to perceive pertinent detail in verbal or tabular material. To observe differences in copy, to proofread words and numbers, and to avoid perceptual errors in arithmetic computation.**
- *K* **MOTOR COORDINATION: Ability to coordinate eyes and hands or fingers rapidly and accurately in making precise movements with speed. Ability to make a movement response accurately and quickly.**
- *F* **FINGER DEXTERITY: Ability to move the fingers and manipulate small objects with the fingers rapidly or accurately.**
- *M* **MANAUL DEXTERITY: Ability to move the hands easily and skillfully. To work with the hands in placing and turning motions.**
- *E* **EYE-HAND-FOOT COORDINATION: Ability to move the hand and foot coordinately with each other in accordance with visual stimuli.**
- *C* **COLOR DISCRIMINATION: Ability to perceive or recognize similarities or differences in colors, or in shades or other values of the same color; to identify a particular color, or to recognize harmonious or contrasting color combinations, or to match colors accurately.**

Figure S18: Aptitude measures, 1965 DoT file

Occupations were initially coded under their respective year's scheme without being harmonized to standard occupation codes. To generate COC1950-based occupation codes that can be merged with the main dataset, the DoT occupation codes (1977 and 1991 editions) are first mapped to the COC1960 system based on the April 1971 Current Population Survey (CPS) Monthly File (with corresponding CPS sample weights) issued by the National Academy of Sciences (2006), where experts assigned each DoT occupation code in their respective years to COC1960-based occupation(s). I then map each COC1960-based occupation to COC1950 using the IPUMS official crosswalk. For future scholars to use these harmonized measures which are (surprisingly) not systematically compiled in public databases, I make the data publicly available through `link`.

E.6.3 Occupation Category-Specific Prestige Trajectory

While the main analysis leverages the parallel trend assumption in TWFE, in this section, I relax the assumption and allow the prestige trajectories (*i.e.*, slopes) to vary by occupation groups (*e.g.*, management versus service class; categories are defined by the Census based on the socio-economic functions of occupations) by interacting years with group-specific fixed effects (Ludwig and Brüderl 2018) (fixed-effects with individual slopes, or FEIS). Substantively, it allows cultural feminization to be partly "selected" based on the growth/decline of prestige; this may be plausible, for example, when well-educated women excessively enter professional or managerial occupations expecting a higher wage (and prestige) growth potential in the coming decades (Harris 2022), or occupations with more flexible schedules but with a lower wage (and prestige) growth potential (Blau and Kahn 2017; Goldin 2014). Indeed, recent empirical studies have shown non-trivial consequences of relaxing the parallel trend in the studies of wage changes relative to family and gender dynamics (Ludwig and Brüderl 2018). Controlling for group-specific slopes, however, does not change this paper's results: according to Figure S19, no substantial differences to the main finding are found; a cultural association between female typing and lower general prestige and potency, but not moral standing or liveliness still persists.

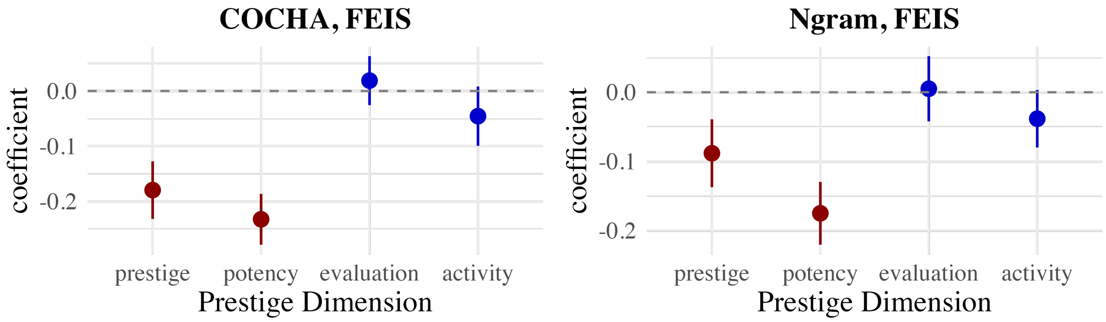


Figure S19: Effects of female typing on occupational prestige, two-way fixed-effects with occupation category-specific slope

E.6.4 Devaluation Effect for Male and Female Incumbents of the Occupation

In studying the wage effect of actual occupational feminization, scholars mostly differentiate male and female wages rather than using the overall (median) wage as the key dependent variable (Harris 2022). One reason is that female workers on average earn less than men do, and an increasing proportion of women in replacement of men typically drives down the overall wages, but not necessarily gender-specific wages. In parallel, female words and the associated gender dimension may be linked with a lower prestige in the macro semantic space (or "mechanical correlation") (Kozlowski et al. 2019) (also see Appendix B.2), and the association between female typing and lower prestige may similarly emerge at the micro occupation level artificially.

While wages can be easily separated by gender, it is hard to differentiate the prestige of the same occupation by the incumbent's gender from the text. Such cases do exist, however, when people use gender-specific words to represent occupations, such as waiter vs. waitress, where incumbents of the occupation are arguably men or women exclusively. In Table S8, I report the effect of the changes in female typing of gender-neutralized titles (*e.g.*, server) on the prestige of gender-specific titles of the *same* occupation (waiter and waitress).[3] Neg-

[3] The gender-specific occupations I use in analysis include: actor (male) vs. actress (female) from entertainer (neutral); charman vs. charwoman from cleaner; steward vs. stewardess from attendant; launderer

ative associations between female typing and general prestige and potency are consistently reported across corpora and appear on both male and female incumbents of the feminized occupations, with mostly comparable magnitude with the main regression. Importantly, while the test focuses mainly on the coefficient direction and size rather than statistical significance (as the sample size is smaller than 10 for these gender-specific occupation titles), most devaluation effects appears to be statistically significant at $p = 0.05$. Taken together, the results suggest that an occupation's movements towards femininity would affect the perceived prestige and potency of both men and women incumbents of the occupation.

Table S8: TWFE estimates of the effect of female typing on occupational general prestige and potency of male and female incumbents of the occupation; the original main results are appended (*i.e.*, gender-neutral occupation titles with both male and female incumbents) for references

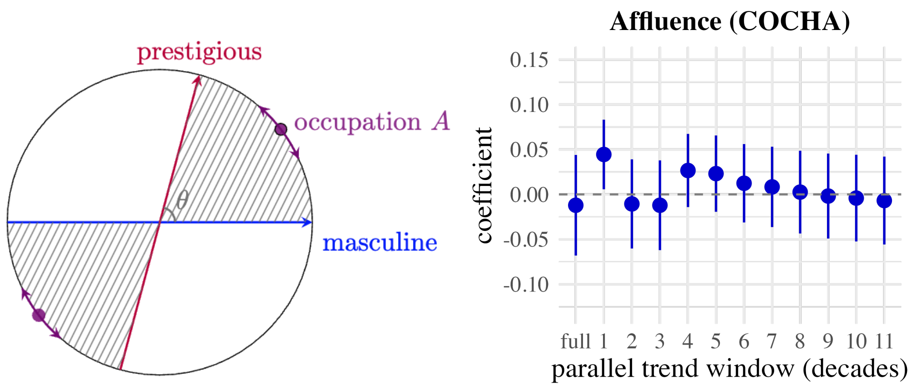
Corpus	Incumbents	Coefficient	s.e.	p	
Ngram	both	-0.091***	0.025	0.000	General Prestige
Ngram	male	-0.060	0.113	0.616	
Ngram	female	-0.044	0.215	0.843	
COCHA	both	-0.174***	0.027	0.000	
COCHA	male	-0.628***	0.094	0.000	
COCHA	female	-0.069**	0.015	0.009	
Ngram	both	-0.171***	0.022	0.000	Potency
Ngram	male	-0.136	0.113	0.266	
Ngram	female	-0.231*	0.094	0.049	
COCHA	both	-0.223***	0.024	0.000	
COCHA	male	-0.404**	0.088	0.002	
COCHA	female	-0.268*	0.076	0.024	

Note: controls of affluence and cultivation are included in all models.
*p<0.05; **p<0.01; ***p<0.001 (two-tailed test).

Indeed, the "mechanical correlation" does not naturally appear at the occupation level by construction. For example, Panel A of Figure S20 that describes the two cultural dimensions

vs. laundress from washer/cleaner; fireman vs. firewoman from firefighter; mailman vs. mailwoman from mailperson/messenger; waiter vs. waitress from server; tailor vs. tailoress/seamstress from tailor; policeman vs. policewoman from police; salesman vs. saleswoman from sales. In some cases, only one gender-specific titles in the same pair appear. For example, firewoman never appears in the three corpora, while fireman appears consistently.

and occupations' relative position associated with them in a 2-D space shows that, at least in the shaded area, any movement that increases a occupation's cultural association with men (women) would lead to a decrease (increase) in its semantic affinity with prestige. The dynamics at the occupation level depend essentially on the occupation's relative positions with each cultural dimension. This may also be seen in the case of the affluence dimension: the female dimension is mechanically positively correlated with the affluence dimension (see discussions in Appendix B.2 and Kozlowski et al. (2019)); yet an occupation's movement towards femininity is associated with a decline in its perceived affluence longitudinally (see the TWFE estimation labeled as "full" in Panel B).



Panel A

Panel B

Figure S20: Movement of occupation A along the two dimensions that are mechanically associated (Panel A) and the effect of female typing on affluence in COCHA (Panel B)

References

Althobaiti, Shahad, Ahmad Alabdulkareem, Judy Hanwen Shen, Iyad Rahwan, Morgan Frank, Esteban Moro, and Alex Rutherford. 2022. "Dictionary of Titles."

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and An Application to Employment Equations." *The Review of Economic Studies* 58(2):277–297.

Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3):789–865.

Bose, Christine E., and Peter H. Rossi. 1983. "Gender and Jobs: Prestige Standings of Occupations as Affected by Gender." *American Sociological Review* 48(3):316–330.

Crawley, Donna. 2014. "Gender and Perceptions of Occupational Prestige: Changes Over 20 Years." *Sage Open* 4(1):2158244013518923.

England, Paula. 1979. "Women and Occupational Prestige: A Case of Vacuous Sex Equality." *Signs: Journal of Women in Culture and Society* 5(2):252–265.

Freeland, Robert E, and Jesse Hoey. 2018. "The Structure of Deference: Modeling Occupational Status Using Affect Control Theory." *American Sociological Review* 83(2):243–277.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes." *Proceedings of the National Academy of Sciences* 115(16):E3635–E3644.

Giesselmann, Marco, and Alexander W Schmidt-Catran. 2022. "Interactions in fixed effects regression models." *Sociological Methods & Research* 51(3):1100–1127.

Glick, Peter, and Susan T. Fiske. 1996. "The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism." *Journal of Personality and Social Psychology* 70(3):491.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104(4):1091–1119.

Gordon, Jonathan, and Benjamin Van Durme. 2013. "Reporting Bias and Knowledge Acquisition." Pp. 25–30 in *Proceedings of the 2013 Workshop on Automated Knowledge Base Construction*. New York, NY: Association for Computing Machinery.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. "Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change." Pp. 2116 in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, volume 2016. Austin, Texas: Association for Computational Linguistics.

Harris, Jorgen. 2022. "Do Wages Fall When Women Enter an Occupation?" *Labour Economics* 74:102102.

Kmetty, Zoltán, Júlia Koltai, and Tamás Rudas. 2021. "The Presence of Occupational Structure in Online Texts Based on Word Embedding NLP Models." *EPJ Data Science* 10(1):55.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class Through Word Embeddings." *American Sociological Review* 84(5):905–949.

Krueger, Sabine, Christian Ebner, and Daniela Rohrbach-Schmidt. 2022. "Gender Composition and the Symbolic Value of Occupations: New Evidence of a U-shaped Relationship between Gender and Occupational Prestige Based on German Microdata." *Work, Employment and Society* 0(0):1–20.

Leszczensky, Lars, and Tobias Wolbring. 2022. "How to Deal with Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study." *Sociological Methods & Research* 51(2):837–865.

Levanon, Asaf, Paula England, and Paul Allison. 2009. "Occupational Feminization and Pay: Assessing Causal Dynamics Using 1950–2000 US Census Data." *Social Forces* 88(2):865–891.

Ludwig, Volker, and Josef Brüderl. 2018. "Is There a Male Marital Wage Premium? New Evidence From the United States." *American Sociological Review* 83(4):744–770.

MacKinnon, Neil J., and Tom Langford. 1994. "The Meaning of Occupational Prestige Scores: A Social Psychological Analysis and Interpretation." *The Sociological Quarterly* 35(2):215–245.

Mandel, Hadas. 2018. "A Second Look at the Process of Occupational Feminization and Pay Reduction in Occupations." *Demography* 55(2):669–690.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. "Distributed representations of words and phrases and their compositionality." *Advances in Neural Information Processing Systems* 26.

Moral-Benito, Enrique, Paul Allison, and Richard Williams. 2019. "Dynamic Panel Data Modelling Using Maximum Likelihood: An Alternative to Arellano-Bond." *Applied Economics* 51(20):2221–2232.

Nakao, Keiko, and Judith K. Treas. 1994. "Updating Occupational Prestige and Socioeconomic Scores: How the New Measures Measure up." *Sociological Methodology* 24:1–72.

National Academy of Sciences. 2006. "Dictionary of Occupational Titles (DOT): Part I - Current Population Survey, April 1971, Augmented With DOT Characteristics and Dictionary of Occupational Titles (DOT): Part II - Fourth Edition Dictionary of DOT Scores for 1970 Census Categories."

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica: Journal of the Econometric Society* 49(6):1417–1426.

Pechenick, Eitan Adam, Christopher M. Danforth, and Peter Sheridan Dodds. 2015. "Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-cultural and Linguistic Evolution." *PloS One* 10(10):e0137041.

Roget, Mark P. 1911. *Roget's Thesaurus of English Words and Phrases*. New York, NY: TY Crowell Company.

Shi, Kejian, Leyi Yan, and Chuwei Xu. 2021. "Social Bias in Masked-LMs Pretrained on Scientific Corpora." The Workshop on Natural Language Processing for Scientific Texts.

Siegel, Paul M. 1971. "Prestige in the American Occupational Structure." PhD dissertation, Department of Sociology, The University of Chicago, Chicago, IL.

Smith, Charles J. 1910. *Synonyms Discriminated: A Dictionary of Synonymous Words in the English Language*. New York, NY: Henry Holt.

Treiman, Donald J. 1977. *Occupational Prestige in Comparative Perspective*. Quantitative Studies in Social Relations. New York, NY: Academic Press.

Treiman, Donald J., and Kermit Terrell. 1975. "Sex and the Process of Status Attainment: A Comparison of Working Women and Men." *American Sociological Review* 40(2):174–200.

Vaisey, Stephen, and Andrew Miles. 2017. "What You Can—and Can't—Do with Three-Wave Panel Data." *Sociological Methods & Research* 46(1):44–67.

Valentino, Lauren. 2020. "The Segregation Premium: How Gender Shapes the Symbolic Valuation Process of Occupational Prestige Judgments." *Social Forces* 99(1):31–58.

Van Loon, Austin, and Jeremy Freese. 2023. "Word Embeddings Reveal How Fundamental Sentiments Structure Natural Language." *American Behavioral Scientist* 67(2):175–200.