AMERICAN SOCIOLOGICAL ASSOCIATION

American Sociological Review
2019, Vol. 84(5) 905–949

DOI: 10.1177/0003122419877135
journals.sagepub.com/home/asr

SAGE

The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings

Austin C. Kozlowski,[a] Matt Taddy,[b] and James A. Evans[a,c]

Abstract

We argue word embedding models are a useful tool for the study of culture using a historical analysis of shared understandings of social class as an empirical case. Word embeddings represent semantic relations between words as relationships between vectors in a high-dimensional space, specifying a relational model of meaning consistent with contemporary theories of culture. Dimensions induced by word differences (*rich* – *poor*) in these spaces correspond to dimensions of cultural meaning, and the projection of words onto these dimensions reflects widely shared associations, which we validate with surveys. Analyzing text from millions of books published over 100 years, we show that the markers of class continuously shifted amidst the economic transformations of the twentieth century, yet the basic cultural dimensions of class remained remarkably stable. The notable exception is education, which became tightly linked to affluence independent of its association with cultivated taste.

Keywords

word embeddings, *word2vec*, culture, computational sociology, methodology, text analysis, content analysis

People classify objects along myriad axes of meaning to interpret the social world. In addition to core social categories such as gender and race, a diverse array of cultural dimensions—fair/unfair, beautiful/ugly, new/old—play key roles in patterning interactions and structuring institutions. Although dominant theories of culture posit such a multidimensional matrix of meanings, empirical investigations commonly limit their attention to one or two facets due to the analytic and methodological difficulties associated with incorporating higher dimensionality.

Social class, a central sociological construct, is itself a complex and multidimensional attribution. Stratification scholars commonly treat class as a composite of several distinct factors, including affluence, education, and occupation, as well as status and cultivated taste. Decades of social science research has produced extensive knowledge of how these various socioeconomic dimensions are materially and causally interrelated (Chan and

[a]University of Chicago
[b]Amazon
[c]Santa Fe Institute

Corresponding Author:
James A. Evans, Department of Sociology, University of Chicago, 1126 E. 59th Street, Chicago IL, 60637
Email: jevans@uchicago.edu

Goldthorpe 2007; DiMaggio 1982; Hout 2012). Yet the multiple dimensions of class are not only attributes used by analysts to articulate an individual's economic standing; they also serve as axes of cultural distinction actors deploy in daily life. People, groups, and everyday objects carry cultural associations of affluence, education, cultivation, and status, which together comprise profiles of classed meaning (Bourdieu 1984; Warner, Meeker, and Eells 1949). These profiles are evoked when individuals make decisions regarding what to purchase, where to spend the evening, how to present themselves, and who to befriend. Stratification scholars have developed strong conceptions of the material relations between the multiple dimensions of class, but understanding how the *meanings* of these dimensions relate to one another and co-evolve over time remains underspecified.

In this article, we apply an emerging computational approach—neural-network word embedding models—to analyze the cultural dimensions of social class and their evolution over the twentieth century. Word embedding algorithms input large collections of digitized text and output a high-dimensional vector-space model[1] in which each unique word is represented as a vector in the space (Mikolov, Yih, and Zweig 2013; Pennington, Socher, and Manning 2014). This means each word appearing in the analyzed documents is ascribed a set of coordinates that fix its location in a geometric space in relation to every other word. Words are positioned in this space based on their surrounding "context" words in the text, such that words sharing many contexts are positioned near one another, and words that inhabit different linguistic contexts are located farther apart. Previous work with word embeddings in computational linguistics shows that words frequently sharing contexts, and thus located nearby in the vector space, tend to share similar meanings.

We provide new evidence that the dimensions of word embedding vector space models closely correspond to meaningful "cultural dimensions," such as *rich-poor, moral-immoral*, and *masculine-feminine*. We show that a word vector's position on these dimensions reflects the word's respective cultural associations. For example, projecting occupation names on an "affluence dimension," we find that traditionally well-compensated occupations, such as banker and lawyer, are positioned at one end of the dimension, and poorly paid occupations, such as nanny and carpenter, lie at the other. This occurs because with each discursive context that "banker" shares with wealthy words like "affluent," "moneyed," and "rich," it is nudged toward the rich pole of the affluence dimension, and each time "nanny" shares a context with terms like "needy," "destitute," and "poor," it is nudged toward the poor pole.

After empirically validating word embeddings' ability to capture widely shared cultural associations, we apply this method to the question of how collective understandings of social class evolved in the United States over the course of the twentieth century. To gain new leverage on this question, we train word embedding models on text from millions of books published over the entire twentieth century digitized in the Google Ngram corpus. We then identify dimensions in these models corresponding to five cultural dimensions of class described by classical and contemporary sociological theory as well as two other cultural dimensions frequently invoked in association with class: affluence, employment, status, education, cultivation, morality, and gender.[2]

Comparing texts from each decade of the twentieth century, we discover that the cultural dimensions of class comprise a complex yet remarkably stable semantic structure. We find that affluence and status serve as cultural mediators between a cluster of education, cultivation, and morality on one hand and associations of employment and ownership on the other. This persistent and intransitive structure requires high dimensionality to represent without distortion. Furthermore, we find that the cultural markers signifying positions within this robust structure are in continual flux, with terms distinguishing high and low class shifting over the decades,

following steady patterns of cultural circulation and turnover.

MULTIDIMENSIONALITY OF CLASS

Social class, the systematic and hierarchical distinction between persons and groups in social standing, has long been recognized to operate along multiple distinct dimensions. Affluence is often treated as a core aspect of class, with income commonly serving as a proxy for socioeconomic status. This is not an arbitrary selection; money is quickly and easily convertible into many forms of capital, power, and influence, making it a particularly salient element of class (Simmel [1900] 2004).

Nevertheless, scholars long have argued that the economics of class cannot be reduced to affluence alone. Analysts in the Marxist tradition foreground socio-structural position and relation to capital as the basis of social class instead (Gramsci 1992; Marx [1867] 2004; Wright 1979). From this perspective, it is not the accumulation of wealth, but rather one's position as an owner or worker, that determines a shared interest with respect to politics, culture, and social life (Marx and Engels 1970). In addition to occupational position and wealth, social scientists frequently include education as a third element of socioeconomic status. Education became particularly central to the study of class after World War II, when the expansion of mass schooling and the demands of a changing labor market turned education into a critical axis of social division (Fischer and Hout 2006).

Theorists have also noted that a full conception of social class requires accounting for its symbolic manifestations. In an early articulation of this distinction, Weber (1978) contrasted economic class with status (*Stand*), which operates via social honor and prestige. Because status refers to actors' ability to make a credible claim of esteem rather than their power in a market, it need not always coincide with affluence (Chan and Goldthorpe 2007). Recent research confirms the empirical relevance of this theoretical distinction, finding that status shapes associational networks independently of economic factors, and individuals commonly distinguish prestige from earnings in their subjective evaluations of occupational social standing (Chan and Goldthorpe 2004; Freeland and Hoey 2018).

Another line of research establishes how cultivated tastes serve as a crucial marker of class distinct from individual or collective status. Veblen ([1899] 1912) and Elias (1978) articulated this connection between cultivation and class early in the twentieth century, and Bourdieu (1984) recentered this association at century's end with the concept of cultural capital. Numerous studies show how actors parlay cultural capital into economic gains (DiMaggio and Mohr 1985), but Bourdieu's (1984) original conception draws a more complex connection between cultivated taste and affluence, with cultural elites such as artists and intellectuals comprising their own high-status social groups that stand in opposition to the economic elite.

The cultural associations of class are entwined with many diverse dimensions of social classification. For example, a growing literature on valuation and moralized markets outlines how socioeconomic attributions are shaped by moral classifications (Fourcade and Healy 2007; Zelizer 1979). This scholarship details how moral distinctions become mapped onto socioeconomic positions (Svallfors 2006) and how moral sentiments shape economic valuation (Fourcade 2011). In this vein, Lamont (1992, 2000) illustrates how middle- and working-class Americans deploy moral and socioeconomic distinctions in tandem when forming judgments about their neighbors, their friends, and themselves. Classed associations similarly interact with understandings of gender. Feminist scholars have shown how gender permeates class in the labor process (Hochschild 2012; Salzinger 2003), consumption patterns (Cohen 2003; Illouz 1997; Mears 2010), and the macro system of economic stratification (Cha and Weeden 2014; Gilman 1999; Ridgeway 2011). Arising from historical processes that

differentially distribute power and prestige by gender, classed meanings are frequently also gendered meanings (Veblen [1899] 1912).

Together, contemporary and classical work paint class as a complex construct with many facets at once connected yet analytically and culturally distinct. The precise ways these cultural dimensions of class relate to one another, however, and how these interrelations have evolved over time, remain open empirical questions.

Social Class in the Twentieth Century

The twentieth century was a period of dramatic class transformation in the United States and beyond. Large organizations came to dominate the Western world's industrial and economic landscape, mass education heightened the importance of formal credentials for occupational attainment, and the gender composition of the workforce shifted radically as women entered historically male jobs and the incidence of divorce spiked (Collins 1979; Fischer and Hout 2006). Nevertheless, it is unclear whether the system of class-based meanings used by lay actors underwent parallel transformations. Despite voluminous scholarship focused on how shared understandings of class operate on the micro-level in particular times and places (e.g., Bourgois 2003; Khan 2010; Willis 1977), macro-historical analyses of the dimensions of meaning undergirding class remain rare.

Commentators offer competing narratives about the cultural trajectory of class in the twentieth century. Some characterize the twentieth century as the eclipse of social-structural positions by identities and lifestyles. According to this line of inquiry, noneconomic identifiers, such as gender, race, education, and consumption patterns, form the new backbone of political organization and group solidarity (Clark 2018; Hunter 1992; Pakulski and Waters 1996). This literature coincides with the popularization of cultural capital in anglophone sociology, which stresses the rising importance of symbolic attributions in determining class (DiMaggio and Mohr 1985).

Other scholars argue against the "death of class" narrative, claiming that occupation and position in class structure continue to play key roles in determining wealth and shaping collective identity (Weeden and Grusky 2005; Wright 2000). Yet most research on the durable importance of occupational position and control of capital focuses on their relations to observable life chances and is not directly concerned with the cultural matrix of class. It remains unclear whether sociology's increasing attention to identity and lifestyle in transforming social class reflects concurrent trends in how class is understood in public discourse.

A third possibility is that symbolic factors like cultivation and status have always been central to how class is collectively understood. For instance, Accominotti, Kahn, and Storer's (2018) analysis of New York Philharmonic attendance recounts how cultivated taste developed into a currency of cultural capital among a middle-class intelligentsia in the nineteenth century. Moreover, classical accounts of status and cultivation suggest these symbolic components have been structuring class since at least the end of the Industrial Revolution (Elias 1978; Veblen [1899] 1912; Weber 1978).

These considerations suggest the possibility that collective understandings of class are founded on a durable system of meanings resilient to large-scale economic transformations. Empirical investigation into whether class's cultural components remained stable over the twentieth century has been stymied by methodological difficulties associated with macro-cultural analysis. Following a line of successful inquiry (Bearman and Stovel 2000; Franzosi 2004; Mohr, Wagner-Pacifici, and Breiger 2015), we propose formal text analysis as a promising avenue for recovering widely-shared understandings of class from historical populations no longer available for direct observation.

FORMAL TEXT ANALYSIS IN THE STUDY OF CULTURE

Cultural scholars from sociology, anthropology, and socio-linguistics have commonly

theorized that a group's language reflects its cultural system (Lévi-Strauss 1963; Whorf 1956). Following this insight, text has served as a key source of data for scholars investigating cultural categories and meaning structures. Text is particularly well-suited to historical-cultural analysis, as it is often the most semantically-rich record a group leaves behind. In sociology, analysis of text has historically been dominated by qualitative approaches, the two most common being interpretivist close-reading and systematic qualitative coding.

Interpretive text analysis, in which the researcher draws insights from a holistic deep reading, has produced great advances in sociological understandings of culture, but it suffers from clear limitations in reproducibility (Ricoeur 1981). Qualitative coding, in which the researcher selects a number of themes and systematically tracks their deployment in text (Glaser and Strauss 1967), can be more reproducible than a singular close reading, but it suffers from low inter-coder reliability when themes are complex or subtle. Because these dominant techniques are not easily replicable and rely on the analyst's intuition and finesse, the study of culture in sociology has largely remained a "virtuoso affair" (DiMaggio 1997). Furthermore, both interpretive text analysis and qualitative coding are limited by the pace of human reading, so neither are well suited for the analysis of very large corpora or entire socio-cultural domains.

Limitations of qualitative textual analysis have motivated scholars of culture in the social sciences and humanities to develop an array of formal and quantitative methods of text analysis (Evans and Aceves 2016). Two such methods that have gained popularity in recent years are semantic network analysis and topic modeling. Semantic networks are typically constructed by treating words as nodes in a network and textual co-occurrences as links (Carley 1994; Hoffman et al. 2017; Kaufer and Carley 1993; Lee and Martin 2015). Examining structural characteristics of a semantic network, such as central words or words that bridge semantic or cultural holes, can provide insight into the relationship between individual words and the overall conceptual structure undergirding a text (Corman et al. 2002; Pachucki and Breiger 2010; Vilhena et al. 2014).

Alternatively, topic modeling is a more recent approach that uses a well-formed probability model to enable inductive discovery of "topics" structuring a corpus, each learned as a sparse distribution over words that tend to co-occur in text (Blei, Ng, and Jordan 2003; Mohr and Bogdanov 2013). Topic modeling can detect polysemy by tracing words that exist in multiple topics, and heteroglossia, the multiple voices of a single text, by inducing the mixture of distinct topics across documents (Blei 2012; DiMaggio, Nag, and Blei 2013).

Both methods can generate important insights into the cultural system that produced a text, but there remain many sociologically important questions for which these methods are poorly suited. When corpora grow sufficiently large, standard semantic network analysis metrics fail to distinguish between concepts that are close or distant by considering topological information alone.[3] Topic modeling sorts words into a predetermined number of clusters, or topics, based on co-occurrence in text, and such discrete clusters do not capture continuous relationships between words.

As such, both networks and topic models are ill-suited for representing the multifarious associations and cultural valances that characterize all words in a corpus. Questions regarding how masculine or feminine, good or bad, high- or low-class a given object is within a cultural system remain difficult to answer using existing formal methods for text analysis. Furthermore, investigation into the relations between cultural dimensions, such as how closely a culture's rich/poor distinction relates to its masculine/feminine dimension, is beyond the scope of prior approaches.

WORD EMBEDDING MODELS AND COMPLEX SEMANTIC RELATIONSHIPS

Recent work in natural language processing has made great strides by representing relationships between words in a corpus not as

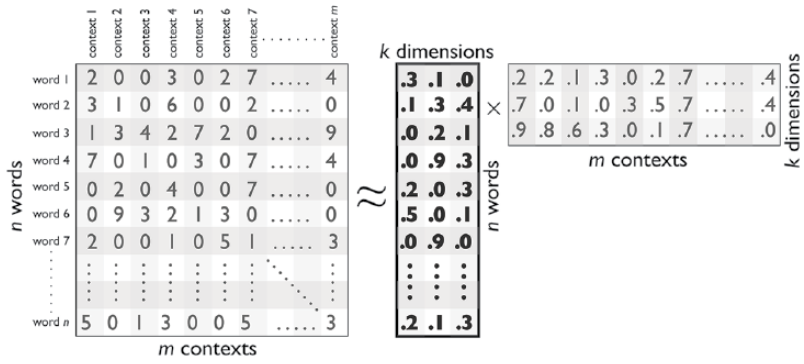

Figure 1. Schematic Illustration of the Descriptive Problem Neural Word Embeddings Solve—How to Represent All Words from a Corpus within a *k*-Dimensional Space That Best Preserves Distances between Words in Their Local Contexts

networks or topical clusters but as vectors in a dense, continuous, high-dimensional space (Joulin et al. 2016; Mikolov, Yih, et al. 2013; Pennington et al. 2014). These vector space models, known collectively as *word embeddings*, have attracted widespread interest among computer scientists and computational linguists due to their ability to capture and represent complex semantic relations.

In a word embedding model, each word is represented as a vector in shared vector space. Words sharing similar contexts within the text will be positioned nearby in the space, whereas words that appear only in distinct and disconnected contexts will be positioned farther apart. Figure 1 schematically illustrates the structure of the descriptive problem that word embeddings attempt to solve: how to represent all words from a corpus within the *k*-dimensional space that best preserves distances between *n* words across *m* semantic contexts. The solution, which we illustrate in subsequent figures, is an *n*-by-*k* matrix of values, where $k \ll m$, bolded here where $k = 3$.

An early approach to word embeddings, Latent Semantic Analysis (LSA), used singular-value decomposition (SVD) to factorize this word-context matrix when contexts were large—entire documents containing hundreds, thousands, or tens of thousands of words. The first singular value explained the most variation in the original *n*-by-*m* word-context matrix, the second component the second most, and so on, such that *k* was typically trimmed when the marginal *k*th singular value explained arbitrarily little variation in the matrix.

From efficiency considerations, SVD placed strict upper limits on the number of documents and lower limits on the size of semantic contexts they could factorize. Neural word embeddings use heuristic optimization of a neural network with at least one "hidden-layer" of *k* internal, dependent variables. This enables factorization of much larger word-context matrices constructed from vast numbers of documents containing many distinct words (large *n*) but very local word contexts (large *m*).[4]

In these models, $k \ll m$, but substantial natural language corpora require $k \geq 300$ to minimize the error of word-context matrix reconstruction (Mikolov, Yih, et al. 2013). Note that because the optimal distance between two vectors is a function of shared context rather than strict co-occurrence, words need not co-occur for their vectors to be positioned close together. If "doctor" and "lawyer" both appear near the word "work" or "office," then the vectors for "doctor" and "lawyer" would be located near each other in the embedding, even if they never appear together in text.

Distance between words in an embedding space is typically assessed using "cosine similarity," the cosine of the angle between two word vectors. This is preferred to the

Euclidean (straight-line) distance due to properties of high-dimensional spaces that violate intuitions formed in two or three dimensions. For example, as the dimensionality of a hypersphere grows, its volume shrinks relative to its surface area as more of that volume resides near the surface.[5] We normalize all word vectors (Levy, Goldberg, and Dagan 2015) such that they lie on the surface of a hypersphere of the same dimensionality as the space.

Word2vec, the most widely used word embedding algorithm and the primary approach we apply in the following analyses, uses a shallow, two-layered neural network architecture that optimizes the prediction of words based on shared context with other words.[6] Because words are located together in the embedding model if they appear in similar local contexts in the corpus, abutting words in the vector space tend to share similar meanings.

A word's nearest neighbors are often either its synonyms or syntactic variants. A word's broader neighborhood in the embedding space is typically populated by a host of terms with related meanings. Therefore, a great deal of semantic and cultural information is available simply by examining the word vectors that surround a word of interest. Kulkarni and colleagues (2015) have used word embedding models in this way to trace shifts in the meaning of the word "gay" over the course of the twentieth century, from a location in the vector space beside "cheerful" and "frolicsome" to one near "lesbian" and "bisexual." Hamilton, Leskovec, and Jurafsky (2016) similarly used word embedding models to investigate how a word's rate of semantic change, measured as change in the word's overall position in space, depends on its frequency and polysemy, finding that words occurring with high frequency change meaning more slowly and polysemous words change more rapidly.

Past work with word embedding models also shows that semantically meaningful relations can be found between words not directly proximate in the space. *Word2vec* initially attracted a great deal of attention by virtue of its intriguing ability to solve analogy problems by applying simple linear algebra to word vectors (Mikolov, Chen, et al. 2013). For example, the analogy "man is to woman as king is to _____" can be solved with a model trained on a large body of text by performing the arithmetic operation with the word vectors $\overrightarrow{king} - \overrightarrow{man} + \overrightarrow{woman}$, with the resulting vector most proximate to the word vector for $\overrightarrow{queen}$. *Word2vec* can achieve success rates as high as 74 percent (Ji et al. 2016) on a challenging analogy test comprising 20,000 questions involving semantic comparisons ranging from currency-country (*kwanza* is to Angola as *rial* is to Iran) and male-female (man is to woman as waiter is to waitress) to syntactic comparisons involving opposites, plural nouns, comparatives, superlatives, and verb conjugations (e.g., past tense, present participle) (Mikolov, Chen, et al. 2013). We provide a more detailed technical discussion of word embedding models in Appendix Part A.[7]

CULTURAL DIMENSIONS OF WORD EMBEDDINGS

In this article, we present a novel method for applying word embedding models to the sociological analysis of culture. We show that derived dimensions of word embedding vector spaces correspond closely to "cultural dimensions," such as affluence, gender, and status, which individuals use in everyday life to classify agents and objects in the world. By discovering and examining these culturally meaningful dimensions in a word embedding, analysts can reveal individual words' associations on those dimensions and determine how these dimensions are positioned relative to one another in that space.

For instance, an analyst can use a word embedding model to determine whether "opera" is considered more affluent than "jazz" by projecting the word vectors corresponding to "opera" and "jazz" onto the dimension of the space corresponding to affluence. Similarly, the researcher can determine if "jazz" is more masculine or feminine than "opera" by projecting these words onto

the dimension corresponding to gender in the same space. This dimensional approach emphasizes that semantic meaning is contained not only in the distance between two word vectors but also in the *direction* of that distance.

The technique we present for discovery of cultural dimensions in a word embedding vector space builds on logic for solving analogies with word embeddings. One interpretation for why $\overrightarrow{king} + \overrightarrow{woman} - \overrightarrow{man} \approx \overrightarrow{queen}$ in word embedding models is because $(\overrightarrow{woman} - \overrightarrow{man})$ closely corresponds to a "gender dimension." Adding $(\overrightarrow{woman} - \overrightarrow{man})$ to $\overrightarrow{king}$ has the effect of starting at $\overrightarrow{king}$ and taking one step on the gender dimension in the direction of femininity. Similarly, adding $(\overrightarrow{affluence} - \overrightarrow{poverty})$ to a word has the effect of taking one step in the direction of affluence. Following this intuition, we find with an embedding trained on contemporary Google News text that $\overrightarrow{hockey} + \overrightarrow{affluence} - \overrightarrow{poverty} \approx \overrightarrow{lacrosse}$. Conversely, $(\overrightarrow{poverty} - \overrightarrow{affluence})$ corresponds to one step in the direction of poverty on the same dimension.

An approximation of the affluence dimension is captured not only by $(\overrightarrow{affluence} - \overrightarrow{poverty})$, but also by any other pairs of words whose semantic difference corresponds to that cultural dimension of interest, such as $\overrightarrow{rich} - \overrightarrow{poor}$, $\overrightarrow{priceless} - \overrightarrow{worthless}$, or $\overrightarrow{prosperous} - \overrightarrow{bankrupt}$. Because we expect these similar word pairs to approximate the same cultural dimension of affluence, we calculate a single, robust affluence dimension by simply taking the arithmetic mean of a set of such pairs.[8] Other cultural dimensions, such as gender or race, can be similarly constructed with sets of antonym pairs such as $\overrightarrow{masculine} - \overrightarrow{feminine}$ or $\overrightarrow{black} - \overrightarrow{white}$, respectively.[9]

The process we propose for identifying cultural associations with word embeddings is diagrammed in Figure 2. To identify the cultural valence of a word, we calculate the orthogonal projection of the word vector onto the cultural dimension of interest. Because vectors are normalized, the projection of a word vector onto a "cultural dimension" vector is equivalent to the cosine of the angle between the two vectors. For instance, to determine the affluence association for the word "tennis," we project $\overrightarrow{tennis}$ onto the class dimension of $(\overrightarrow{affluence} - \overrightarrow{poverty}) + (\overrightarrow{rich} - \overrightarrow{poor}) + (\overrightarrow{priceless} - \overrightarrow{worthless}) + \ldots$. In this case, a more positive projection would indicate an association with affluence, and more negative values an association with poverty.[10] By comparing the projections of multiple words on a single cultural dimension, we can compare their connotations within the given spectrum of meaning.

Panel A of Figure 2 shows the construction of an affluence dimension by averaging the differences of several related antonym pairs. Panel B depicts how, by projecting the names of several sports onto the affluence dimension, we find that "boxing" and "camping" project onto the poor side of the dimension, "baseball" and "basketball" are nearly orthogonal to affluence, indicating no strong class association, and "golf," "tennis," and "volleyball" all project rich. Panel C shows how this process can be repeated for another dimension, in this example gender, and how words may be simultaneously positioned along multiple cultural dimensions. The angle between these dimensions can be calculated to capture the similarity between axes of cultural meaning, and it can be evaluated at multiple time points to trace shifts in categorical relations. Induced dimensions like affluence or gender will be approximately orthogonal if those dimensions are semantically and contextually unrelated.[11] When the angle between dimensions deviates from 90 degrees, it suggests a meaningful relationship between them, as we will demonstrate.

Our technique for identifying cultural dimensions is closely related to recent work using word embedding models to detect bias[12] in texts. Caliskan, Bryson, and Narayanan (2017) show that a word's position relative to gendered or racialized labels in a word embedding model is strongly associated with that word's associations measured by Implicit Association Tests (IAT) capturing unconscious bias (Greenwald, McGhee, and Schwartz 1998). They use this evidence to argue that word embedding models reveal

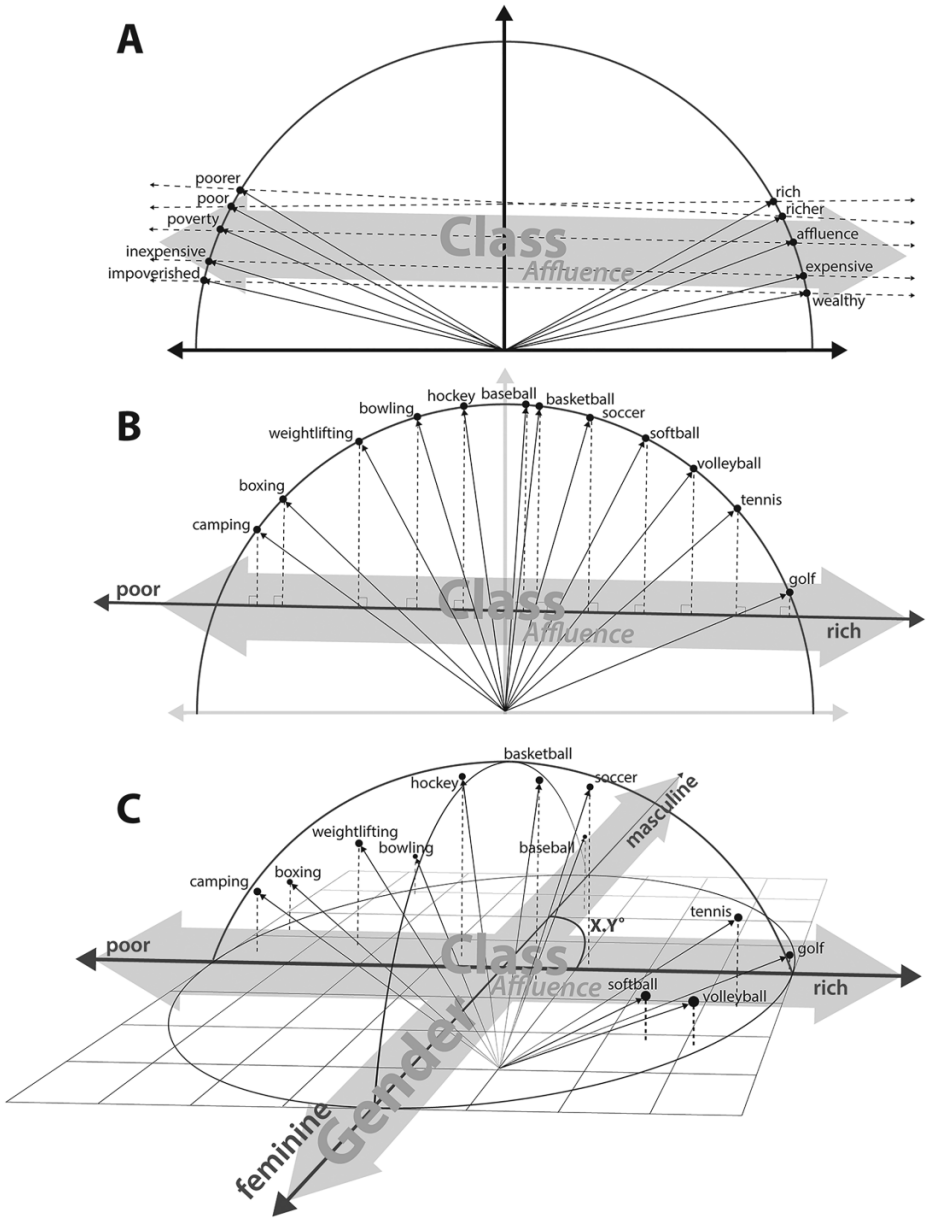

Figure 2. Conceptual Diagram of (A) the Construction of a Cultural Dimension; (B) the Projection of Words onto That Dimension; and (C) the Simultaneous Projection of Words onto Multiple Dimensions

negative racial and gender stereotypes implicit in texts. Bolukbasi and colleagues (2016) deploy a related approach to neutralize such biased associations in text.

Our work builds on these studies in several ways. First, we show that word position within the embedding model correlates not only with the unconscious associations but also with widely shared, conscious associations measured by surveys. Second, we argue that the method presented here detects not only hidden biases but a vast array of cultural valances. Many associations we find here are indeed biased: "criminal" is consistently

found to be more "poor" than "rich," and "scientist" more "masculine" than "feminine." Word embeddings include harmful stereotypes, however, only because they accurately reflect cultural systems that are themselves rife with such stereotypes. Thus, it is rarely in cultural analysts' interest to "debias" a word embedding model as Bolukbasi and colleagues (2016) propose. Rather, it is by interrogating these biases, as well as the neutral cultural associations present in the models, that analysts can cultivate an understanding of the multifaceted word meanings and cultural categories deployed in text.

Garg and colleagues (2018) begin to move in this direction by using word embeddings to study change in gender and ethnic stereotypes over time. By examining change in stereotypes, they recognize them as more than simply distortions of the semantic system, but rather meaningful characteristics that reflect the culture in which texts were produced. Their analysis, however, like Bolukbasi and colleagues (2016) and Caliskan and colleagues (2017), remains couched in the analysis of bias rather than cultural categories in general. Our approach builds on these studies by interpreting the dimensions of embedding models as representative of meaningful cultural categories rather than simply biases, distortions, or deficits in the semantic system. We then use these dimensions as tools to illuminate complex cultural relations associated with class in a given social context, across contexts, and over time. More broadly, this article is the first to specifically demonstrate the utility of word embedding models for sociological and cultural inquiry.

WORD EMBEDDINGS AND CULTURAL THEORY

Word embedding models at once align and contend with dominant theories of culture in a number of significant ways. First, word embedding models are fundamentally *relational* in how they represent meaning. "Posh" only has meaning in that it is positioned near "wealth" but closer to "style," near "fashion" but closer to "rich," and distant from "plain" and "cheap." At the same time, "wealth," "style," "fashion," "plain," and "cheap" themselves achieve meaning through their position relative to "posh" and other words in the space.

This purely relational approach to modeling meaning parallels a diverse body of cultural theorizing, including the structuralist models of meaning developed by Saussure (1916), which posit that individual signifiers are arbitrary and acquire meaning only through placement in a complex system of signification. The fundamental insight that meaning is not immanent within words and phrases but rather coheres within a broader cultural system is inherent in any word embedding analysis. This theoretical congruence makes word embeddings an effective tool for advancing empirical research within relational frameworks popular among contemporary theorists of culture (DiMaggio 2011; Emirbayer 1997; Mische 2011).

Meaning and Dimensionality

Dominant theories of culture often conceptualize meaning in terms of semantic dimensions. Considering objects' multiple, cross-cutting valences along dimensions such as good/bad, rich/poor, and masculine/feminine not only resonates with structuralist thought (Douglas 1966; Lévi-Strauss 1963), but it is central to contemporary intersectionality, affect control, and field theories. Inducing labeled cultural dimensions from word embeddings thus makes it possible to operationalize and engage with these prominent theoretical traditions using large-scale text. Distances between terms can also be fruitfully analyzed without imposing labeled semantic dimensions onto the space. Word embeddings may therefore be applied to "non-dimensional" theories of meaning, such as those based on cognitive prototypes or family resemblances (Rosch and Mervis 1975; Tversky and Gati 1978).

The ability of word embedding models to simultaneously locate objects on multiple cultural dimensions, including race, gender, class, and many others, makes them a powerful tool for studies of intersectionality. The fundamental insight of the intersectionality literature is

that cultural categories, particularly those of identity, cannot be isolated and understood independently (Crenshaw 1991; McCall 2005). Rather, analysts must always consider ways in which the meanings of cultural categories change as they overlap and intersect one another. Interrogation of the intersection of cultural categories becomes empirically tractable through word embedding models.

For example, comparing words that project high on both affluence and masculinity to those that project high on affluence and femininity will reveal how markers of class differ across gender lines within the cultural world in which the texts were produced. The theory that identity is defined by numerous cross-cutting and overlapping categories is itself predicated on a "high-dimensional" model of culture similar to that modeled by Euclidean word embeddings. Indeed, the empirical success of word embedding models to represent cultural dimensions promotes a radical view of intersectional identity, modeled not as a low-dimensional matrix, but rather a high-dimensional array composed of hundreds or thousands of interacting cultural associations.

Our use of word embeddings also shares much in common with Osgood's semantic differential method, which similarly rates words along cultural dimensions. In the semantic differential method, respondents are asked in an interview to place words on culturally meaningful spectra: for example, "Is 'dictator' closer to 'smooth' or 'rough?'" (Osgood, Suci, and Tannenbaum 1957). A key finding from this method is that much of the variance across all dimensions tested can be explained by just three core factors: evaluation (good versus bad), potency (powerful versus weak), and activity (lively versus torpid). Osgood's insight matured within sociology into Heise's (1979, 1987) affect control theory, which posits that individuals interpret events and plot courses of action by accounting for culturally based affective meanings, operationalized as the position of words on evaluation, potency, and activity (EPA) dimensions (see also Schröder, Hoey, and Rogers 2016).

Osgood and colleagues' (1957) work has at times been used to argue for the low dimensionality of meaning systems, but this interpretation overlooks key findings of the semantic differential research program. When Osgood and colleagues (1957) had respondents rate words on a set of semantic dimensions purposely selected to be unrelated in meaning, they found the EPA dimensions captured a relatively small portion of the total variance. Motivated by such results, Osgood (1969) concluded that the semantic differential only effectively captures the "affective" components of objects' meanings while systematically missing more denotative elements. The recent discovery that word embedding models require upward of 200 dimensions to successfully recover complex semantic relationships suggests that although three dimensions may be able to coarsely bin concepts and predict approximate human responses, higher dimensionality enables fine-grained classification along a rich set of distinctions particularly useful for sociologists, who are often concerned with subtle nuances of meaning between specific dimensions, such as gender, status, and education.

Word embedding models also operationalize and extend key elements of Bourdieu's field theory. Bourdieusian cultural fields offer a model of how individuals, objects, and positions in social structure are located relative to one another in structurally homologous "social spaces," with relations between entities described in terms of "distances" (Bourdieu 1989). Bourdieu (1984) frequently represented these social spaces geometrically using the method of correspondence analysis (Greenacre 2017), rendering distances between entities and meaningful dimensions of the field visible by placing them in a two-dimensional plane. By overlaying the space of economic relations with the homologous space of cultural relations, Bourdieu underscores how social class operates at once materially and symbolically.

The vector-space models produced by word embeddings similarly position objects relative to one another in a shared space based on cultural similarity. By leveraging the wealth of information contained in a large corpus, however, word embeddings are able

to position words in a semantically-rich, high-dimensional space that need not be reduced to low dimensionality for interpretation. Indeed, the low-dimensional projection of correspondence analysis operationalizes a theory of cultural capital that is itself low-dimensional: social actors struggle to obtain and maintain dominant positions within a cultural field through a single currency of cultural capital and a single dimension of status-distinguishing tastes and preferences (Bourdieu 1984).[13] In this vein, Lamont (1992) criticizes Bourdieu's approach for overemphasizing distinctions based on aesthetic cultivation such as common/rare while neglecting moral distinctions such as honest/dishonest or fair/unfair.

By preserving higher dimensionality in a cultural space, word embeddings can facilitate the development and testing of high-dimensional theories for how actors acquire and exploit varied cultural capitals along multiple dimensions of distinction. Moreover, identifying cultural dimensions using antonym pairs does not require interpreting orthogonal dimensions like correspondence analysis, but instead allows analysts to examine relations between correlated but distinct semantic dimensions. The high dimensionality of word embeddings thus leaves room for complex interrelations between multiple axes of cultural distinction and opens the relationship between these axes as grounds for empirical investigation.

DATA AND METHODS

Our investigation relies on multiple data sources,[14] first for validation of our method and second for examination of historical trends in the cultural dimensions of class. To determine the ecological validity of our general approach, we compare results from word embedding models to human-rated cultural associations assessed by surveys, both contemporary and historical. Having established the validity of our method, we train word embedding models on Google Ngrams text from books published over the span of the twentieth century, and we use these models to interrogate broadly shared understandings of social class.

Surveys of Cultural Association

To establish a basis of comparison between human-reported associations and associations represented in word embedding models, we fielded a survey of cultural associations to 398 respondents on Amazon Mechanical Turk. The survey was fielded in 2016 and 2017 and was open only to Mechanical Turk users located in the United States. Although our sample cannot be said to be representative of the general U.S. population, responses to basic demographic questions indicate wide diversity in age, gender, and racial composition (Levay, Freese, and Druckman 2016). To improve representativeness, we apply post-stratification weights to the sample, weighting on race (white, black, or other), education (bachelor's degree or less), and sex (male or female). The results presented here include post-stratification weighting, but unweighted models produce substantively similar findings. This survey and the weighting procedures are detailed in Appendix Part B.

In the survey, respondents were asked to rate 59 different items on scales representing association along class, race, and gender lines. All questions followed the format, "On a scale from 0 to 100, with 0 representing *very working class* and 100 representing *very upper class*, how would you rate a *steak*?" For measuring race and gender associations, the survey posed similarly worded questions, replacing "working class" and "upper class" with "white" and "African American," or "feminine" and "masculine," respectively. A full list of items asked on the survey is available in Appendix Table B1. Words were selected in seven topical domains: occupations, foods, clothing, vehicles, music genres, sports, and first names. A diverse array of topical domains were chosen to test the capacity of word embedding models to detect cultural associations across very different subjects. Specific terms were selected within each topical domain to ensure high variance across dimensions.[15] We calculate the weighted mean of responses for each item, and we use these means as our estimates of a general cultural association. The end product

is thus a rating between 0 and 100 on a class dimension, a race dimension, and a gender dimension for each of the 59 words listed in Table B1. Measurement of broadly shared cultural associations with a Mechanical Turk survey is likely to suffer from bias and measurement error, but these weaknesses should only attenuate the correspondence between the surveyed associations and those recovered from word embedding models. Therefore, the associations presented here between survey and word embedding models can be interpreted as conservative estimates.

For historical validation, we draw on a similar dataset collected in the 1950s by semantic differential researchers. To produce a standard set of word scores for social psychologists to use across studies, Jenkins, Russell, and Suci (1958) had 30 college students rate 360 common terms on 20 semantic dimensions, such as *hard-soft* and *good-bad*, and published a table reporting the average rating for every word on each semantic dimension. We use these average scores as measures of self-reported cultural associations from the 1950s, enabling us to at once test a broader range of semantic dimensions and validate word embeddings for historical analysis. We exclude 11 terms from the analysis either because they are two-word phrases (e.g., "neurotic man") or they did not appear frequently enough in the Google Ngrams text to be rendered in the vector space (e.g., "briny"), resulting in a total of 349 words used in the analysis, each scored on 20 semantic dimensions.

Word Embedding Data

We analyze several word embedding models trained on multiple textual archives. The majority of our analyses utilize embedding models trained on publicly-available Google Ngram texts. The Google Ngram corpus, the product of a massive project in text digitization across thousands of the world's libraries, distills text from 6 percent of all books ever published (Lin et al. 2012; Michel et al. 2011). Any sequence of five words that occurs more than 40 times over the entirety of the scanned texts appears in the collection of 5-grams, along with the number of times it occurred each year. Because word embeddings require local context to determine the meaning of words, we limit our analysis to the collection of 5-grams, and we exclude data on the occurrence of 4-grams, 3-grams, 2-grams, and single words.[16] All characters were converted to lowercase in preprocessing to increase the frequency of rare words. Although the Google Ngrams corpus does not represent one single, identifiable voice, it includes a vast number of documents spanning a variety of genres, including novels, government documents, academic texts, and technical reports, making it sensitive to subtle associations that appear diffusely in general discourse. Google Ngrams are poorly suited for identifying subcultural or contextually-specific meanings, but they are able to successfully capture pervasive and widely-shared meanings that characterize terms across contexts.

The Google Ngram corpus is a uniquely powerful source of textual data, but it suffers from various weaknesses. Google Ngrams have been subject to criticism because the composition of the corpus in a given year may not be representative of total literary output (Pechenick, Danforth, and Dodds 2015). We also recognize that authors whose books and periodicals appear in Google Ngrams are by no means a culturally representative sample of the U.S. general public. Instead, we must limit our generalizations to a relatively elite, "literary public"; a group whose cultural framework of class is consequential given its wide dissemination but possibly different from more marginalized populations underrepresented in the corpus. Word embedding models require very large collections of text to reproduce accurate semantic relationships, and Google Ngrams provide the largest and most extensive sampling of historical English texts. Furthermore, our contemporary and historical validations suggest Google Ngrams over the twentieth century are able to produce cultural associations that mirror human reports on numerous diverse semantic dimensions. We therefore proceed with Google Ngrams as our primary source of historical text and reflect on limitations of our analyses in the discussion.

We train word embedding models on Google Ngrams texts for both the historical analysis of class and contemporary validations. The Google Ngrams corpus contains metadata specifying the year of publication for each string of text, making it possible to trace semantic changes over time. We divide the corpus by decade, training separate models on texts from 1900 to 1909, 1910 to 1919, and so on through 1990 to 1999, resulting in 10 independently constructed word embedding models. By comparing these models side-by-side, we are able to trace macro-cultural trends over this 100-year period. Only words that appear at least 25 times are rendered in the model for a given decade, thus excluding words mentioned too rarely to be accurately placed.

For contemporary validation, we train an embedding model on Google Ngrams of publications dating from 2000 through 2012. We use this range of years because Google Ngrams do not include publications more recent than 2012, and this duration is similar to those used in our historical analyses. For additional validation, we compare the performance of the Google Ngrams embedding to two widely used, pre-trained embeddings: one trained on contemporary Google News text with *word-2vec* and one trained on a broad scraping of website text from the Common Crawl with *GloVe*. These alternative embeddings are discussed in greater detail in Appendix Part A.

For validation with the 1950s semantic differential survey data, we use the same embedding model trained on 1950 to 1959 Google Ngrams that we use in our historical analysis. We train all word embeddings with *word2vec* skipgram architecture with 300 dimensions, following standards that prior research found to be effective in solving analogy tasks (Mikolov, Chen, et al. 2013). We also test the validity of our approach across different corpora and word embedding algorithms, including large samples of twenty-first-century news and webpages, which we detail in Appendix Part A.

We identify a diverse set of cultural dimensions in our embedding models for validation and for historical analysis. For contemporary validation, we construct cultural dimensions corresponding to three core sociological axes of classification: affluence, gender, and race (black/white). For historical validation, we construct 20 cultural dimensions corresponding to those measured by Jenkins and colleagues (1958). Finally, for our historical analysis of collective understandings of class, we construct cultural dimensions corresponding to those identified in the literature as being constitutive of, or deeply intertwined with, social class. For these analyses, we again construct dimensions for affluence and gender, and we add dimensions of education, employment (owner/worker), status, cultivation, and morality.

Measuring Cultural Dimensions

To identify cultural dimensions in word embedding models, we average numerous pairs of antonym words. Cultural dimensions are calculated by simply taking the mean of all word pair differences that approximate a given dimension, $\frac{\sum_{p}^{|P|} \overrightarrow{p_1} - \overrightarrow{p_2}}{|P|}$, where p are all antonym word pairs in relevant set P, and $\overrightarrow{p_1}$ and $\overrightarrow{p_2}$ are the first and second word vectors of each pair.[17] The projection of a normalized word vector onto a cultural dimension is calculated with cosine similarity, as is the angle between cultural dimensions.

We bound our estimates with 90 percent confidence intervals constructed through a nonparametric subsampling approach. This method involves splitting the corpus into 20 non-overlapping subsamples, independently constructing embedding models on these 20 subcorpora, and calculating the desired estimates on all 20 embedding models. The variance between these estimates is then used to quantify how sensitive the estimates are to particular usages in the text. If a word is used infrequently and appears in several very different contexts, it will produce a wider error bound than a word used frequently in consistent contexts. Technical details regarding our calculation of these confidence intervals is available in Appendix Part C.

To assemble effective lists of antonym terms, we used five thesauri: three contemporary (*Bartlett's Roget's Thesaurus* 1996; *Oxford Thesaurus* 1992; *Webster's Collegiate Thesaurus* 1976) and two historical (Roget 1912; Smith 1903). Drawing words from historical thesauri ensures our list of terms is robust for the early and more recent decades of the twentieth century. Indeed, certain terms only appear in the early decades of the century (e.g., "luxuriant" and "penurious") and others only appear at the end (e.g., "privileged" and "underprivileged"). Antonym pairs that do not appear in a given decade's embedding are excluded from calculation of the average cultural dimension. As a result, the terms that comprise a cultural dimension shift as the terms used in discourse to designate the cultural dimension themselves shift.

Some cultural dimensions are characterized by a much larger set of words in the English language than others, leading to substantial differences in the number of antonym pairs included for each. Furthermore, selection of antonym pairs requires some discretion on the part of the analyst, because thesauri often contain a wide range of loosely synonymous terms inappropriate for the given analysis. We present supplemental analyses suggesting that cultural dimensions constructed from fewer antonym pairs may be less robust, but results do not differ substantially between those constructed from 10 pairs and those trained on 40. We further find that the exact ways words are paired (e.g., $\overrightarrow{rich - poor}$ instead of $\overrightarrow{rich - impoverished}$) has a minimal effect on the effectiveness of the dimension in predicting human-rated associations. The full sets of antonym pairs we use for all cultural dimensions analyzed in this study are listed in Appendix Part D, and robustness checks are presented in Part E. Corpus sizes are listed in Appendix Part F.

We contextualize our cultural analysis of class by comparing associations held in the general public to those expressed in sociological literature. To produce clear grounds for formal comparison, we compute word embedding models trained on a corpus of all sociology articles published in the twentieth century in the JSTOR collection. The class-based associations we find in this corpus generally accord with widely recognized disciplinary trends (see Appendix Part H).

RESULTS

Validation of Cultural Dimensions

We validate the ability of word embedding models to reflect widely shared cultural associations by calculating the Pearson's correlation between a word's mean rating on a given survey scale and the word's projection on the corresponding cultural dimension in an embedding model. Correlations are calculated using the 59 terms listed in Appendix Table B1. We compare the validation results from the Google Ngrams embedding to two widely-used, pre-trained embedding models to illuminate the strengths and weaknesses of Google Ngrams compared with other corpora. Results are presented in Table 1.

The first column of Table 1 presents correlations between survey responses and word vector projections for class. We see that association for the Google Ngrams embedding is .53, and correlations with the two alternative embeddings are .57 and .58 (details in Appendix Part A). The second column displays the correlation between gendered associations in survey response and projection on the embedding's gender dimension. For gender associations, the Google Ngrams embedding correlates with surveyed ratings at .76, and alternative embeddings correlate at .88 and .90. These correlations attest to how well a gender dimension elicited from the word embedding model corresponds to contemporary individuals' understandings of masculinity and femininity. The third column shows correlations between word embedding projections and survey ratings for racial associations. The Google Ngrams corpus does relatively poorly in this test, correlating at only .27 with survey response. Other embeddings range widely from .42 to .75.

There are many possible explanations for the Google Ngrams' relatively poor performance in picking up racial associations. The

Table 1. Pearson Correlations between Survey Estimates and Word Embedding Estimates for Gender, Class, and Race Associations

	Class (Affluence)	Gender	Race
Google Ngrams *word2vec* Embedding†	.53	.76	.27
Google News *word2vec* Embedding	.58	.88	.75
Common Crawl *GloVe* Embedding	.57	.90	.44

Note: $N = 59$, except †$N = 58$ where one word measured in the survey did not occur frequently enough in the text to appear in the word embedding.

subject matter of news articles and general internet postings may be imbued with more racial associations than the Ngrams corpus, which contains significant non-fiction, including technical reports and scientific publications without narrative content that could invoke ambient, contemporary racial associations within that embedding model's projections. Additionally, as noted earlier, the Google Ngrams text were reduced to lowercase in preprocessing, which decreased the available number of antonym word pairs for constructing the race dimension from seven to five, possibly resulting in decreased accuracy of the dimension. This poses pronounced difficulties for analyses of race, given that the semantic dimension *black-white* will likely capture a host of associations related to color but unrelated to race. Because of these difficulties in recovering racial associations from the Ngram corpus, we refrain from analyses of race in our subsequent analyses of class associations over time.

Figure 3 plots the correspondence between word embedding models and our survey of cultural associations. The figure reveals how several music genres—jazz, rap, opera, punk, techno, hip hop, and bluegrass—are arrayed on the cultural dimensions of class and race by survey response and the word embedding trained on Google News, with the average survey rating of a word depicted in black and the projection in gray. Comparing survey ratings to word embedding projections, we see striking similarity in the relative positions of words. In both methods, opera holds the association of being both high class and white. Techno, punk, and bluegrass are similarly white but of distinctly lower class than opera. On the right end of the panel, jazz is associated with both African Americans and high class, whereas hip hop and rap tend toward the working class. Projecting words simultaneously into multiple dimensions, it is clear how word embeddings can be used to examine intersectionality by revealing how class markers vary across racial lines.

We next validate results from an embedding trained on 1950s Google Ngrams text on data from a semantic differential survey fielded in 1958 (Jenkins et al. 1958). This validation assesses the ability of Google Ngrams embeddings to capture historical associations and their capacity to reflect a wide variety of semantic dimensions beyond core sociological categories. We take the same set of 349 words and 20 cultural dimensions measured by Jenkins and colleagues and produce a corresponding embedding-derived dataset by projecting the respective word vectors onto corresponding cultural dimensions from the embedding model. The sets of antonym pairs used to construct these cultural dimensions in the embedding are listed in Appendix Table D2.

Figure 4 depicts Pearson correlations between word embedding projections and human ratings for 20 semantic dimensions. We find a statistically significant ($p < .01$), positive association between human-rated associations and embedding projection on all dimensions. Many correlations are impressively high; correlations on six dimensions exceed .60, including *kind-cruel, good-bad, beautiful-ugly*, and *true-false*. We see more modest correlations on other dimensions, but we also find that lower correlations generally correspond to lower variance in average human ratings on those dimensions. This

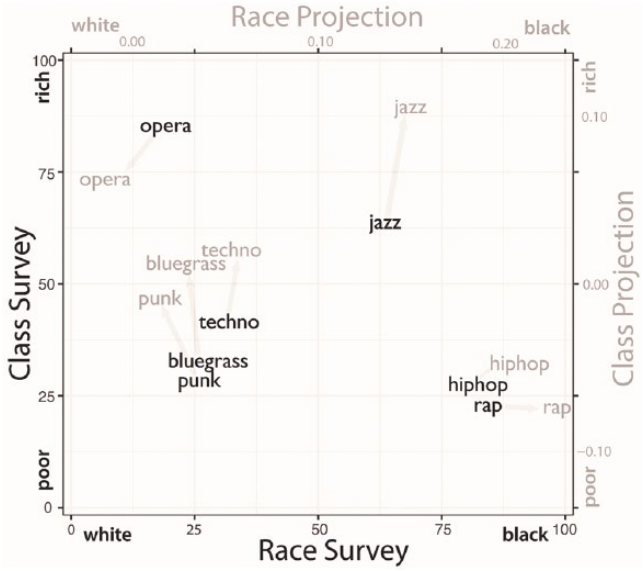

Figure 3. Projection of Music Genres onto Race and Class Dimensions of the Google News Word Embedding (Gray) and Average Survey Ratings for Race and Class Associations (Black)

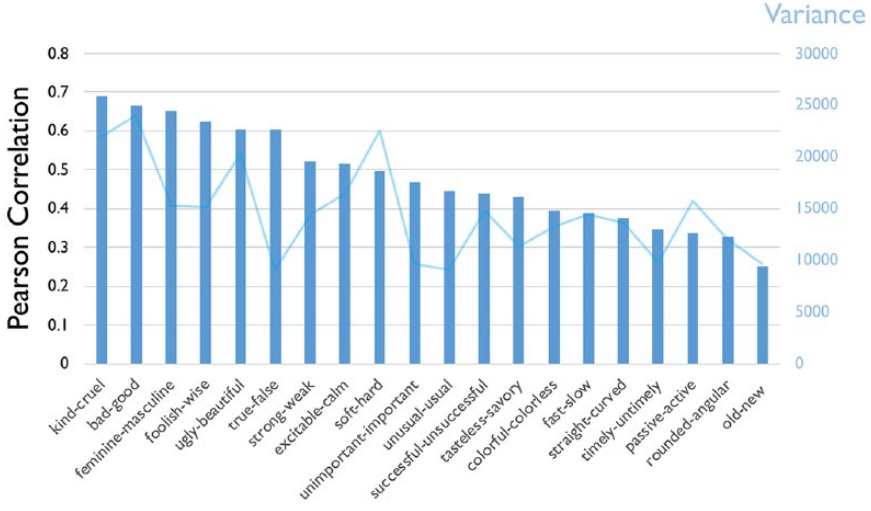

Figure 4. Correlations between Word Embedding Projections and Human-Rated Associations on 20 Semantic Dimensions, Alongside Variance of Average Human-Ratings on Those Dimensions; 1950 to 1959 Google Ngrams Corpus

means dimensions with many strongly-rated words on both ends of the spectrum are more successfully captured by word embedding models. For example, subjects tended to rate most words near the middle on the *rounded-angular* dimension, suggesting they do not register strong associations. Unsurprisingly, these subtle and potentially more noisy associations are more difficult to capture from text. We engage semantic differential theory more deeply with supplemental analyses in Appendix Part G, showing that subspaces of word embeddings can reproduce the dimension reduction typical of semantic differential

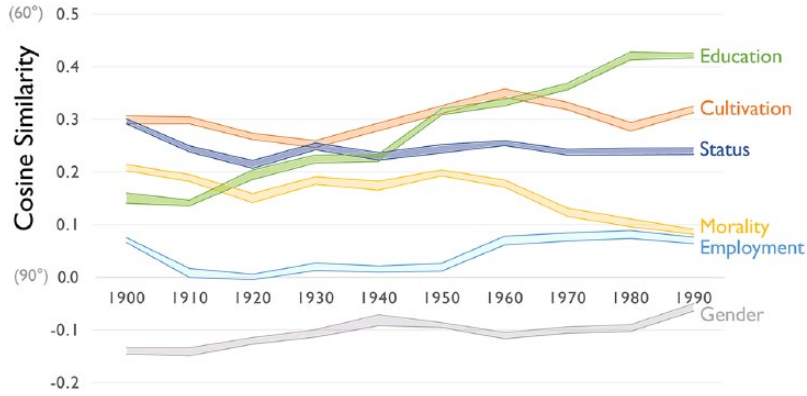

Figure 5. Cosine Similarity between the Affluence Dimension and Six Other Cultural Dimensions of Class by Decade; 1900 to 1999 Google Ngrams Corpus
Note: Bands represent 90 percent bootstrapped confidence intervals produced by subsampling.

analysis, but the full spaces cannot be represented in lower dimensionality without considerable loss of information.

Meanings of Class across the Twentieth Century

Having validated word embeddings' capacity to capture meaning along many semantic dimensions, we apply this method to unpack multiple dimensions of class and explore their interrelation in the United States over the twentieth century. Specifically, we seek to discover how shared understandings of social class evolved during a period of dramatic economic transformation and which class components remained stable in spite of these developments. We analyze five dimensions of class identified prominently in sociological theory: affluence, employment (owner/worker), status, education, and cultivation.[18] For additional comparison, we also construct dimensions for two categories that theorists have noted as deeply intertwined with class: morality (Lamont 1992; Lerner and Miller 1978; Skeggs 1997) and gender (Reay 1998; Veblen [1899] 1912).

First, we focus on the cultural dimension of affluence, the ubiquitous class marker that anchors modern understandings of socioeconomic inequality (Piketty 2014). We begin by investigating how affluence has changed its relations to the other components of class. To accomplish this, we calculate the angle between each class dimension and the other six dimensions of interest, and then we explore how these angles shift over the course of the twentieth century.[19] Figure 5 displays the angle, measured in cosine similarity, between the affluence dimension and each of our other six cultural dimensions: employment, status, cultivation, education, morality, and gender. We observe general stability in the relations between affluence and other cultural dimensions, with a few key exceptions. Interestingly, the dimensions most parallel to affluence at the start of the twentieth century are cultivation and status. These are closely followed by morality, gender, and education, respectively. Affluence notably manifests the most modest association with employment position.

It is illuminating to consider places where popular cultural associations run counter to understandings of class expressed within sociology. For example, gender's association with affluence is weakly negative within general discourse, implying an association between affluence and femininity. This finding runs contrary to the sociological expectation that masculinity would be associated with affluence, given that men in the United States earn greater income and control more wealth than women. Such disjunctions between sociological and conventional

understandings of class can be verified by comparing results from embeddings trained on Google Ngrams to those trained on sociological literature. We provide this empirical comparison in Appendix Part H.

The popular association of femininity with affluence in general discourse is less surprising when affluence is considered from a historical perspective. Veblen ([1899] 1912) documented how wives and daughters were frequently used as vessels for men's "vicarious consumption," and how women's distance from toil in the workplace served as a marker of class in affluent society. Similarly, Zelizer (1989) notes that women's money in the early twentieth century was commonly considered "pin money," earmarked for extravagant and indulgent purchases, whereas men's money was reserved for mundane necessities. Projections in historical Google Ngram embeddings reinforce this interpretation. Among the 10 nouns most highly projecting on the affluence dimension in the first decade of the twentieth century are "fragrance," "perfume," "jewels," and "gems," all of which project strongly feminine, suggesting that upper-class women were cultural mannequins for the display of wealth.

Employment position, either as a worker or owner, is similarly prominent in sociological understandings of wealth accumulation in the late twentieth century, yet its relationship with affluence in general discourse is weak. Across the entire century, the employment association is dwarfed by affluence's relationship with the symbolic factors of cultivation and status. Again, although these findings do not align with how sociologists conceive of social class, they accord with certain key theories of class representation. Bourdieu's concept of "misrecognition" and Marx's earlier concept of fetishism both describe how relations of production undergirding systems of economic stratification are obscured while the outward trappings of class, displayed through consumption patterns, remain visible and culturally salient. This perspective also anticipates the tight association between affluence and cultivated tastes in popular discourse throughout the twentieth century.

Most cultural dimensions of class remain remarkably stable over the century, yet we observe a striking change in the relationship between dimensions of affluence and education. Although their association is only weakly positive at the dawn of the twentieth century, it surpasses all other dimensions by the century's close, suggesting that education and affluence became increasingly synonymous. It is possible, however, that this relationship is mediated by notions of cultivation. Cultural capital scholars have long argued that education reproduces patterns of economic stratification by providing students with cultural knowledge and dispositions that exert signaling effects in the market (Collins 1979; Lamont and Lareau 1988; Lareau and Weininger 2003). In embedding terms, this would imply that words with strong, positive educational valence only have an association with affluence insofar as they also project strongly on cultivation. To determine the extent to which education's semantic connection to affluence is mediated by cultivation, we use regression to model their relationship and parse the geometry of this cultural space. OLS regression estimates the expected slope along one dimension of the vector space while holding others fixed. Given that non-independence is inherent to word embedding models, we do not intend the quasi-experimental interpretation of regression common in sociological analysis.[20]

Figure 6 presents results from OLS regressions of cultivation and education projections predicting affluence projections. Interestingly, when adjusting for cultivation, projection on the education dimension actually exhibits a weakly negative association with affluence in the first half of the twentieth century. In other words, for two words with the same cultivation projection, the word with a greater education projection would have a *lower* expected affluence projection, suggesting education's cultural association with affluence was a byproduct of its association with cultivation, sophistication, and refinement. Indeed, at the beginning of the twentieth century, education at times implied a necessity to participate in the world of work rather than living comfortably on rentier income (Veblen [1899] 1912).

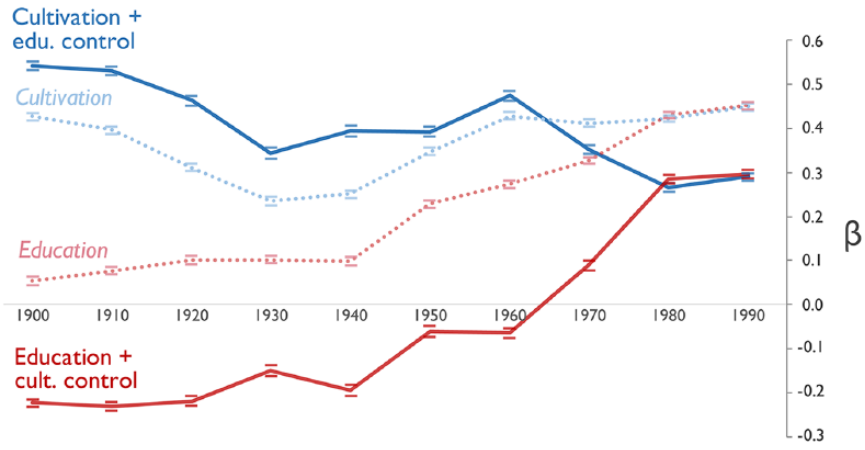

Figure 6. Standardized Coefficients from OLS Regression Models in Which Word Projections on Cultivation and Education Dimensions Predict Projection on the Affluence Dimension; 1900 to 1999 Google Ngrams Corpus
Note: A separate OLS regression model is fit for each decade; $N = 50{,}000$ most common words in each decade.

This relationship transforms over the course of the century. By the 1990s, education projections strongly associate with affluence, independent of cultivation. This finding suggests that by the end of the twentieth century, education represents a marginally distinct cultural marker of affluence, no longer redundant with cultivation. Education's cultural association with affluence was mediated by cultivation at the beginning of the twentieth century, but meanings associated with education and affluence intertwined as education became increasingly essential for socioeconomic attainment.

This finding is ironic when considered against concurrent trends in sociological theories of class. With the rise of cultural capital theory, critical scholars suggested that education influences income by bestowing forms of cultural distinction rather than by providing practical knowledge and skills (see Appendix Part H). Yet, at the moment sociologists came to see education as operating via cultivation, the opposite occurred in public perception, where education became imbued with independent connotations of affluence as its demand among elite, well-paid occupations rose (see Appendix Part I).

In Figure 7, we broaden our focus away from affluence to comprehensively view relations between the multiple dimensions of class, displaying each dimension's angles with all others. In spite of the rapid and encompassing economic transformations of the twentieth century, we find that relations between the cultural dimensions of class remain remarkably constant. Most dimensions that begin close together remain close, and those orthogonal retain their independence. The rank ordering of most angles is preserved for 100 years. Examining which cultural dimensions are correlated and which are independent, we see that cultivation, morality, and education are consistently close together, moderately related to status and affluence, and almost orthogonal to employment position. In fact, employment shows an association with morality in the opposite direction, with bosses carrying an odious cultural valence relative to workers. Despite its negative relationship with morality, however, employment shares modest but positive associations with affluence and status.

Taken together, these results demonstrate a remarkably stable and complex structure among the cultural dimensions of class, with dimensions most closely associated with social distinction—morality, cultivation, and education—clustered on one end, employment position on the other, and status and affluence mediating these otherwise unrelated domains. Observing this structure holistically

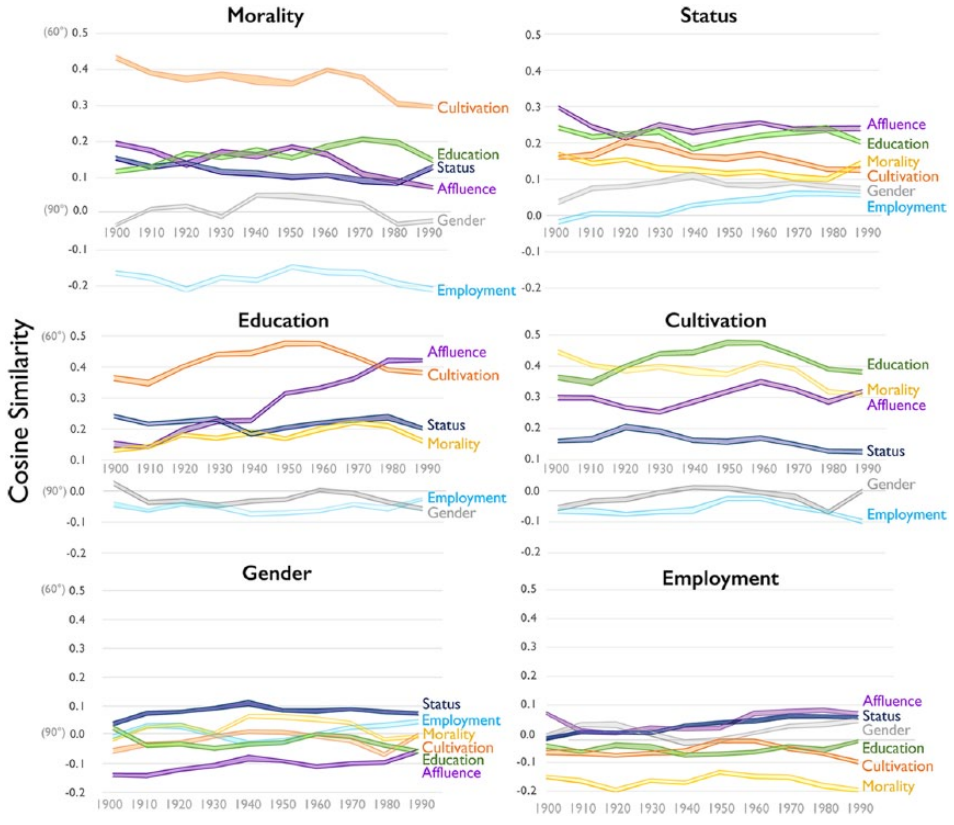

Figure 7. Cosine Similarity between Each Class Dimension and All Others by Decade; 1900 to 1999 Google Ngrams Corpus
Note: Bands represent 90 percent confidence intervals produced by subsampling.

helps clarify the cultural relationship between these dimensions, which are rarely considered simultaneously. Status and affluence are at once colored by two distinct cultural valances. On one side, they carry connotations of ownership and power. On the other, they are signaled by refinement, virtue, and edification—characteristics with little association to power and industry.

This complex semantic structure requires high dimensionality for representation. In Figure 8, conceptual diagrams illustrate that two dimensions are not enough to reproduce the angles between any three dimensions of class without significant distortion. If the relation between employment and cultivation is held at its measured value of 90°, then it is impossible to keep the angle between cultivation and status at 85.5° while also maintaining that between employment and status at 79.3°. Thus, even when considering cultural categories closely related to class, high dimensionality is necessary to preserve crucial distinctions between meanings.

Finally, we turn from relations between class dimensions to focus on the stability of meanings *within* dimensions. We operationalize stability as the correlation between words' projection on a given dimension in one decade and their projection in subsequent decades. Figure 9 displays the stability of projections for the 50,000 most common words on each class dimension. The first line represents the average correlation of word projections in the 1900s with their projections in the 1910s, 1920s, and so on through the 1990s. Similarly, the second line shows the correlation between projections in the 1920s with those in the 1930s, 1940s, and so on. For each decade, a word's projection is highly correlated with its projection the following decade, in most cases greater than .9. This correlation diminishes by

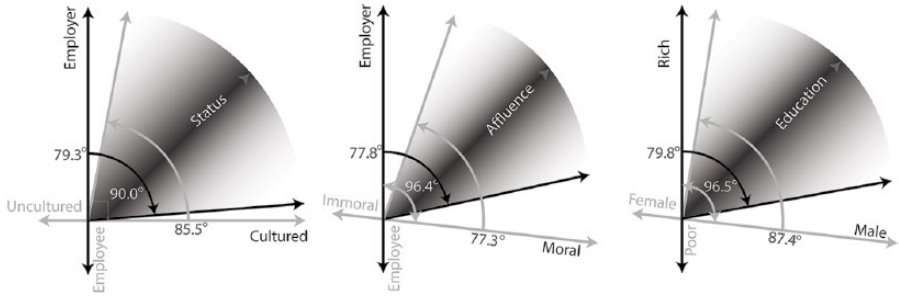

Figure 8. Conceptual Diagram of the Distortions Introduced When Reducing High-Dimensional Embeddings to Two-Dimensions

decade, however, such that the correlation between a word's projection in the 1900s and 1990s falls between .7 and .6. This pattern reveals that beneath the historic stability of class's dimensional structure, there is continuous flux in how cultural markers are positioned along these spectra.

To clarify this process of cultural circulation, we look more closely at how particular words change their projections on the employment dimension over the twentieth century. We select employment because it shows the greatest decline in correlation between the beginning and end of the century in Figure 9. Figure 10 displays the four highest and lowest loading words on the employment dimension in the beginning (1900s to 1910s) and end (1980s to 1990s) of the century, along with exemplary terms that display informative semantic trajectories. The top-left of the figure shows that many terms most strongly associated with the employer position were titles of formal office: "lords," "governor," "mayor," "earl," "bishop," and "secretary." As the century progresses, however, these titles lose ascendency to terms associated with power in an industrial and financialized economy: "promoter," "speculator," "rival," "designer," and "mogul."

The bottom of Figure 10 shows terms associated with the position of worker or employee. The strongest association at the start of the century is with "wage" and "earners"; this attenuates as a greater share of the U.S. workforce becomes contracted and salaried employees. The words "soldier," "muscle," and "bodied" project strongly on the "employee" end of the employment dimension during a period when manual labor comprised a large proportion of the workforce and World War I saw a large share of able-bodied workers enlisted into armed service. These words are displaced over time, with preeminent markers of "employee" at century's end including "retirement," "qualified," and "student." This suggests an emerging cultural image of the worker as white-collar and middle-class. Widespread perceptions of worker problems also shift with time. "Suffering" ceases to be a strong marker, but "unemployed" becomes prominent.

Other results of this analysis are not so easily interpretable. The words "patient" and "expectancy" are among the strongest negative projections on the ownership dimension at the end of the century, suggesting a powerful "employee" valence for both terms. Imaginative explanations for such findings are always conceivable—perhaps a growing recognition of workers as subject to ailment or injury led to an equivalence between "workers" and "patients." Yet this style of post hoc interpretivism is vulnerable to misleading conclusions drawn from statistical flukes. These ambiguous findings provide an instructive example of how inductive approaches must be applied cautiously to word embedding analyses.

DISCUSSION

Summary of the Argument and Results

In this article we introduce word embedding models as a productive method for the analysis of cultural categories and associations. By

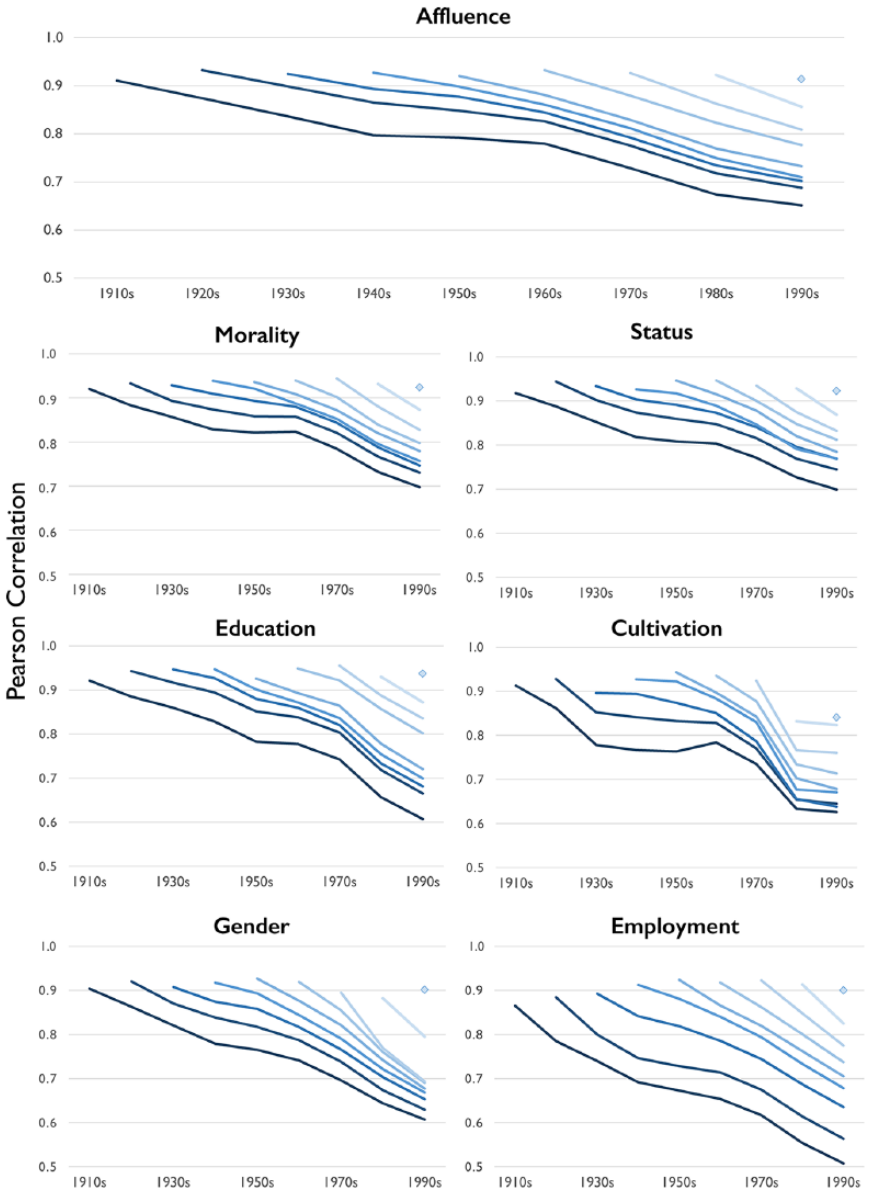

Figure 9. Correlation of 50,000 Most Common Words' Projection in One Decade with Their Projection in Each Subsequent Decade for Seven Cultural Dimensions of Class; 1900 to 1999 Google Ngrams Corpus

representing the relationship between words as the relationship between vectors in a high-dimensional vector space, word embedding models distill vast collections of text into a singular representation while preserving much of the richness and complexity of their semantic relations. We describe how dimensions of word embedding models correspond closely to "cultural dimensions" such as *rich-poor, good-evil*, and *masculine-feminine*, and how the positions of words arrayed on salient cultural dimensions of a word embedding

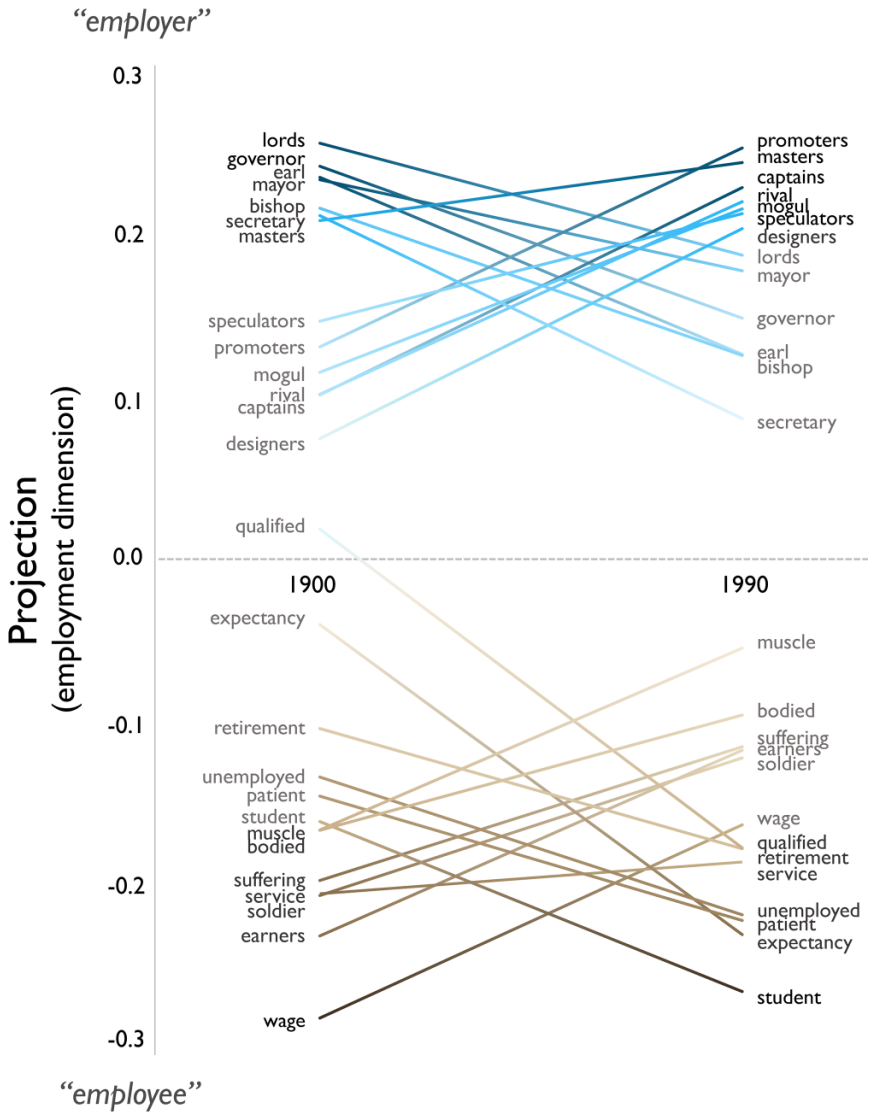

Figure 10. Words That Project High and Low on the Employment Dimension of Word Embedding Models Trained on Texts Published at the Beginning and End of the Twentieth Century; 1900–1919 and 1980–1999 Google Ngrams Corpus

reflect patterns of association and classification within a given cultural system. Furthermore, by calculating angles between cultural dimensions, we are able to investigate relationships between the axes of classification themselves.

After validating our method by comparing multiple word embeddings to contemporary and historical surveys of cultural associations, we apply it to a macro-historical investigation of shared understandings about social class in the United States over the twentieth century. We take up five facets of class and two related cultural dimensions that have been extensively theorized in the past. For each, we identify corresponding dimensions in word embedding models trained on texts produced over the twentieth century. We then measure relations between these class dimensions, bringing to light their dynamics,

but also their stability, in the face of economic and industrial transformation. Our findings reveal that the multiple dimensions of class identified in sociological theory comprise a complex yet stable semantic structure that can only be represented faithfully in high dimensionality. We find persistent, close relations between dimensions of cultivation, morality, and education, and these interrelated spectra are nearly orthogonal or negatively associated with cultural conceptions of the classic Marxian owner/worker relation. Nevertheless, both share a connection to status and affluence, which intermediate them, serving as a cultural nexus between the outward trappings of class and the social relations that produce and reproduce class in the modern world.

The relationships between the cultural dimensions of class remain stable over the century, but locations of individual words on those dimensions are in constant flux. Collectively, these findings suggest that many of the basic dimensions through which class is understood were robust against the twentieth century's tectonic shifts in the organization of economy, industry, and employment. What evolved were symbols used to signify locations in the multidimensional architecture of class.

General Implications of the Study of Culture

The full range of potential applications for word embedding models reaches far beyond the class example presented in this article. Following the general approach piloted here, analysts could use word embedding models to compare the cultural systems represented by literary genres, texts produced by distinct authors, or texts written in different languages (Lev, Klein, and Wolf 2015). A wide array of social collectives, including scientific disciplines, political elites, and contributors to online forums, can be analyzed and compared by training word embedding models on the text they produce. Furthermore, while this article focused on insights produced by identifying, extracting, or comparing "cultural dimensions" from the vector space, we do not maintain this is the only method for utilizing word embedding models to advance social science. Simply calculating the proximity of word vectors can also provide a strong indicator of the similarity or distance between word meanings (Kulkarni et al. 2015).

Word embedding models can further be used to classify and predict which group produced a text, given multiple corpora produced by distinct social groups (Taddy 2015a). Finally, future word embeddings that use hyperbolic or elliptical geometries could be used to systematically capture nonlinear relations in language, such as hierarchy or clustering (Chamberlain, Clough, and Deisenroth 2017; Nickel and Kiela 2017; see Appendix Part A). We argue that a wide range of techniques for productively developing and applying word embedding models to social and cultural inquiry are possible but yet to be developed. Nevertheless, Euclidean word embeddings are conducive to modeling and evaluating intersecting dimensions of culture in a way that maps onto a wide range of cultural theory.

Caveats and Limitations

As well as identifying broad potential, our investigation exposed clear limitations of word embedding models for cultural analysis. First, word embeddings must be trained on very large corpora if the output vector space is to capture subtle and complex associations of interest to culture analysts. Previous studies indicate that analogy tests can only be reliably solved when input text comprises several million words or more (Hill et al. 2014). As a result, groups that do not leave extensive textual records are difficult to study with word embeddings.

Second, the exact algorithmic processes undergirding the training of word embedding models can be highly complex and therefore elude theoretically parsimonious description. Although the word embedding models we present (*word2vec* and *GLoVe*) rely on two-layered neural networks with a single hidden layer, state-of-the-art deep-learning models

deploy many-layered neural architectures with hundreds of millions of parameters for improved performance on natural language and intelligence tasks like question-answering (e.g., Devlin et al. 2018). Added algorithmic complexity can produce more sensitive and informative models, but it may also diminish the researcher's understanding of how the model is generated and what distortions it is likely to produce.[21]

Moreover, word embeddings are not able to adjudicate the suitability of a given corpus for an investigation. Just as rigorous sampling is crucial in interview-based methods, the ability to make cultural inferences about a given group with word embeddings depends on the sample of text utilized in model training. In our analysis, we opted to use a broad sampling of U.S. texts over time. The magnitude of our corpora enables recovery of subtle and diffuse semantic relations, but it requires combining texts produced by very different groups across vastly different social and cultural contexts. The resulting model captures broadly shared meanings that characterize U.S. culture in a given decade, but it levels the cultural heterogeneity of the individuals and groups that articulated them.

Furthermore, we acknowledge that the voices and worldviews published in books digitized by Google are not a random sample of U.S. culture. Poor and marginalized populations are unlikely to have their discourses published in the books, periodicals, and pamphlets that comprise the Google Ngrams corpus. We therefore must limit our population of inference to the U.S. "literary public" in any given decade. The correspondence between these word embedding findings and surveyed Americans on Mechanical Turk suggests the models' associations are prevalent in the general public, but identifying exactly where this generalization succeeds and fails falls beyond the scope of this investigation.

A set of texts should not be taken as a pure or complete reflection of the culture that produced it. Authors may strategically emphasize or obscure semantic associations depending on their goals in producing the text (Jakobson 1960). Factors including the genre, purpose, and audience must be considered when utilizing texts for analysis and inference. Moreover, various elements of culture, such as tacit knowledge and embodied practices, are not inscribed in written discourse and therefore remain overlooked by formal models of text such as word embeddings (Lizardo 2017).

Finally, word embeddings cannot identify the cultural dimensions most important for a given semantic system or social process. Analysts can identify the cultural dimensions of the model that explain the most variance, either across the entire semantic space or within a circumscribed vocabulary. But although high explained variance indicates that terms have strong positive and negative valences along the dimension, it reveals little about how these valences are deployed in social life and to what ends. It is possible that subtle cultural associations may be deeply consequential for action, and thus explained variance could be misleading as an indicator of social significance. We argue that selection of cultural dimensions for analysis should be motivated by theoretical considerations, as we ultimately did here with class, rather than emergent and sometimes arbitrary qualities of the embedding space.[22]

Concluding Remarks

Despite their limitations, word embedding models can serve as a powerful tool for analysts of culture. Word embedding algorithms require large corpora to create informative models, but the amount of digitized text produced and available to analysts is growing exponentially (Evans and Aceves 2016; Salganik 2017). Although social scientists have widely recognized that these vast archives contain a cache of cultural information with great potential for analysis, scholars have remained limited by the available tools for integrating and analyzing large-scale text. Neural word embeddings present a method for producing rich models of semantic relationships from corpora too large for techniques such as topic modeling or semantic network

analysis. When adequate text is available, contemporary word embedding approaches can distill detailed and precise semantic information and relationships with a fidelity that exceeds prior methods and approaches human performance (e.g., Devlin et al. 2018).

Cultural dimensions from word embeddings do not simply provide a means of deriving textual measures compatible with intersectionality, affect control, and field theory. The very success of these models provides suggestive validation for relational approaches to cultural theorizing. Furthermore, by operationalizing a high dimensional model of culture, word embedding models allow researchers to extend and contend with theories of culture in novel ways. Embeddings present a much vaster set of potential axes along which individuals and social groups may compete, cooperate, fracture, or coalesce than low-dimension theories of cultural constraint allow. By simultaneously capturing the multiplicity of associations expressed in language, word embedding models are able to represent a complex geometry of culture, which, like a many-faceted crystal, amplifies the subtle and shifting framings that enable coordinated and spontaneous social action.

APPENDIX

Part A: Word Embeddings: History, Varieties, and Implementations

Word embedding models are naïve as to what words signify, lacking intrinsic word referents. They position words relative to one another based purely on how they are used in relation to one another. This process of identifying a word's meaning from context resonates with a tradition of practice-oriented theories of language in which word meanings are always understood through usage (Searle 1969; Wittgenstein 1953). The theory of meaning implicit in word embedding algorithms is well summarized by linguist J. R. Firth's (1957) dictum: "you shall know a word by the company it keeps."

Early approaches to word embedding, including Latent Semantic Analysis (LSA) or Indexing (LSI), have existed since the 1970s (Dumais 2004), but they initially involved the factorization of word-document matrices with singular value decomposition (SVD). Recent breakthroughs in autoencoding neural networks and advances in computational power have enabled a new class of word embedding models (Mikolov, Sutskever, et al. 2013) that can heuristically factorize much larger matrices (Levy and Goldberg 2014). This allows them to incorporate information about local semantic contexts from surrounding word windows rather than the entire documents. This one change, shifting from global to local context, has resulted in a punctuated increase in their accuracy on a wide range of tasks, from the analogy tests we detail to word classification (Taddy 2015b), question-answering (Zhou et al. 2015), and automated translation (Johnson et al. 2017). As a result, contemporary neural word embedding models distil an encyclopedic breadth of subtle and complex cultural associations from large collections of text by training the embedding model with local word associations a human might learn through ambient exposure to the same collection of language (Nagy, Herman, and Anderson 1985).

Word2vec can operate under two distinct model architectures: continuous bag-of-words (CBOW) or skip-gram. Under the CBOW architecture, the corpus is read line-by-line in a sliding window of k words, with k determined by the analyst. Previous studies have found windows of ~8 words produce the most consistent results (Le and Mikolov 2014). For each word in the corpus, the algorithm aims to maximize classification of the center word n, given its surrounding words within a context window of size k. The skip-gram architecture works similarly, except instead of predicting a word with context, it predicts context given a word. Related embedding approaches take into account additional information such as the "global" proximity of words within an overarching document (*GLoVe*, Pennington et al. 2014) or even

sub-word letter sequences in surrounding words (*fastText*, Joulin et al. 2016).[23]

To test the robustness of Google Ngrams and the *word2vec* algorithm, we evaluated our approach across different corpora and word embedding algorithms. Specifically, we compared the results from our survey of cultural associations with two widely used, publicly-available pre-trained embedding models. The first model we use to represent contemporary cultural associations is trained using the *word2vec* algorithm with CBOW architecture on 100 billion words scraped from Google News articles published by U.S. news outlets (Mikolov, Chen, et al. 2013). The second model was produced with the *GloVe* algorithm, which accounts for local and global dependencies, and is trained on a corpus collected as part of the Common Crawl, a broad scraping of millions of webpages (Pennington et al. 2014). A weakness of both of these publicly-available embeddings is that they lack satisfying documentation regarding the exact conditions of inclusion for texts in the corpus. This is unfortunately common among pre-trained embeddings, and it limits their utility for analysis of culture. Nevertheless, we include results from validations on these models to allow comparability with contemporary research in natural language processing and because embeddings greatly benefit from training upon massive quantities of text, and both of these embeddings are exceptional in this regard.

We note that the varieties of word embeddings analyzed and discussed here are all "Euclidean," meaning the space is defined by non-intersecting, parallel dimensions. Embedding in other geometries is possible and may be better suited for modeling semantic patterns other than "cultural dimensions." In a hyperbolic space, infinitely many lines may go through p without intersecting ℓ, and a central node may be close to many peripheral nodes without those nodes being close to each other. Embedding corpora in a hyperbolic geometry makes discovery of the semantic dimensions underlying them less straightforward, but it facilitates modeling semantic hierarchy.[24] Embedding semantic networks in hyperbolic space has facilitated automatic discovery of hypernyms—words with broad meanings under which specific instance words lie—such as the relationship between *animal, rodent*, and *rat*, or *color* and *red, green*, and *blue* (Chamberlain et al. 2017; Handler 2014; Nickel and Kiela 2017; Rei and Briscoe 2014). This might also enable discovery of holonyms and meronyms—words constituting wholes and their parts—like *hand, flesh*, and *fingers*. In this way, Euclidean word embeddings are tuned to capture semantic dimensions, but altering hidden parameters, such as the curvature of the underlying geometry, would allow them to capture other associations, like semantic hierarchy.[25]

Part B: Survey of Cultural Associations

Here we detail the Survey of Cultural Associations we fielded to produce a set of current, human-rated cultural evaluations for comparison against results from word embedding models. The survey was fielded through Amazon Mechanical Turk, an online service through which "requesters" can post a task and workers find and select tasks to complete in exchange for monetary compensation. Our survey was listed as "Sociological Survey," with the description "a fifteen-minute survey of cultural associations" and compensation of $1.75. The task was only available to Mechanical Turk workers located in the United States. The survey was fielded in two waves, October 2016 and December 2017 to samples of 206 and 200, respectively, of which a total of 398 respondents completed the survey. We pool the two waves in our analyses. Respondents were posed with the task of rating words on three scales, gender (very masculine to very feminine), race (very African American to very white), and class (very upper-class to very working-class). The set of 59 words they rated are listed in Table B1.

A number of previous studies have found that Mechanical Turk surveys fare well when compared to surveys with probability sampling, particularly when researchers measure and account for the sociodemographic

characteristics of the sample (Levay et al. 2016). Although Mechanical Turk's population of workers cannot be said to represent the general U.S. population, it is characterized by considerable diversity along racial, gender, and socioeconomic lines (Huff and Tingley 2015).

To mitigate any bias in estimates due to disproportionate representation of sociodemographic groups in the sample, we use post-stratification weighting to make our sample match the U.S. general population. We took population estimates from the U.S. Census Current Population Survey (CPS) of 2017 as population estimates for weighting our sample. We weighted along three strata: sex, education, and race. Sex is treated as two categories: male and female; education is divided into two categories: bachelor's degree or less than bachelor's degree; and race is divided into three strata: white, African American, or other. Results presented in this article include post-stratification weighting; however, additional analyses available upon request confirm that the inclusion of weights does not substantively alter results. Table B2 displays basic demographic characteristics of the sample.

In Table B3 we provide a more detailed summary of the correspondence between associations produced in Google News word embedding models and those reported by survey respondents, and we examine differences in performance between word domains. For all pairs of words that have a statistically significant difference in mean survey rating ($p < .01$) for class, race, and gender associations within a substantive domain, we calculate the proportion of pairs that are correctly ordered by the word embedding model trained on Google News text. For instance, if "steak" is significantly more upper-class than "hamburger" in the survey, we test if *steak* projects more masculine than *hamburger* in the embedding, and we then calculate the percentage of all such pairs of words that are correctly matched.

Table B3 shows that within most substantive domains, the rate of correct classification is above 80 percent and in many cases above 90 percent. It is also clear that embedding does a better job in domains with stronger cultural associations. For instance, there is very little difference in racial association between the clothing items included in the survey (standard deviation of 4.68), and in this domain the embedding has a low 55.0 percent rate of matching the survey. In first names, however, where signals are stronger (standard deviation of 32.46), the same dimension of the word embedding correctly matches 94.7 percent of differences in the survey.

Table B1. List of Words Rated in Cultural Associations Survey

Occupations	Clothing	Sports	Music Genres	Vehicles	Food	First Names
Banker	Blouse	Baseball	Bluegrass	Bicycle	Beer	Aaliyah
Carpenter	Briefcase	Basketball	Hip hop	Limousine	Cheesecake	Amy
Doctor	Dress	Boxing	Jazz	Minivan	Hamburger	Connor
Engineer	Necklace	Golf	Opera	Motorcycle	Pastry	Jake
Hairdresser	Pants	Hockey	Punk	Skateboard	Salad	Jamal
Journalist	Shirt	Soccer	Rap	SUV	Steak	Molly
Lawyer	Shorts	Softball	Techno	Truck	Wine	Shanice[a]
Nanny	Socks	Tennis				Tyrone
Nurse	Suit	Volleyball				
Plumber	Tuxedo					
Scientist						

[a]Word did not appear frequently enough in the 2000 to 2012 Google Ngrams to appear in the embedding model and is therefore excluded from 2000 to 2012 Google Ngrams analyses.

Table B2. Descriptive Statistics for Mechanical Turk Sample and Census CPS Sample

	Mechanical Turk	Census CPS
Gender (1 = female)	43.47%	51.76%
Education		
High school, GED, or less	12.31%	39.99%
Some college	26.88%	18.83%
Associate's degree	10.05%	9.75%
Bachelor's degree	43.47%	20.03%
Graduate degree	7.29%	11.39%
Race/Ethnicity		
African American	6.53%	12.52%
White	79.15%	78.22%
Other	14.32%	9.26%
Hispanic	9.82%	15.92%
Age (mean)	34.40	47.20
N	398	135,137

Table B3. Percentage of Statistically Significant ($p < .01$) Survey Differences Correctly Classified in Google News Word Embedding Model

	Sports	Food	Music	Occupations	Vehicles	Clothes	Names	All Domains
Gender	87.9%	88.2%	72.2%	93.6%	82.4%	74.4%	95.2%	84.8%
Class	96.3%	93.8%	88.9%	60.9%	94.1%	90.0%	77.3%	75.3%
Race	90.0%	68.8%	100%	51.5%	87.5%	55.0%	94.7%	69.1%

Part C: Statistical Significance of Distances and Associations

We propose well-established nonparametric bootstrapping and subsampling methods to show the stability and significance of word associations within our embedding model. This approach allows us to establish conservative confidence or credible intervals for both (a) distances between words in a model and (b) projections of words onto an induced dimension (e.g., *affluence-poverty*). If we assume the texts underlying our word embedding model are observations drawn from an independent and identically distributed (i.i.d.) population of cultural observations, then bootstrapping allows us to estimate the variance of word distances and projections by measuring those properties through sampling the empirical distribution of texts with replacement (Efron 2003; Efron and Tibshirani 1994).

To estimate bootstrapped 90 percent confidence intervals, the analyst draws documents with replacement from the corpus to construct 20 new corpora, each the size of the original corpus. The analyst then estimates either word similarities or angles between vectors on all 20 of these new corpora. The 2nd order (2nd smallest) estimated statistic $s_{(2)}$ is taken as the confidence interval's lower bound and the 19th order statistic $s_{(19)}$ as its upper bound. The distance between $s_{(2)}$ and $s_{(19)}$ across 20 bootstrap samples span the 5th to the 95th percentiles of the statistic's variance, bounding the 90th confidence interval. A 95 percent confidence interval would span $s_{(2)}$ and $s_{(39)}$ in word embedding distances or projections estimated on 40 bootstrap samples of a corpus, tracing the 2.5th to 97.5th percentiles. Due to the limits of corpus size, we use this bootstrapping approach to conduct statistical significance tests for our JSTOR models.

If the corpus is very large, however, we may take a subsampling approach, which randomly partitions the corpus into non-overlapping samples, then estimates the word

embedding models on these subsets and calculates confidence or credible intervals as a function of the empirical distribution of distance or projection statistics and number of texts in the subsample (Politis, Romano, and Wolf 1997). Subsampling relies on the same i.i.d. assumption as the bootstrap (Politis and Romano 1992, 1994). For 90 percent confidence intervals, we randomly partition the corpus into 20 subcorpora, then calculate the error of our embedding distance or projection statistic s for each subsample k as $B^k = \sqrt{\tau_\kappa}(s^k - \overline{s})$, where τ_κ is the number of texts in subsample k, s^k is the embedding distance or projection for the k_{th} sample, and $\overline{s}$ is the mean of the 20 estimates. The 90 percent confidence interval spans the 5th to 95th percentile variances, inscribed by $\overline{s} - \frac{B^K_{(19)}}{\sqrt{\tau}}$ and $\overline{s} - \frac{B^K_{(2)}}{\sqrt{\tau}}$ where τ is the number of texts in the total corpus. As with bootstrapping, a 95 percent confidence interval would require 40 subsamples; a 99 percent confidence would require 200 (.5th to 99.5th percentiles). We use this subsampling approach to construct confidence intervals for our Google Ngrams models.

A great benefit of bootstrapped and subsampled confidence intervals is that they reflect how robust an association is across texts. If a word occurs only rarely or is used in a diffuse set of very distinct contexts, the word's position in the vector space will be radically different between subsamples and therefore will produce larger confidence or credible intervals. On the other hand, words that are frequently used in consistent contexts will hold more stable positions across the subsamples and hence produce smaller confidence or credible intervals.

Part D: Word Pair Lists

Table D1. Word Pairs Used to Construct Affluence, Gender, and Race Dimensions for Amazon Mechanical Turk Survey Validation

Affluence		Gender	Race
rich-poor	precious-cheap	man-woman	black-white
richer-poorer	priceless-worthless	men-women	blacks-whites
richest-poorest	privileged-underprivileged	he-she	Black-White
affluence-poverty		him-her	Blacks-Whites
affluent-destitute	propertied-bankrupt	his-her	African-European
advantaged-needy	prosperous-unprosperous	his-hers	African-Caucasian
wealthy-impoverished	developed-underdeveloped	boy-girl	Afro-Anglo
costly-economical		boys-girls	
exorbitant-impecunious	solvency-insolvency	male-female	
expensive-inexpensive	successful-unsuccessful	masculine-feminine	
exquisite-ruined	sumptuous-plain		
extravagant-necessitous	swanky-basic		
flush-skint	thriving-disadvantaged		
invaluable-cheap	upscale-squalid		
lavish-economical	valuable-valueless		
luxuriant-penurious	classy-beggarly		
luxurious-threadbare	ritzy-ramshackle		
luxury-cheap	opulence-indigence		
moneyed-unmonied	solvent-insolvent		
opulent-indigent	moneyed-moneyless		
plush-threadbare	rich-penniless		
luxuriant-penurious	affluence-penury		
	posh-plain		
	opulence-indigence		

Table D2. Word Pairs Used to Reconstruct 20 Semantic Differential Dimensions from Jenkins and Colleagues (1958) for Historical Survey Validation

soft-hard	**foolish-wise**	**unimportant-important**	**fast-slow**
supple-tough	dumb-smart	inconsequential-consequential	quick-lagging
delicate-dense	irrational-rational	secondary-principal	rapid-unhurried
pliable-rigid	stupid-thoughtful	irrelevant-major	speedy-sluggish
fluffy-firm	unwise-sensible	trivial-crucial	swift-gradual
mushy-solid	silly-reasonable	negligible-critical	quickly-slowly
softer-harder	ridiculous-enlightened	insignificant-significant	swiftly-gradually
softest-hardest	unintelligent-intelligent	unnecessary-essential	faster-slower
		peripheral-central	fastest-slowest
unusual-usual	**excitable-calm**	**strong-weak**	**colorful-colorless**
different-customary	volatile-tranquil	powerful-powerless	brilliant-uncolored
abnormal-normal	nervous-still	muscular-frail	bright-pale
irregular-regular	tempestuous-serene	brawny-feeble	radiant-drab
odd-standard	fiery-peaceful	strapping-puny	vivid-pallid
atypical-typical	emotional-restful	sturdy-fragile	vibrant-lackluster
unexpected-expected	jumpy-sedate	robust-flimsy	colored-bleached
unconventional-conventional	unsettled-settled	vigorous-languid	
rounded-angular	**passive-active**	**true-false**	**ugly-beautiful**
circular-cornered	immobile-mobile	true-untrue	unattractive-attractive
round-pointed	lethargic-energetic	verifiable-erroneous	unsightly-pretty
dull-sharp	frail-vital	veracious-fallacious	hideous-handsome
smooth-jagged	subdued-vigorous	accurate-inaccurate	grotesque-gorgeous
spherical-edged	static-dynamic	faithful-fraudulent	repulsive-cute
	subdued-lively	correct-incorrect	
feminine-masculine	**bad-good**	**successful-unsuccessful**	**old-new**
woman-man	worst-best	victorious-failed	aged-recent
women-men	deficient-fine	triumphant-abortive	ancient-contemporary
she-he	inferior-superior	winning-losing	decrepit-fresh
her-him	unsatisfactory-satisfactory	thriving-failing	elderly-young
her-his	unacceptable-acceptable	fruitful-fruitless	historic-modern
hers-his	awful-excellent	prosperous-ineffectual	adult-child
girl-boy	terrible-superb	success-failure	older-newer
girls-boys	dreadful-outstanding	win-lose	oldest-newest
female-male	unexceptional-exceptional		
kind-cruel	**straight-curved**	**timely-untimely**	**tasteless-savory**
tender-callous	linear-nonlinear	punctual-late	bland-tasty
compassionate-heartless	unswerving-swerving	ready-unready	flavorless-flavorful
humane-inhumane	unbending-bent	prompt-delayed	unappetizing-delectable
merciful-merciless	untwisted-twisted	reliable-unreliable	mild-piquant
gentle-brutal	direct-meandering	early-late	insipid-succulent
nice-unpleasant	undeviating-serpentine	earlier-later	dull-delicious
kindest-cruelest	straighter-curvier	earliest-latest	blandest-tastiest

Note: Terms used by Jenkins and colleagues to specify dimensions are in bold.

Table D3. Word Pairs Used to Construct Class Dimensions (Along with Affluence and Gender in Table D1)

Cultivation	Employment	Education	Status	Morality
cultivated-uncultivated	employer-employee	educated-uneducated	prestigious-unprestigious	good-evil
cultured-uncultured	employers-employees	learned-unlearned	honorable-dishonorable	moral-immoral
civilized-uncivilized	owner-worker	knowledgeable-ignorant	esteemed-lowly	good-bad
courteous-discourteous	owners-worker	trained-untrained	influential-uninfluential	honest-dishonest
proper-improper	industrialist-laborer	taught-untaught	reputable-disreputable	virtuous-sinful
polite-rude	industrialists-laborers	literate-illiterate	distinguished-commonplace	virtue-vice
cordial-uncordial	proprietor-employee	schooled-unschooled	eminent-mundane	righteous-wicked
formal-informal	proprietors-employees	tutored-untutored	illustrious-humble	chaste-transgressive
courtly-uncourtly	capitalist-proletarian	lettered-unlettered	renowned-prosaic	principled-unprincipled
urbane-boorish	capitalists-proletariat		acclaimed-modest	unquestionable-questionable
polished-unpolished	manager-staff		dignitary-commoner	noble-nefarious
refined-unrefined	managers-staff		venerable-unpretentious	uncorrupt-corrupt
civility-incivility	director-employee		exalted-ordinary	scrupulous-unscrupulous
civil-uncivil	directors-employees		estimable-lowly	altruistic-selfish
urbanity-boorishness	boss-worker		prominent-common	chivalrous-knavish
politesse-rudeness	bosses-workers			honest-crooked
edified-loutish	foreman-laborer			commendable-reprehensible
mannerly-unmannerly	foremen-laborers			pure-impure
polished-gruff	supervisor-staff			dignified-undignified
gracious-ungracious	superintendent-staff			holy-unholy
obliging-unobliging				valiant-fiendish
cultured-uncultured				upstanding-villainous
genteel-ungenteel				guiltless-guilty
mannered-unmannered				decent-indecent
polite-blunt				chaste-unsavory
				righteous-odious
				ethical-unethical

Part E: Constructing Cultural Dimensions

Here we perform tests to reveal design principles for the construction of cultural dimensions. We begin by considering the set of antonym pairs needed to effectively approximate a cultural dimension. The English language contains a vast vocabulary for denoting affluence and poverty, and drawing on five thesauri, we assembled a list of 42 pairs of terms that closely correspond to this cultural dimension. The very selection of antonym pairs presents two methodological difficulties. First, it is not clear how many antonym pairs are required to approximate the cultural dimension of interest. Second, terms do not always have a single obvious antonym, so constructing pairs requires subjective judgment on the part of the researcher. We investigate both of these issues in our validation of the affluence dimension.

First, we test if using a greater number of antonym pairs in constructing a cultural dimension is associated with improvements in correlation between projections on that dimension and human-rated associations from survey data. To accomplish this, we found the average correlation between surveyed class association and projection on an affluence dimension constructed with a single antonym pair, two antonym pairs, three antonym pairs, through all 42 pairs. Results are presented in Figure E1.[26] Cultural dimensions constructed from single antonym pairs fare relatively poorly, with their projections correlating on average at .2 with surveyed response. Correlations with survey response rise as a greater number of antonym pairs are used to construct the cultural dimension, but the gains in correlation from adding additional antonym pairs shrinks. In the following analyses, we use the full 42 antonym pairs for our affluence dimensions to improve robustness and decrease chance variability.

Next we test the extent to which the precise pairing of words affects correlation with survey data. To do this, we take our sets of 42 "rich" synonyms and 42 "poor" synonyms, and we re-pair them in random permutations. For instance, "rich" may be paired with "impoverished" instead of "poor." We then construct the affluence dimension using this set of randomly paired, roughly antonym terms, and we correlate its projections with the survey data. On average, it performs only marginally worse than our curated pairs of antonyms.

Finally, to eliminate the element of analyst judgment in pairing, we try a third strategy of averaging all "rich" synonyms together and subtracting the average of all the "poor" synonyms, an approach we label the "grouped pairs" method. The result of this operation is very similar to the one we propose, but it is mathematically distinct because it involves the averaging of vectors *before* performing the nonlinear operation of cosine similarity. Once again, we find substantively similar results, as displayed in Figure E1.

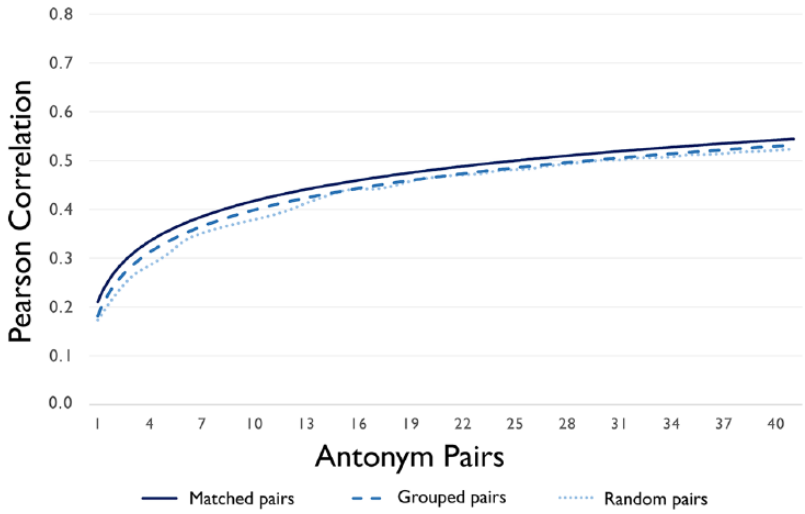

Figure E1. Average Correlation between Survey-Rated Class Associations and Word Embedding Projections on the Cultural Dimension of Affluence, Constructed with 1 to 42 Antonym Pairs; Google Ngrams 2000 to 2012 Word Embedding, Smoothed to Clarify Trend

Part F: Corpus Sizes

Table F1. Sizes of Google Ngrams and JSTOR Corpora by Decade

Decade	Google Ngram Word Count	JSTOR Word Count	JSTOR Article Count
1900s	3.0×10^{10}	4.7×10^{7}	1,294
1910s	2.9×10^{10}	5.7×10^{7}	2,020
1920s	2.4×10^{10}	8.4×10^{7}	3,266
1930s	2.1×10^{10}	1.1×10^{8}	4,228
1940s	2.2×10^{10}	1.6×10^{8}	5,923
1950s	2.9×10^{10}	2.2×10^{8}	7,442
1960s	5.0×10^{10}	3.5×10^{8}	10,152
1970s	5.9×10^{10}	6.8×10^{8}	17,855
1980s	7.2×10^{10}	8.9×10^{8}	19,830
1990s	1.2×10^{11}	1.2×10^{9}	20,698
2000–12	2.5×10^{11}		

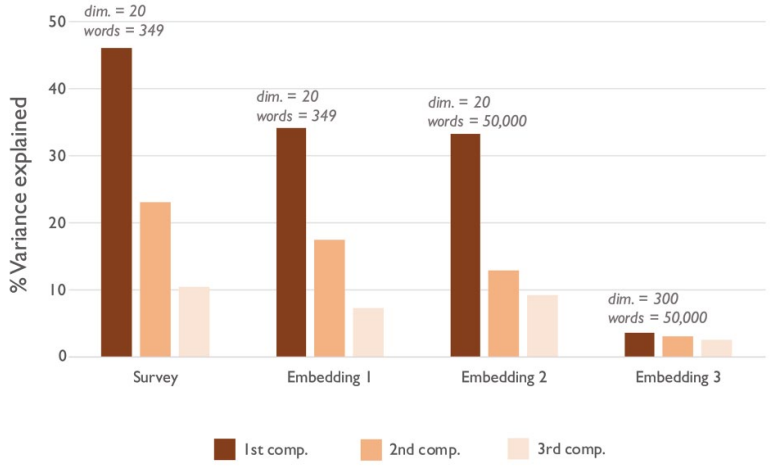

Figure G1. Variance Explained in Principal Components Analysis of 1958 Semantic Differential Survey Data and from Three Datasets of Projections from the 1950s Google Ngrams Embedding

Part G: Semantic Differential Validation

We draw on Jenkins and colleagues' (1958) semantic differential dataset again to conduct a close comparison between the semantic spaces produced by word embeddings and those produced by the semantic differential method. A major finding of Osgood and colleagues' (1957) research program was that the data produced with the semantic differential method could be reduced to relatively low dimensionality with only modest loss of information. They consistently found that when they subjected the matrix of word associations to principal components analysis (PCA), the first three components captured upward of 70 percent of the total variance of the semantic spaces. We validate this finding but show that the same kind of successful dimension reduction is not possible across all dimensions with word embedding models, suggesting the importance of higher dimensionality in analyzing culture. Results are presented in Figure G1.

First, we conduct PCA on Jenkins and colleagues' 1958 dataset of human-rated associations. As anticipated by semantic differential theory, the great majority of the variance of the 20-dimensional space is explained by the first three components. Second, we construct an embedding-derived dataset that mirrors the dataset of human ratings by projecting the same set of 349 terms onto 20 cultural dimensions corresponding to those measured by Jenkins and colleagues (1958) (see Table D2). Conducting PCA on this embedding-derived dataset, we find a comparably high percent of the total variance is explained by the first three principal components. Third, we expand the embedding-derived dataset from the set of 349 words used by Jenkins and colleagues to the set of 50,000 most commonly used words in the 1950s Ngrams corpus, while restricting to the same 20 semantic dimensions specified by Jenkins and colleagues. Again, most of the variance is explained by the first three dimensions. Finding that the projections of 50,000 common terms can similarly be reduced to low dimensionality suggests the ability to compress the space to three dimensions does not result from the particular set of terms rated.

Finally, we perform PCA on the full, 300 dimensional *word2vec* output model for the 50,000 most common words. Figure G1 shows the first component explains only 3.4 percent of the variance in the entire vector space. The stark difference in results between this and the previous analyses suggests the information in semantic spaces produced by word

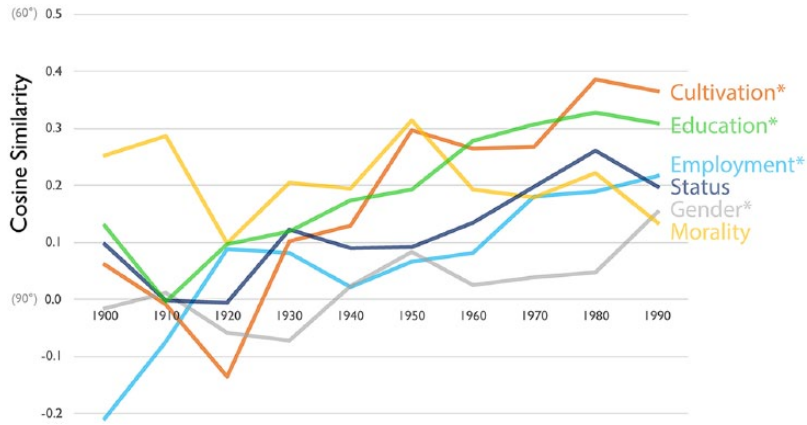

Figure H1. Cosine Similarity of the Affluence Dimension with Six Other Class Dimensions in Sociology Texts; 1900 to 1999 JSTOR Sociology Corpus
Note: Asterisks represent statistically significant difference between angles in the 1900s and 1990s ($p < .10$).

embeddings is diffusely spread across its many dimensions. This finding suggests these additional dimensions of the word embedding provide information not contained in the semantic differential spaces.[27] These additional dimensions likely make it possible to capture subtle or unusual semantic dimensions, such as *American-French* or *city-state* (Mikolov, Chen, et al. 2013), which would be missed by the standard semantic differential approach.

Part H: Embedding Sociological Discourse

We contextualize our cultural analysis of class by comparing associations held in the general public to those expressed in sociological literature. Here we display results from word embedding models trained on a corpus of all sociology articles published in the twentieth century in the JSTOR collection. This corpus includes 121 English-language periodicals, ranging from *American Sociological Review* and *Sociological Methods & Research* to *Poetics, Social Problems*, and *Symbolic Interaction*. As with the Google Ngrams, we divide the JSTOR sociology corpus into 10-year windows, training *word2vec* embedding models for each decade of the twentieth century. The class-based associations we find in this corpus generally accord with widely recognized trends in the discipline, so we use this analysis not to produce new discoveries but to allow formal comparison with the embeddings trained on general discourse. Details regarding the size of the JSTOR corpus are available in Appendix Part F.

Figure H1 is analogous to Figure 5 in the main text, but it presents results from the JSTOR corpus rather than the Google Ngrams corpus. As described in the main text, there are several places where sociological understandings of class depart from conventional associations. First, while femininity maintains a persistent association with affluence in general discourse, masculinity becomes identified with affluence in the second half of the twentieth century in sociological texts, evincing the discipline's growing concern for gender inequality. Second, within sociology, the latter half of the twentieth century witnesses a heightened association between the affluence and employment dimensions, with owners and bosses becoming increasingly marked as wealthy relative to workers and staff. This strong association between employment and affluence at the end of the century contrasts with the middling association found in the Ngrams embeddings, and it may reflect sociology's focus on structural sources of stratification and the influence of Marxian thought. Finally, cultivation only emerges as a strong

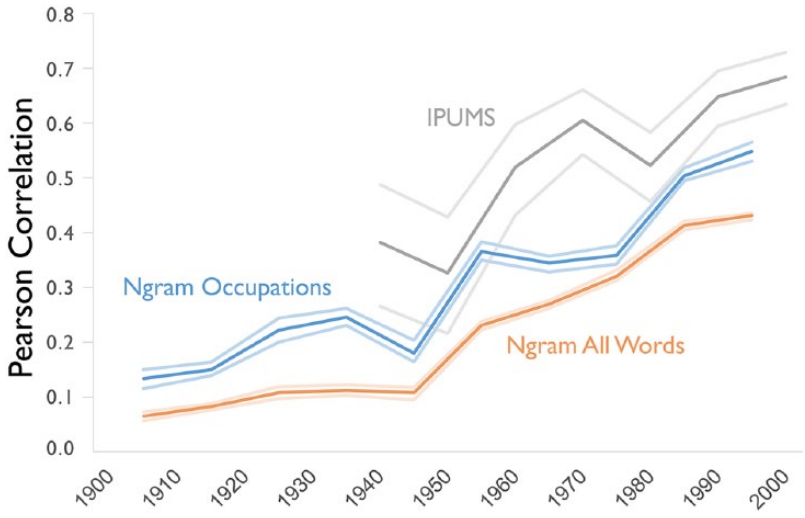

Figure I1. Correlations of Affluence and Education from IPUMS Surveys and Google Ngrams Text

Note: Correlation of occupations' average income and average education by decade; correlation of occupation names' projections on affluence and education dimensions; and correlation of all words' projections on affluence and education dimensions.

marker of affluence at the end of the century in sociological texts, whereas the two dimensions display a persistent association in general discourse. This tightening relationship coincides with the discipline's growing awareness of how self-presentation and cultural capital fuel class reproduction.

Part I: Cultural versus Material Changes in Education

The most striking change in the transformation of class associations revealed in Figures 5 and 6 occurs between the dimensions of affluence and education. Their association is weakly positive at the dawn of the twentieth century, but it becomes visibly stronger in the second half of the century. Figure I1 shines light on the growing semantic connection between affluence and education by displaying multiple indicators of this relationship over time, drawing on word embedding projections and their complex relationship with census data from IPUMS (Ruggles et al. 2019).

First, for all occupations reported in the census within a given decade, we calculate the correlation between the average reported income and average years of formal education for all individuals from that occupation in the IPUMS data. We observe a distinct spike in the correlation between an occupation's average income and education level between 1950 and 2000. Unfortunately, the census did not collect data on education level prior to 1940, but extensive historical records confirm that, while education has long shown returns to income, it played a substantially smaller role in shaping social stratification in the beginning of the twentieth century (Collins 1979; Goldin and Katz 1999).

To compare the material trend to the cultural trend, we take the names of all occupations reported by the census in a given decade and project them on both the affluence and education dimensions of the word embedding trained on the respective decade of Google Ngrams text. We then calculate the correlation between occupations' projections on education and affluence for each decade. We find an upward trend in semantic association between education and affluence among the occupations beginning in the 1950s that mirrors the socioeconomic trend. Finally, we correlate the projections on the education and

affluence dimensions of the 50,000 most common words in each decade. We find that the correlation among all words follows a parallel trend, exhibiting a steady increase that begins mid-century. Its overall levels of correlation are consistently lower than when the vocabulary is limited to occupations. These findings suggest a growing correspondence between the cultural associations of education and affluence coincided with material shifts that drew these dimensions of class materially together in the economy. Furthermore, while this semantic convergence is particularly apparent in the domain of occupations, it is also evident to a lesser extent in the general lexicon.

Editors' Note

Figure 3, Figure 10 and Table B3 were incorrect in the Online First version. They have now been corrected online and in print, along with two related values in the text.

Acknowledgments

We thank John Levi Martin, Etienne Ollion, and Marion Fourcade for their helpful comments. An earlier version of this paper was presented at the 2017 American Sociological Association Annual Meeting in Montreal, QC.

ORCID iDs

Austin C. Kozlowski https://orcid.org/0000-0001-8458-1129
James A. Evans https://orcid.org/0000-0001-9838-0707

Notes

1. Word embedding models are sometimes considered "low dimensional" relative to the number of words used in text (e.g., 50,000) because they reduce this *very* high dimensional word space. Nevertheless, considered from the perspective of one-, two-, or three-dimensional models common in the analysis of culture, these spaces are much more complex and reproduce much more accurate cultural associations, as we will show.
2. The cultural dimensions of race are also strongly linked to collective understandings of class. However, historical analyses of race with word embeddings present methodological challenges that require a unique and careful treatment that is beyond the scope of this investigation, as we will detail.
3. Such networks could be made dense and their links weighted, encoding a myriad of word collocations, but analysis of the resulting hairball would require a calculus that deviates widely from standard network analysis, such as one based on random walks (Rosvall and Bergstrom 2008; Shi, Foster, and Evans 2015) or simulated flows over implied curvature (Jost and Liu 2014).
4. Scientists have attempted to perform these parametrically, as with exponential family embedding models, but their performance has not yet approached that of autoencoders (Rudolph et al. 2016).
5. The surface area of a unit circle surpasses its volume in three dimensions. As a hypersphere's dimension approaches infinity, its volume approaches zero.
6. A shallow, two-layer neural network word embedding like *word2vec* constrains semantic dimensions to be linear as they are in PCA or SVD. Deeper neural network embedding models allow estimation of nonlinear semantic dimensions (Devlin et al. 2018).
7. Multiple word embedding approaches have become widespread in recent years, but the analyses we present here primarily utilize the skip-gram models in *word2vec*, cross-validated with those from *GloVe* (Pennington et al. 2014). The methodological principles outlined here, however, reach beyond neural-network autoencoders and are generally applicable to word embedding models constructed with other algorithms, including Latent Semantic Analysis based on SVD (Dumais 2004) and Bayesian nonparametric estimation (Rudolph et al. 2016).
8. We find that this calculus produces nearly identical results to a similar approach of first averaging the words on each side of the semantic dimension and then taking the difference between the two averages.
9. Because the cultural category of race is itself multidimensional, its representation in word embeddings is multidimensional as well. We restrict our analyses to the *black-white* dimension, but other word pairs, such as *hispanic-white* or *hispanic-black*, similarly capture meaningful semantic relations.
10. The signs may be flipped, of course, making positive values reflect low-class associations if poverty terms are subtracted from affluence terms, that is, by using *poverty – affluence* instead of *affluence – poverty*.
11. We know this from the Gaussian Annulus theorem, that two random points from a *d*-dimensional Gaussian with unit variance in each direction are approximately orthogonal (Blum, Hopcroft, and Kannan 2016).
12. "Bias" in this literature refers to harmful negative stereotypes, not the statistical definition of the term.
13. In a few instances, Bourdieu (1984:266, 343) notes other dimensions that structure the social topography, such as upward or downward trajectory and a preference for the traditional versus the innovative, but these dimensions have not enjoyed the same systematic treatment or theoretical elaboration as economic and cultural capital.
14. Code used in this analysis is available at: https://github.com/KnowledgeLab/GeometryofCulture.
15. First names were sampled from lists of names found to be most predictive of belonging to an African American person and most predictive of belonging

to a non-Hispanic white person for each sex from data of all children born in California from 1961 to 2000 (Fryer and Levitt 2004). Terms in other domains were selected based on known race, class, and gender markers that have been examined in previous literature.

16. In preliminary analyses, we trained word embeddings on collections of both 5-grams and 4-grams, but we found they performed poorer on survey validation than models trained on 5-grams alone. Because all information in word embedding models comes from neighboring words, it is unsurprising that smaller context windows produce weaker models.
17. We note that an analyst could discover differential weights for each word pair by estimating them with exploratory factor analysis or within a linear model that predicts surveyed cultural associations. A weighted sum necessarily improves the correlation of our dimension with surveyed associations, but it would be fragile for analysis of historical culture where the weights likely change, and we can field no surveys in the past.
18. Each of these cultural dimensions refers to a specific vector within an embedding model. Nevertheless, we reserve vector notation for individual word vectors, indexed by the precise word under the vector symbol (e.g., $\overrightarrow{man}$ refers to the vector associated with the word "man" in the embedding). For consistency with theoretical discussions earlier and later in the text, we do not use vector notation but assume it for vectors such as affluence, status, and cultivation, which comprise the average of the difference between many specific word vectors (see Tables D1 and D3).
19. Substantively similar results are produced when we correlate projections of all words on the two dimensions instead of calculating the angle between the dimensions. This correlational approach more closely derives from our validations, but we chose to display angles between dimensions to underscore the geometric rendering of cultural meaning inherent to word embedding models.
20. We acknowledge that because our embedding space was constructed with a single optimization algorithm, word projections on one semantic dimension are not independent from those on another. Nevertheless, as we show in Figure 1, there are many dimensions and degrees of freedom that limit the influence of this singular dependence even for semantically proximate associations. Moreover, because we do not seek to generalize beyond the texts within our substantial sample, we do not violate the assumptions of the OLS framework, which allows us to directly ask the degree to which word shifts in the projection along one dimension are a function of their position on another, holding constant their position on a third.
21. Complex neural network models are not statistical objects, in that their heuristic methods of optimization cannot (yet) be characterized by a σ-algebra, which details the full range of parameters searched on the path to the final, fitted model. This means fitted models, including those in this article, lack proof that they are the best models of their kind, despite successful performance on language and culture tasks.
22. Early in this project, we attempted the quixotic feat of inductively identifying the "most important" semantic dimension in word embedding space. To accomplish this, we collected every pair of antonyms in English from the digital dictionary WordNet (Miller 1995) and calculated the total variance in the vector space explained by each pair. Our analysis revealed one dimension that explained more semantic variance across the entire lexicon than any other: *steroidal–nonsteroidal*. After feeling like the crew from Douglas Adams's *The Hitchhiker's Guide to the Galaxy* when they find that the "Answer to the Ultimate Question of Life, the Universe, and Everything" computed by the supercomputer Deep Thought over 7.5 million years is "42" (Adams, Brett, and Perkings 1978), we caution against "theory-free" approaches to meaning discovery.
23. Although we do not use them in this analysis, the m contexts by k dimensions matrix in Figure 1 also retains a great deal of semantic information and has been used in concert with word embeddings to identify words that are complements versus substitutes in text (Nalisnick et al. 2016; Ruiz, Athey, and Blei 2017).
24. Words have highly unequal frequencies, with some central to common language and others peripheral (Zipf 1932).
25. Embedding networks in hyperbolic geometry typically requires fewer dimensions than in Euclidean space, both because the space may better reflect the intrinsic geometry of the data, and because there is "more space" in a d-dimensional Poincare ball, where the volume rises exponentially relative to the surface area, than in a Euclidean hypersphere of the same dimension.
26. Calculating the average correlation for dimensions constructed from all possible combinations of antonym pairs, although possible, is computationally impractical. For example, there are more than 500 billion ways to select 21 of the 42 antonym pairs. Instead, we sample 400 randomly selected combinations of pairs and calculate the average correlation in the sample for each number of antonym pairs.
27. It is possible that the variation on many of the dimensions of word embeddings is composed of noise and lacks meaningful semantic information. However, prior studies that have attempted to train embedding models with fewer than 200 dimensions display substantially lower performance in semantic benchmarking tasks (Mikolov, Chen, et al. 2013). Conclusively determining how many dimensions are required to faithfully reproduce a system of cultural associations is beyond our scope, but the evidence we provide suggests it is much greater than three.

References

Accominotti, Fabien, Shamus R. Khan, and Adam Storer. 2018. "How Cultural Capital Emerged in Gilded Age America: Musical Purification and Cross-Class Inclusion at the New York Philharmonic." *American Journal of Sociology* 123(6):1743–83.

Adams, Douglas, S. Brett, and G. Perkings. 1978. *The Hitchhiker's Guide to the Galaxy*. London, UK: BBC Radio 4.

Bartlett's Roget's Thesaurus. 1996. Edited by A. Grometstein, P. B. Hansen, K. W. McManus, R. G. Pustell, S. W. Reinecke, and J. C. Ritchie. Boston, MA: Little, Brown, and Company.

Bearman, Peter S., and Katherine Stovel. 2000. "Becoming a Nazi: A Model for Narrative Networks." *Poetics* 27(2):69–90.

Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM* 55(4):77–84.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blum, Avrim, John Hopcroft, and Ravindran Kannan. 2016. "Foundations of Data Science." *Vorabversion Eines Lehrbuchs*.

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings." Pp. 4349–57 in *Advances in Neural Information Processing Systems*.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.

Bourdieu, Pierre. 1989. "Social Space and Symbolic Power." *Sociological Theory* 7(1):14–25.

Bourgois, Philippe. 2003. *In Search of Respect: Selling Crack in El Barrio*. Cambridge, UK: Cambridge University Press.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics Derived Automatically from Language Corpora Contain Human-Like Biases." *Science* 356(6334):183–6.

Carley, Kathleen. 1994. "Extracting Culture through Textual Analysis." *Poetics* 22(4):291–312.

Cha, Youngjoo, and Kim A. Weeden. 2014. "Overwork and the Slow Convergence in the Gender Gap in Wages." *American Sociological Review* 79(3):457–84.

Chamberlain, Benjamin Paul, James Clough, and Marc Peter Deisenroth. 2017. "Neural Embeddings of Graphs in Hyperbolic Space" (arXiv:1705.10359).

Chan, Tak Wing, and John H. Goldthorpe. 2004. "Is There a Status Order in Contemporary British Society? Evidence from the Occupational Structure of Friendship." *European Sociological Review* 20(5):383–401.

Chan, Tak Wing, and John H. Goldthorpe. 2007. "Class and Status: The Conceptual Distinction and Its Empirical Relevance." *American Sociological Review* 72(4):512–32.

Clark, Terry N. 2018. *The New Political Culture*. New York: Routledge.

Cohen, Lizabeth. 2003. *A Consumers' Republic: The Politics of Mass Consumption in Postwar America*. New York: Knopf.

Collins, Randall. 1979. *The Credential Society: An Historical Sociology of Education and Stratification*. New York: Academic Press.

Corman, Steven R., Timothy Kuhn, Robert D. McPhee, and Kevin J. Dooley. 2002. "Studying Complex Discursive Systems: Centering Resonance Analysis of Communication." *Human Communication Research* 28(2):157–206.

Crenshaw, Kimberle. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." *Stanford Law Review* 43(6):1241–99.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (arXiv:1810.04805).

DiMaggio, Paul. 1982. "Cultural Capital and School Success: The Impact of Status Culture Participation on the Grades of U.S. High School Students." *American Sociological Review* 47(2):189–201.

DiMaggio, Paul. 1997. "Culture and Cognition." *Annual Review of Sociology* 23(1):263–87.

DiMaggio, Paul. 2011. "Cultural Networks." Pp. 286–310 in *Sage Handbook of Social Network Analysis*, edited by J. Scott and P. J. Carrington. Thousand Oaks, CA: Sage Publications (http://dx.doi.org/10.4135/9781446294413.n20).

DiMaggio, Paul, and John Mohr. 1985. "Cultural Capital, Educational Attainment, and Marital Selection." *American Journal of Sociology* 90(6):1231–61.

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.

Douglas, Mary. 1966. *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. New York: Routledge.

Dumais, Susan T. 2004. "Latent Semantic Analysis." *Annual Review of Information Science and Technology* 38(1):188–230.

Efron, Bradley. 2003. "Second Thoughts on the Bootstrap." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics* 18(2):135–40.

Efron, Bradley, and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. London, UK: CRC Press.

Elias, Norbert. 1978. *The History of Manners,* Vol. 1, *The Civilizing Process*. New York: Pantheon.

Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103(2):281–317.

Evans, James A., and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42:21–50.

Firth, John R. 1957. "A Synopsis of Linguistic Theory, 1930–1955." *Studies in Linguistic Analysis*. Oxford, UK: Blackwell.

Fischer, Claude S., and Michael Hout. 2006. *Century of Difference: How America Changed in the Last One Hundred Years*. New York: Russell Sage Foundation.

Fourcade, Marion. 2011. "Cents and Sensibility: Economic Valuation and the Nature of 'Nature.'" *American Journal of Sociology* 116(6):1721–77.

Fourcade, Marion, and Kieran Healy. 2007. "Moral Views of Market Society." *Annual Review of Sociology* 33(1):285–311.

Franzosi, Roberto. 2004. *From Words to Numbers: Narrative, Data, and Social Science*. Cambridge, UK: Cambridge University Press.

Freeland, Robert E., and Jesse Hoey. 2018. "The Structure of Deference: Modeling Occupational Status Using Affect Control Theory." *American Sociological Review* 83(2):243–77.

Fryer, Roland G., and Steven D. Levitt. 2004. "The Causes and Consequences of Distinctively Black Names." *Quarterly Journal of Economics* 119(3):767–805.

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):E3635–44.

Gilman, Nils. 1999. "Thorstein Veblen's Neglected Feminism." *Journal of Economic Issues* 33(3):689–711.

Glaser, Barney, and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago, IL: Aldine.

Goldin, Claudia, and Lawrence F. Katz. 1999. "Human Capital and Social Capital: The Rise of Secondary Schooling in America, 1910–1940." *Journal of Interdisciplinary History* 29(4):683–723.

Gramsci, Antonio. 1992. *Prison Notebooks*. New York: Columbia University Press.

Greenacre, Michael. 2017. "Ordination with Any Dissimilarity Measure: A Weighted Euclidean Solution." *Ecology* 98(9):2293–300.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74(6):1464–80.

Hamilton, William L., Jure Leskovec, and Dan Jurafsky. 2016. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change" (arXiv:1605.09096).

Handler, Abram. 2014. "An Empirical Study of Semantic Similarity in WordNet and Word2Vec." PhD dissertation, University of New Orleans, New Orleans, LA.

Heise, David R. 1979. *Understanding Events: Affect and the Construction of Social Action*. Cambridge, UK: Cambridge University Press Archive.

Heise, David R. 1987. "Affect Control Theory: Concepts and Model." *Journal of Mathematical Sociology* 13(1–2):1–33.

Hill, Felix, Kyunghyun Cho, Sebastien Jean, Coline Devin, and Yoshua Bengio. 2014. "Not All Neural Embeddings Are Born Equal" (arXiv:1410.0718).

Hochschild, Arlie Russell. 2012. *The Managed Heart: Commercialization of Human Feeling*. Berkeley: University of California Press.

Hoffman, Mark Anthony, Jean-Philippe Cointet, Philipp Brandt, Newton Key, and Peter Bearman. 2017. "The (Protestant) Bible, the (Printed) Sermon, and the Word(s): The Semantic Structure of the Conformist and Dissenting Bible, 1660–1780." *Poetics* 68:89–103.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38:379–400.

Huff, Connor, and Dustin Tingley. 2015. "'Who Are These People?' Evaluating the Demographic Characteristics and Political Preferences of MTurk Survey Respondents." *Research & Politics* 2(3)(https://doi.org/10.1177/2053168015604648).

Hunter, James Davison. 1992. *Culture Wars: The Struggle to Control the Family, Art, Education, Law, and Politics in America*. New York: Basic Books.

Illouz, Eva. 1997. *Consuming the Romantic Utopia: Love and the Cultural Contradictions of Capitalism*. Berkeley: University of California Press.

Jakobson, Roman. 1960. "Linguistics and Poetics." Pp. 350–77 in *Style in Language*. Cambridge, MA: MIT Press.

Jenkins, James J., Wallace A. Russell, and George J. Suci. 1958. "An Atlas of Semantic Profiles for 360 Words." *American Journal of Psychology* 71(4):688–99.

Ji, Shihao, Nadathur Satish, Sheng Li, and Pradeep Dubey. 2016. "Parallelizing Word2Vec in Shared and Distributed Memory" (arXiv:1604.04661).

Johnson, Melvin, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation." *Transactions of the Association for Computational Linguistics* 5:339–51.

Jost, Jürgen, and Shiping Liu. 2014. "Ollivier's Ricci Curvature, Local Clustering and Curvature-Dimension Inequalities on Graphs." *Discrete & Computational Geometry* 51(2):300–322.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. "Bag of Tricks for Efficient Text Classification" (arXiv:1607.01759).

Kaufer, David S., and Kathleen M. Carley. 1993. "Condensation Symbols: Their Variety and Rhetorical Function in Political Discourse." *Philosophy & Rhetoric* 26(3):201–26.

Khan, Shamus Rahman. 2010. *Privilege: The Making of an Adolescent Elite at St. Paul's School*. Princeton, NJ: Princeton University Press.

Kulkarni, Vivek, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. "Statistically Significant Detection of Linguistic Change." Pp. 625–35 in *Proceedings of the 24th International Conference on World Wide Web, WWW '15*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Lamont, Michèle. 1992. *Money, Morals, and Manners: The Culture of the French and the American Upper-Middle Class*. Chicago: University of Chicago Press.

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

Lamont, Michèle, and Annette Lareau. 1988. "Cultural Capital: Allusions, Gaps and Glissandos in Recent Theoretical Developments." *Sociological Theory* 6(2):153–68.

Lareau, Annette, and Elliot B. Weininger. 2003. "Cultural Capital in Educational Research: A Critical Assessment." *Theory and Society* 32(5):567–606.

Le, Quoc, and Tomas Mikolov. 2014. "Distributed Representations of Sentences and Documents." *Proceedings of the 31st International Conference on Machine Learning*, Beijing, China.

Lee, Monica, and John Levi Martin. 2015. "Coding, Counting and Cultural Cartography." *American Journal of Cultural Sociology* 3(1):1–33.

Lerner, Melvin J., and Dale T. Miller. 1978. "Just World Research and the Attribution Process: Looking Back and Ahead." *Psychological Bulletin* 85(5):1030–51.

Lev, Guy, Benjamin Klein, and Lior Wolf. 2015. "In Defense of Word Embedding for Generic Text Representation." Pp. 35–50 in *Natural Language Processing and Information Systems*, edited by C. Biemann, S. Handschuh, A. Freitas, F. Meziane, and E. Métais. New York: Springer International Publishing.

Levay, Kevin E., Jeremy Freese, and James N. Druckman. 2016. "The Demographic and Political Composition of Mechanical Turk Samples." *SAGE Open* (https://doi.org/10.1177/2158244016636433).

Lévi-Strauss, Claude. 1963. *Structural Anthropology*. New York: Basic Books.

Levy, Omer, and Yoav Goldberg. 2014. "Neural Word Embedding as Implicit Matrix Factorization." Pp. 2177–85 in *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger. Red Hook, NY: Curran Associates, Inc.

Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. "Improving Distributional Similarity with Lessons Learned from Word Embeddings." *Transactions of the Association for Computational Linguistics* 3:211–25.

Lin, Yuri, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman, and Slav Petrov. 2012. "Syntactic Annotations for the Google Books Ngram Corpus." Pp. 169–74 in *Proceedings of the ACL 2012 System Demonstrations, ACL '12*. Stroudsburg, PA: Association for Computational Linguistics.

Lizardo, Omar. 2017. "Improving Cultural Analysis: Considering Personal Culture in Its Declarative and Nondeclarative Modes." *American Sociological Review* 82(1):88–115.

Marx, Karl. [1867] 2004. *Capital: A Critique of Political Economy*. London, UK: Penguin.

Marx, Karl, and Friedrich Engels. 1970. *The German Ideology*. New York: International Publishers.

McCall, Leslie. 2005. "The Complexity of Intersectionality." *Signs: Journal of Women in Culture and Society* 30(3):1771–800.

Mears, Ashley. 2010. "Size Zero High-End Ethnic: Cultural Production and the Reproduction of Culture in Fashion Modeling." *Poetics* 38(1):21–46.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331(6014):176–82.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space" (arXiv:1301.3781).

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." Pp. 3111–9 in *Advances in Neural Information Processing Systems 26*, edited by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. Red Hook, NY: Curran Associates, Inc.

Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. 2013. "Linguistic Regularities in Continuous Space Word Representations." *Proceedings of NAACL-HLT 2013*, 746–51.

Miller, George A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38(11):39–41.

Mische, Ann. 2011. "Relational Sociology, Culture, and Agency." Pp. 80–97 in *Sage Handbook of Social Network Analysis*, edited by J. Scott and P. J. Carrington. Thousand Oaks, CA: Sage Publications.

Mohr, John W., and Petko Bogdanov. 2013. "Introduction—Topic Models: What They Are and Why They Matter." *Poetics* 41(6):545–69.

Mohr, John W., Robin Wagner-Pacifici, and Ronald L. Breiger. 2015. "Toward a Computational Hermeneutics." *Big Data & Society* 2(2)(https://doi.org/10.1177/2053951715613809).

Nagy, William E., Patricia A. Herman, and Richard C. Anderson. 1985. "Learning Words from Context." *Reading Research Quarterly* 20(2):233–53.

Nalisnick, Eric, Bhaskar Mitra, Nick Craswell, and Rich Caruana. 2016. "Improving Document Ranking with Dual Word Embeddings." Pp. 83–84 in *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Nickel, Maximillian, and Douwe Kiela. 2017. "Poincaré Embeddings for Learning Hierarchical Representations." Pp. 6338–47 in *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Red Hook, NY: Curran Associates, Inc.

Osgood, Charles E. 1969. "On the Whys and Wherefores of E, P, and A." *Journal of Personality and Social Psychology* 12(3):194–9.

Osgood, Charles Egerton, George J. Suci, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Pachucki, Mark A., and Ronald L. Breiger. 2010. "Cultural Holes: Beyond Relationality in Social Networks and Culture." *Annual Review of Sociology* 36(1):205–24.

Pakulski, Jan, and Malcolm Waters. 1996. *The Death of Class*. London, UK: Sage.

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 (https://doi.org/10.1371/journal.pone.0137041).

Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "Glove: Global Vectors for Word Representation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–43.

Piketty, Thomas. 2014. "Capital in the Twenty-First Century: A Multidimensional Approach to the History of Capital and Social Classes." *British Journal of Sociology* 65(4):736–47.

Politis, Dimitris N., and Joseph P. Romano. 1992. "A Circular Block-Resampling Procedure for Stationary Data." *Exploring the Limits of Bootstrap* 263–70.

Politis, Dimitris N., and Joseph P. Romano. 1994. "The Stationary Bootstrap." *Journal of the American Statistical Association* 89(428):1303–13.

Politis, Dimitris N., Joseph P. Romano, and Michael Wolf. 1997. *Subsampling*. New York: Springer.

Reay, Diane. 1998. "Rethinking Social Class: Qualitative Perspectives on Class and Gender." *Sociology* 32(2):259–75.

Rei, Marek, and Ted Briscoe. 2014. "Looking for Hyponyms in Vector Space." Pp. 68–77 in *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, Baltimore, MD.

Ricoeur, Paul. 1981. *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. Cambridge, UK: Cambridge University Press.

Ridgeway, Cecilia L. 2011. *Framed by Gender: How Gender Inequality Persists in the Modern World*. New York: Oxford University Press.

Roget, Peter M. 1912. *Thesaurus of English Words and Phrases*, edited by A. Boyle. New York: E. P. Dutton.

Rosch, Eleanor, and Carolyn B. Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7(4):573–605.

Rosvall, Martin, and Carl T. Bergstrom. 2008. "Maps of Random Walks on Complex Networks Reveal Community Structure." *Proceedings of the National Academy of Sciences of the United States of America* 105(4):1118–23.

Rudolph, Maja, Francisco Ruiz, Stephan Mandt, and David Blei. 2016. "Exponential Family Embeddings." Pp. 478–86 in *Advances in Neural Information Processing Systems 29*, edited by D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett. Red Hook, NY: Curran Associates, Inc.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. *IPUMS USA: Version 9.0*. Minneapolis, MN: IPUMS.

Ruiz, Francisco J. R., Susan Athey, and David M. Blei. 2017. "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements" (arXiv:1711.03560).

Salganik, Matthew J. 2017. *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

Salzinger, Leslie. 2003. *Genders in Production: Making Workers in Mexico's Global Factories*. Berkeley: University of California Press.

de Saussure, Ferdinand. 1916. *Course in General Linguistics*. New York: Columbia University Press.

Schröder, Tobias, Jesse Hoey, and Kimberly B. Rogers. 2016. "Modeling Dynamic Identities and Uncertainty in Social Interactions: Bayesian Affect Control Theory." *American Sociological Review* 81(4):828–55.

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.

Shi, Feng, Jacob G. Foster, and James A. Evans. 2015. "Weaving the Fabric of Science: Dynamic Network Models of Science's Unfolding Structure." *Social Networks* 43(Supplement C):73–85.

Simmel, Georg. [1900] 2004. *The Philosophy of Money*. New York: Routledge.

Skeggs, Beverley. 1997. *Formations of Class & Gender: Becoming Respectable*. London, UK: Sage.

Smith, Charles J. 1903. *Synonyms Discriminated: A Dictionary of Synonymous Words in the English Language*, edited by P. Smith. New York: Henry Holt.

Svallfors, Stefan. 2006. *The Moral Economy of Class: Class and Attitudes in Comparative Perspective*. Stanford, CA: Stanford University Press.

Taddy, Matt. 2015a. "Document Classification by Inversion of Distributed Language Representations" (arXiv:1504.07295).

Taddy, Matt. 2015b. "Document Classification by Inversion of Distributed Language Representations." Pp. 45–49 in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Stroudsburg, PA: Association for Computational Linguistics.

Tversky, Amos, and Itamar Gati. 1978. "Studies of Similarity." *Cognition and Categorization* 1:79–98.

Veblen, Thorstein. [1899] 1912. *The Theory of the Leisure Class: An Economic Study of Institutions*. New York: B. W. Huebsch.

Vilhena, Daril A., Jacob G. Foster, Martin Rosvall, Jevin D. West, James Evans, and Carl T. Bergstrom. 2014. "Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication." *Sociological Science* 1:221–38.

Warner, W. Lloyd, Marchia Meeker, and Kenneth Eells. 1949. *Social Class in America: A Manual of Procedure for the Measurement of Social Status*. Chicago: Science Research Associates.

Weber, Max. 1978. *Economy and Society*. Berkeley: University of California Press.

Webster's Collegiate Thesaurus. 1976. Springfield, MA: Merriam-Webster.

Weeden, Kim A., and David B. Grusky. 2005. "The Case for a New Class Map." *American Journal of Sociology* 111(1):141–212.

Whorf, Benjamin Lee. 1956. *Language, Thought and Reality, Selected Writings of Benjamin Lee Whorf*, edited by J. B. Carroll. Cambridge, MA: MIT Press.

Willis, Paul. 1977. *Learning to Labor: How Working Class Kids Get Working Class Jobs*. New York: University of Columbia Press.

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford, UK: Blackwell.

Wright, Erik Olin. 1979. *Class Structure and Income Determination*. New York: Academic Press.

Wright, Erik Olin. 2000. *Class Counts: Student Edition*. Cambridge, UK: Cambridge University Press.

Zelizer, Viviana A. 1979. *Morals and Markets: The Development of Life Insurance in the United States*. New York: Columbia University Press.

Zelizer, Viviana A. 1989. "The Social Meaning of Money: 'Special Monies.'" *American Journal of Sociology* 95(2):342–77.

Zhou, Guangyou, Tingting He, Jun Zhao, and Po Hu. 2015. "Learning Continuous Word Embedding with Metadata for Question Retrieval in Community Question Answering." Pp. 250–9 in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*.

Zipf, George Kingsley. 1932. *Selected Studies of the Principle of Relative Frequency in Language*. Cambridge, MA: Harvard University Press.

Austin C. Kozlowski is a doctoral student at the University of Chicago, Department of Sociology. His research applies a diverse set of methodological tools to the analysis of contemporary American culture.

Matt Taddy is Vice President for Economic Technology and Chief Economist at Amazon. He previously was professor of economics and statistics at the University of Chicago Booth School of Business. His research focuses on statistics, machine learning, and their application to a wide range of problems and large-scale data in economics, social science, and business.

James A. Evans is Professor of sociology at the University of Chicago, Department of Sociology, and External Professor at the Santa Fe Institute. His research uses large-scale data, machine learning, and generative models to understand how collectives think, what they know, and what they create. He is especially interested in innovation and the emergence of ideas, shared patterns of reasoning, and processes of attention, communication, agreement, and certainty in science, technology, society, and culture.