

A Century of Immigrant Incorporation: Corpus Analysis of Etiquette Books to Uncover Stability and Change in Representation, Sentiment, and the Ordinariness of Immigrants

Andrea Voyer¹
Associate Professor
Stockholm University
andrea.voyer@sociology.su.se

Zachary D. Kline
Doctoral Student
University of Connecticut

Madison Danton
Doctoral Student
University of Connecticut

Tatiana Volkova
Masters Student
Stockholm University

April 2021

Word count: 9,635 (Excluding tables and references)

Number of Figures: 8

Number of Tables: 4

Keywords: assimilation, immigration, word embeddings, sentiment, etiquette

Declaration of interests: none.

¹ Corresponding author. Department of Sociology, Stockholm University, SE-106 91 Stockholm, Sweden. Tel: +46 (0)72 523 79 99 Email: andrea.voyer@sociology.su.se This article is based on research supported by the United States National Science Foundation Award No. 1851304, The University of Connecticut Office of the Vice President for Research, The University of Connecticut College of Liberal Arts and Sciences, and the University of Connecticut Department of Sociology. The authors wish to thank Lauretta Czarnecki, Alexander Karl, Asia Letlow, Pedro Pinales, Olivia Reid, and Stuart Riepl for their research assistance.

From Immigrant to Normal: Computational Approaches to Examining Incorporation through Shifts in the Mainstream

Abstract

The goal of this article is to contribute to the growing literature on mainstream shifts in the wake of immigration by using computational social scientific tools to examine the changing representation of immigrant groups in the social “mainstream.” Analyzing a corpus of etiquette books published between 1922 and 2017, we examine three elements of mainstream shifts: 1. the representation of different immigrant groups through naming and salience to the texts; 2. the positive and negative sentiments associated with different immigrant groups; and 3. the placement of different immigrant groups on a semantic dimension represented by the poles of “normal” and “strange.” Our preliminary results show the gradual adaptation and adjustment of the social mainstream as a result of immigrant integration, with some caveats. We observe increased representations of immigrants as named person entities in the texts, but these changes are quite recent and are dwarfed by the continuing dominance white, Anglo-American names. We observe declining negative sentiment for all immigrant groups over time, but with differences that we interpret to reflect group-specific historical, political, and cultural circumstances, including persistently less positivity associated with non-white immigrants. Finally, semantic shifts in the text demonstrate that most immigrants move from an association with “strange” and toward associations as “normal” in the texts, although continued ambiguity is attached to “Muslim.” These findings provide a new source of information that both corroborates the reality of mainstream shifts in light of immigration, and also reveals the persistence of racialized exclusions and the ongoing reality of immigrant integration as a form of “working toward whiteness” (Roediger 2018). The research techniques we develop can be adapted to other studies of historical change related to representation, social sentiment, and symbolic inclusion in the social mainstream.

Keywords

assimilation, immigration, word embeddings, sentiment, etiquette

I. Introduction

Developments in computational social science, in particular machine learning, provide new angles for studying immigrant incorporation. The study of immigration has a long history in the social sciences, but that research has generally examined immigrant incorporation by focusing on the assimilation or integration of immigrant minorities to the societies where they settle. Even though many of the more influential theories of immigrant incorporation posit the importance of non-immigrants to immigration (c.f. Alba 2005; Alba and Nee 2003; Massey and Sánchez 2010; Portes 1996; Portes and Rumbaut 2001; Portes and Zhou 1993), these non-immigrant “established” (Jiménez 2017) individuals rarely take center stage in research. When they do, they are generally seen as part of the context of immigrant reception or as gatekeepers in the process of immigrant integration.

Recent research has flipped the script on immigrant incorporation by observing how established individuals are not merely the people holding the keys to the gates of society; they are themselves participants in the process of incorporation through their own adaptations and adjustments in light of the social changes immigration brings. Research shows how established individuals change as a result of immigration, including developing comfort in multicultural spaces, rethinking aspects of how they live their everyday lives and conduct social relationships, relativizing their own social and cultural norms, and adopting a more expansive sense of what it means to be a neighbor and a citizen (Jimenez 2017; Voyer 2013b).

To date, this line of research is carried out primarily through public opinion surveys, qualitative interviews, and observation. But, the co-adaptations of the established are not merely individual-level phenomena among established individuals. The social impacts of immigration shape group boundaries and practices and collective meanings and sentiments associated with the “mainstream” establishment (Voyer 2013a). Methodological innovations from computational social science provide new ways to scale up the study of immigrant incorporation through shifts in the mainstream.

In this article, we present a computational sociological approach to observing mainstream adaptations and adjustments in light of immigration. Analyzing a historical corpus of texts that are explicitly charged with describing white middle-class social norms over the period 1922-2017, we observe social change occurring in light of immigration and in relationship to long-standing groups who once were and may currently still be seen as immigrants and outsiders in the United States: Catholic, Chinese, Cuban, Irish, Italian, Jewish, Mexican, and Muslim (Dumenil 1991; Portes and Back 1985; Takaki 1993). Research has shown that a racial in-betweenness emerged in which some of these groups (e.g., Irish, Italian, Jewish) were increasingly seen as white and American. In contrast, other groups were not provided the same recognition and remained racial others, outside of the white mainstream (Roediger 2005).

Our computational analyses uncover mainstream adaptations in relation to each of these groups. Specifically, we examine shifts in representation, sentiment, and the extent to which different immigrant groups are conveyed as being more “normal” or more “strange” over time. The results show that immigrant incorporation does occur through mainstream shifts, with some caveats. We observe increased representation of immigrants through the gradual inclusion of immigrant nationalities within the names appearing in the texts, but these changes are quite recent and are dwarfed by the continuing dominance of the white Anglo-American mainstream establishment. Likewise, while there is little text in our corpus with a close relationship to immigrant groups, there is more engagement with these groups in the texts over time. The nature of that engagement also changes. We observe declining negative sentiment for most immigrant groups over time in favor of neutrality or positivity, but with persistently less positivity associated with non-white immigrant groups. Finally, semantic shifts in the text demonstrate that most immigrant groups move from an association with “strange” and toward associations as “normal” in the texts, although continued ambiguity is attached to some groups, the “Muslim” category in particular. These findings provide new data and methods that demonstrate their value by corroborating both the reality of mainstream shifts in light of immigration and the persistence of racialized exclusions. The results demonstrate the ongoing reality of immigrant incorporation through normative shifts as a form of “working toward whiteness” (Roediger 2005).

II. The Literature: immigrants and their contexts

The role of widespread social change as an element of immigrant incorporation has deep roots in the literature, even if that role tends to be more implicit. According to Alba and Foner (2015), immigrant incorporation results in social membership.

Full membership means having the same educational and work opportunities as long-term native-born citizens, and the same chances to better their own and their children’s lot. It also means having a sense of dignity and belonging that comes with acceptance... (Alba and Foner 2015: 1)

If acceptance is an element of the encounter between immigrants and others, it follows, then, that immigrant incorporation is a process which requires action from both the immigrants and the mainstream of society.

Many influential theories of immigrant incorporation posit the importance of the mainstream to immigration (c.f. Alba 2005; Alba and Nee 2003; Massey and Sánchez 2010). For example, segmented assimilation theory (Portes 1996; Portes and Rumbaut 2001; Portes and Zhou 1993) postulates that the nature of immigrant incorporation hinges upon different aspects of the migration experience, including the context of immigrant reception, which consists of attitudes towards immigrants and public practices and policies shaping immigrants’ opportunities in the new country. Past work has also identified the symbolic processes underlying immigrant

incorporation. Alba's (2005) bright and blurred boundaries suggest that, in some contexts, the boundary between ethnic minorities and the mainstream majority is blurred, allowing some immigrant group characteristics to fuse with the social mainstream, and, in other contexts, the boundary is bright, maintaining the exclusion of immigrant group characteristics. Wimmer has extended this insight and developed a comprehensive typology of the boundaries between ethnic groups, including a variety of approaches to modifying ethnic group boundaries, including the development of overarching global, civil and local identities (Kroneberg and Wimmer 2012; Wimmer 2008; Wimmer 2009).

The literature tends to be fairly optimistic when it comes to the promise and possibilities for immigrant incorporation, which is typically seen as a multigenerational process (Drouhot and Nee 2019). We can point to immigrant "success" stories, such as how Irish, Italian, Jewish immigrants, who all initially faced hostility and racialized exclusion, over generations came to be seen as white American ethnic groups (Brodkin 1998; Ignatiev 2009; Roediger 2005). Of course, there are other immigrant groups that did not "become white" – including Chinese, Japanese, and Mexican immigrants (Lee 2003; Massey et al. 2002; Takaki 1993). As Roediger (2005) aptly discusses, these non-white immigrants and their descendants, even in cases where they achieved economic and educational successes, are not generally symbolically included in the American mainstream. Instead, their inclusion shifted with the ebb and flow of political and economic tides – from inclusion as racialized Americans to exclusion as potential outsiders who could be exhorted to "go back to their own country." Historically, this in-betweenness has both reflected and reinforced by systematic racism, for example in the way federal red-lining policies and the development of post-war housing moved Irish, Jewish, and Italian immigrants into the white suburbs, while leaving Mexican and Asian immigrants in the cities (Roediger 2005). Recent research shows the continuing importance of race to immigration. Even today, white Americans do not mind having immigrant neighbors. They mind even less if those immigrants are white, speak English fluently, and volunteer in the community, but they see non-white immigrants as less preferable neighbors and as less similar to themselves (Schachter 2016).

In other words, the symbolic boundary of the white American mainstream is crucial to the trajectory of immigrant incorporation, but research tends to focus on immigrants. This imbalance contributes, as Schinkel (2018) notes, to the background reification and invisibility of the mainstream. As a welcome exception, there is growing literature extending both theory and empirical knowledge of mainstream shifts in light of immigration. The fundamental insight of such work is that immigrant incorporation is not a one-way process in which immigrants learn how to fit in. It is rather a two-way and relational process involving the adjustments and adaptations of both immigrants and the established people and practices in the societies where they settle. For example, Voyer (2013b) finds that rapid and substantial Somali immigration into a historical white town prompts a redefinition of community membership that could include black immigrant newcomers, a revision of belonging that blurred racial boundaries. Meanwhile,

in their research among established individuals in Silicon Valley, Jimenez and Horowitz (2013) find that mainstream conceptions of success devalue whiteness in comparison with Asian-ness. Jimenez (2017) observes that established individuals who are in at least the third generation in the United States become comfortable and familiar with the languages, foods, and cultural and religious practices of immigrant newcomers, and ultimately revise their notions of what it means to be an American.

This existing research on adaptations to immigration provides direct evidence of the “remaking of the American mainstream” (Alba and Nee 2009) as an element of immigrant incorporation, but to date this work has largely consisted of ethnographies and interview studies of individuals. Recent advances in computational methods for text analysis provide the opportunity to scale this research up from established individuals to the social establishment and to focus on the construction of the American mainstream over time and in relation to immigrant groups.

III. Analytical Approach. Viewing Integration through the Evolution of the Mainstream

Analysis of mainstream shifts in light of immigration requires a theory of the mainstream and its boundary dynamics. In this article, we draw on civil sphere theory as it has been applied to the topic of immigrant incorporation (Alexander 2001, Voyer 2013a). The underlying assumption of civil sphere theory is that societies include a civil sphere (Alexander 2006) – a symbolic collectivity built upon core values and the practices in which they are reified. We think of this civil sphere as the mainstream of society. Immigrant incorporation occurs in relationship to the civil sphere through the making of symbolic distinctions between “us” and “them” that have no *necessary* relationship to ethnic and immigrant groups, although, in practice, the boundaries of the civil sphere are often drawn along racial, ethnic, religious, and other “primordial” boundaries (Alexander 2001). From this theoretical perspective, to be incorporated is to be widely seen as a recognizably “normal” member of the civil sphere and to be included in symbolic constructions and expressions of the “we” of that civil sphere (Voyer 2013a). Inclusion of this kind can occur in different ways. Immigrants and their descendants can be included through strict assimilation in which the conception of the civil sphere remains unchanged, and those who are incorporated must abandon any trace of their immigrant backgrounds. Immigrants could also be included as a result of the expansion of the boundary of the civil sphere through revision of social membership so that immigrant and ethnic phenotypes, identities, languages, religions, and other characteristics are seen as American (Alexander 2001). The approach we outline here is designed to establish whether or not these mainstream shifts occur.

IV. Data & Method

Computational text analysis, which combines quantitative methods and qualitative epistemology and allows for back-and-forth movement between qualitative readings and computational

analyses of text (Nelson forthcoming; Nelson 2020), is well-suited to analyses of the symbolic and affective inclusion and exclusion of immigrant groups through mainstream shifts. Applied text analysis makes it possible to answer theoretical and qualitative sociological questions by developing large scale and reproducible analyses of the social meanings present in texts.

For our analyses, we employ a corpus of etiquette books. The purpose of etiquette writing is to bring existing mundane practices and meanings from the cultural background to the foreground, and we take this practical and symbolic background as a representation of the social mainstream. Sociologists have long recognized that etiquette literature highlights the mainstream's socially significant cultural behaviors, signals, social valuations, and justifications of social status and hierarchy (Arditi 1999; Bourdieu 1984; Elias 2000; Goffman 1963). Past research examines the role of etiquette books in establishing social control, generating in people an emotional predisposition to act in concert with their social location/social roles, constructing channels for communication and mutual understanding, creating and indicating social distinctions, and legitimating moral evaluations and systems of governance (Arditi 1999; Elias 2000, Hemphill 1999; Menell 2007; Wouters 2007). Recent research has used computational analysis of etiquette books to study general social changes (Abrutyn and Carter 2015), but, to our knowledge, this is the first research to use etiquette literature in a study of immigrant incorporation.

The social world revealed by etiquette advice is grounded in time and space. There is a large and long-standing etiquette industry including, but not limited to books, films, radio programs, podcasts, YouTube channels, magazines, newspaper columns, and etiquette training and consulting services. The industry is internally differentiated, with distinct content for particular social roles (e.g., corporate etiquette versus domestic etiquette), racial groups (e.g., etiquette books for the Black elite and middle class), and situations (e.g., golfing etiquette and wedding etiquette). Through the selection of topics and its presumed audience, etiquette media reveal the assumed characteristics of the mainstream. Computational sociological analysis of an etiquette corpus makes it possible to observe the relationship between immigrants and the mainstream, exploring whether there is evidence for increased representation of immigrants, positive emotions associated with immigrants, and shifts in the “normal-ness” of immigrants.

Data: The Emily Post Corpus

Our data consist of a text corpus of all 20 editions of the best-selling and longest-running American etiquette manual, Emily Post's *Etiquette*. Each edition of *Etiquette* is 500 – 1100 pages long and includes photographs, figures, and charts. The first edition of the book is in the public domain, and all others remain under copyright and are used under fair use doctrine. A physical copy of each edition was scanned into a text document using optical character recognition (OCR) software. The resulting Emily Post corpus includes more than 5 million words and 12,000 pages

of text. Errors arising during the OCR process are managed and corrected within the version control manager GitHub and HTML mark-up assists with the retention of text features like images, tables, and sidebars. Prior to analysis, the text is tokenized, and stop words are removed. Analyses are conducted using text analysis tools based in R and Python programming languages.

Etiquette is considered a “pre-eminent example of American advice literature” (Lees-Maffei 2012: 217). The book was first published in 1922 and subsequently revised about every five years, with its most recent publication in 2017. *Etiquette*’s popularity is enduring. The book has sold at least 30,000 copies a year since its publication (Claridge 2009). Only the bible has been purchased more by schools and libraries (Jacobs 2011). Emily Post, the first author of *Etiquette*, has been characterized as the “arbiter of etiquette,” and that reputation has long-persisted since her death in 1960 (Lees-Maffei 2014). In 1995, *Etiquette* was featured in the New York Public Library’s “Books of the Century” exhibit, and it was commemorated with a postage stamp in 1998 (Lees-Maffei 2012).

We take *Etiquette* as valid data on the American mainstream because presenting the mainstream is the purpose of such texts. The book is tasked with representing the social organization at the time of its publication by keeping pace with the changing social norms of its intended audience. Research shows that *Etiquette* remains thematically and structurally consistent with its contemporary etiquette manuals (Wouters 2007, Arditi 1999). *Etiquette* does not merely reflect the opinions and proclivities of its authors, and we attach no more causal power to this text and its authors when it comes to immigrant incorporation than we would to any other successful authors or widely read books. Naturally, there have been changes in the authorship of *Etiquette*. The Emily Post Institute, which was founded by Post and her children, has maintained the Post family business since Post’s death. Her descendants are the past and current authors of *Etiquette* and several other books taking up specific etiquette topics. The Emily Post Institute also produces syndicated advice columns, etiquette training seminars, etiquette podcasts, and dispenses etiquette advice and knowledge on talk shows and news programs.

Etiquette books are “tailored to suit [their] public’s conceptions of behavior” (Arditi 1999: 28). In many ways, *Etiquette* is a crowd-sourced text. After the publication of the first edition, Post received approximately 50,000 letters from readers describing their views of socially-acceptable behavior (Arditi 1999: 28). From then on, she received thousands of letters each week and used them to develop content for her columns and programs, with some of that material finding its way into future editions of *Etiquette* (Jacobs 2001). The authors that have followed her also collect reader and listener feedback. In other words, *Etiquette* is charged with observing and representing its readers’ conceptions of the American mainstream back to themselves. The book’s enduring popularity is testament to its success in this endeavor.

The close relationship between *Etiquette* and its historical context leads the book to change gradually, even given changes in authorship. Table 1 shows changes in the corpus over time by plotting the cosine similarity between all editions. Generally, editions display a high degree of similarity, with edition-by-edition changes rarely exceeding .3. The pace of change is also largely independent from author type, outside of a small increase in the rate of change proceeding the tenure of the corpus' founding author, Emily Post. Many factors contribute to edition-by-edition change in texts (Giuliani et al. 2006). Because each edition of *Etiquette* is written in explicit reference to the previous edition(s), topical changes, such as the decision to add or delete topics, can certainly reflect mainstream shifts in light of immigrant integration. At the same time, more general and superficial shifts in vocabulary, editorial practices, developments in publishing and printing, and authors' stylistic quirks also shape *Etiquette*'s development and may also be related to immigrant integration. Studying a corpus consisting of multiple editions makes it possible to consider both topical and superficial changes in relationship to mainstream shifts.

TABLE 1 ABOUT HERE

Immigration is more an implicit than an explicit element of *Etiquette*. Our exploratory searches for immigrant keywords turned up few explicit mentions, and no clear trends. Many explicit occurrences were associated with just a few topics such as foreign travel, hosting foreign dignitaries, choosing between the continental or American style of holding one's fork and knife, and how to avoid snobbery when using French expressions while speaking English. Immigrant incorporation is more implicit in the texts. Traditional qualitative reading and interpretation of *Etiquette* demonstrate that mainstream shifts indicating incorporation are evident in the representation of and sentiment associated with immigrant groups in the texts and in background assumptions about whether immigrant groups are treated as included in the texts, presumably mainstream readers. Based on traditional content analysis of the text, we observed that immigrants and immigrant groups were present in the data in three ways: representation; sentiment; and presumed position as "normal" or "strange" relative to the mainstream.

Representation is visible in the introduction of locations, events, institutions, and practices associated with immigrant groups in *Etiquette*'s explanatory and how-to-discussions. For example, in 1937 Post first instructed readers on how to address both priests and rabbis, and in 1997 readers were first counseled how to behave at a religious service in a mosque and informed about what to expect at a quinceñera. Bar and bat mitzvahs, appropriate attire for Catholic church services, and what to expect at 100-day parties for babies were some events and rituals associated with different immigrant groups that were added and developed as *Etiquette* was revised in light of its role as an encyclopedic guide to navigating social situations its readers might encounter. Extracting and counting all such introductions and mentions by reading all

editions of *Etiquette* would be a monumental task, and computational text analysis can facilitate the identification and analysis of representation.

Different sentiments associated with immigrants and immigrant groups are legible in *Etiquette* as well. Sentiments can be explicitly connected to a judgment or valuation, such as the negative sentiment and mainstream exclusion attached to immigrants in this quotation from the 1937 and 1942 editions: “[A child] must never be allowed to hold his fork immigrant fashion, perpendicularly clutched in the clenched fist, and to saw across the food at its base with his knife” (Post 1937: 745). However, there is a great deal of implicit sentiment in texts, and computational tools for analyzing sentiment can assist in establishing the relationship between immigrants and immigrant groups in terms of both explicit and latent sentiment.

Mainstream shifts related to immigrants can also be examined through analysis of the presumed membership of immigrants and immigrant groups in the “normal” mainstream of *Etiquette*, or, alternatively as “strange” from the position of the mainstream. For clarification, let’s return to the inclusion of quinceñaras in 1997. At that time, and again in the 2004 edition, the discussion of quinceñaras was written with the presumption that the reader would likely encounter the event as a guest and not as a host (see Table 2). This is clearly visible in comparison with the discussion of the sweet-sixteen celebration, which describes how one should arrange the invitations and the meal for such an event. Meanwhile, the discussion of quinceñaras focuses on what a guest needs to know, including appropriate attire and whether or not gifts are expected. In this case, then, we see the representation of “foreign” practices, but not the inclusion of people who would hold a quinceñara in *Etiquette’s* readership.

TABLE 2 ABOUT HERE.

However, *Etiquette* does show a shift in the conception of its mainstream readership in 2011 (see Table 3). The 2011 edition of *Etiquette* presumes it has readers who need advice on hosting a quinceñara. An altered description of the event says a “quince” originates in Latin America, instead of being something that “Latin American girls” do. Furthermore, the text offers advice related to hosting the event. The book also includes a sample quinceñara invitation in the resources appendix.

TABLE 3 ABOUT HERE

As in this case, mainstream shifts in light of immigration are evident through the meanings within the text. The traditional qualitative reading and interpretation of *Etiquette* just presented is one way to access these meanings, but computational social scientific approaches to the analysis of text data extend these qualitative insights. A computational approach makes it possible to include a larger volume of text than is feasible for hand-reading, subjects qualitative

interpretation of text to additional scrutiny through the application of quantitative methods designed with interpretation in mind, and leads to the development of transparent and reproducible research that can be applied to other empirical questions and other sources of text data (Nelson 2020).

B. Method

Our research methods incorporate multiple natural language techniques, which we describe below. But two important methodological points must be addressed in advance.

First, the reader must keep in mind that our study examines the incorporation of specific immigrant groups in an interpretive and symbolic sense and does not include measures or estimates of incorporation in a material sense. Our analytic objective is to understand how the mainstream depicted in *Etiquette* changes in relation to immigration, particularly immigrant groups represented as Catholic, Chinese, Cuban, Irish, Italian, Jewish, Mexican, and Muslim. We meet this objective by looking at the representation of things (i.e., names and passages) that can be reasonably associated with these historical immigrant groups, the sentiment in text passages that our analyses find to be associated with these groups, and the relationship between these text passages and semantic dimensions represented by the binary poles “normal” and “strange.”

Second, the reader should also keep in mind what immigrant incorporation means when it comes to our particular corpus of etiquette manuals concerned with everyday life. Politics and public policy, including the topics of immigration and citizenship, are rarely addressed in *Etiquette* apart from the occasional reminder that such topics are generally to be avoided in social situations and that people who persist on making such things an issue can be problematic. We are thus unable to draw conclusions about the representation of, sentiment associated with, and conceptions of shared social membership with immigrant groups in American politics or news media as that would require a different corpus of texts. However, as Jimenez’s (2017) investigation of established individuals shows, mainstream adaptations in light of immigration occur precisely within the mundane world of everyday life: workplace relationships; food and fashion; norms of communication; familiarity with different religions; and the rituals that punctuate the life course from birth to death. The *Etiquette* corpus is very well suited to revealing the scope of immigrant incorporation in relationship to these elements of the American mainstream.

V. Analysis and Results

We present the analysis and results in terms of our interests in the representation, sentiment, and presumed mainstream membership of the immigrant groups we examine.

Representation

We took two approaches to representation.

Our first study of representation analyzes the ethnicity and national origins of last names appearing in the *Etiquette* corpus. Names are widely used in the books. Authors deploy hundreds of different names and surnames in sample messages and invitations, pseudonyms or characters used to describe correct and incorrect behavior, and even the occasional mention of real public figures or readers who have corresponded with questions. An analysis of these names could provide evidence of mainstream shifts through changes in representation. For example, as discussed above, a sample quinceñara invitation was included in the 2011 edition: an invitation from Mr. and Mrs. Jorge Delgado to celebrate their daughter, Ana Theresa, with the ceremony including a Catholic religious service (Post et al. 2011: 742).

Name Analysis

There are substantial challenges in identifying the ethnicity and nationality of first names, so we chose to analyze surnames. First, we extracted surnames from the corpus and established the frequency of those surnames. We used the open-source Natural Language Toolkit (NLTK) package 3.5 (NLTK Project 2020), which uses python coding language. NLTK is a computational linguistics package that can identify a variety of text features. Our code was written to extract surnames only. Given normal error as well as the impact of certain text features (for example, that names may be presented in the form of an addressed envelope) on extraction, we improved the accuracy of our extracted names by creating a list of stop words that precluded some of the most common misrecognitions such as days, months, nationalities, languages, holidays, titles, and numbers. A spreadsheet was created which contained all extracted names and their frequency, edition by edition. With these procedures, we extracted between 300 and 800 names per *Etiquette* edition. In the 1984 edition, surnames were used 765 times, and that is the highest occurrence. And in the 2017 edition surnames were used 316 times, and that is the lowest occurrence. On average, surnames appear in each edition between 400 and 600 times.

We identified two sources of quality data on the nationality or ethnicity of names. For ethnicity, we used a dataset containing all 162,253 surnames occurring at least 100 times in the 2010 US Census and specifying the distribution of each name across the Census racial and ethnic categories (Census 2016). The census data has the advantage of representing the US population, but the data does not contain information on the nationalities and religions we were interested in as immigrant groups. For this reason, we also used data from NamePrism API, a classification tool that has been designed to establish nationality and ethnicity using names (Ye et al. 2017). NamePrism API is trained on global data, which makes it less accurate for classifying names within the United States, but it did allow us to get information on most of the groups we were interested in. Both the census data and NamePrism had the limitation of being quite recent. It could be the case that data limitations are causing us to miss relevant historical contingency in

terms of the common nationality and ethnicity of surnames in the United States. The results should be interpreted in light of these facts and potential limitations.

Figure 1 presents the census race and ethnicity categories (white, black, Asian and Pacific Islander, and American Indian) for the names we extracted from each *Etiquette* edition. The category distribution for each name was taken from the census data set and the distributions were averaged across all the names included in each group. This table drives home the overwhelming whiteness of the names appearing in *Etiquette*, but also rise of non-white names, in particular names that are Asian and Hispanic. Specifically, if we imagined that the names we extracted were “the mainstream,” that mainstream would be 90-95% white in most editions. The mainstream would have increasing Non-Hispanic Asian and Native Hawaiian and Other Pacific Islander members beginning with the 1997 edition, and increasing Hispanic and Latinx members beginning in 1945. The representation of Asian and Hispanic ethnic groups within *Etiquette* was most pronounced in the 21st century, and the name analysis suggests that this increasing representation came more at the expense of the black than the white share of the mainstream. There was decreasing black inclusion in the mainstream for most years, beginning in 1969.

FIGURE 1 ABOUT HERE

NamePrism data made it possible to look more closely at the ethnicities and nationalities represented in the extracted names from *Etiquette* (Ye 2017). We were able to extract Chinese, Hispanic, Italian, Jewish, and Muslim names.² For each name on our list, we assigned it to our target immigrant groups if that group was the more prevalent source of the name (for example, if the name was prevalent as both an Italian and a Hispanic name, it would be credited to only the group it was most prevalent in). We examined the occurrences of names by immigrant group, weighting names that occurred multiple times.

The results are presented in figure 2. The first thing to observe is that there are very few (between 0 and 20) names for each group, by edition year. Given the small number of

² These groupings merged some different nationalities and regions based on the available data: Italian = "European,Italian,Italy" + "European,Italian,Romania"; Muslim = "Muslim,Pakistanis,Bangladesh" + "Muslim,Nubian" + "Muslim,Turkic,CentralAsian" + "Muslim,Persian" + "Muslim,Maghreb" + "Muslim,Pakistanis,Pakistan" + "Muslim,ArabianPeninsula" + "Muslim,Turkic,Turkey"; Jewish = "Jewish"; Hispanic = "Hispanic,Spanish"; Chinese = "EastAsian,Chinese." There was not name data for Ireland specifically. Instead, there were “Celtic, English” names, but this name list includes also Welsh, Scottish, and other names (Smith, for example) from the British Isles. This was conceptually problematic since it includes names we would expect to see in the mainstream (Smith, for example). Given the lack of specificity to Ireland, and Irish immigration to the US in particular, we excluded the Irish immigrant group from this part of the analysis but hope to identify a better source of Irish-American name data for future analyses.

occurrences, it is difficult to interpret clear trends, but it does seem that there is a general trend from no or little representation in the first 50 years of *Etiquette*, to increasing representation beginning with the 1975 edition. There may be some general downward movement again in the 2011 and 2017 editions. Jewish names occurred more after the 1950 edition, and occurrences increased between 1969 and 1984, and 2004 and 2011. The occurrences of Italian names increased between 1975 and 1997. While for Hispanic names we can see two peaks: between 1969 and 1975; and 2004 and 2011. Occurrences of Muslim names tend to be fairly consistently higher, with the appearance of a downward trend beginning in 2011. Occurrences of Chinese names are highest in the 21st century, and in the 1937 edition.

FIGURE 2 ABOUT HERE

Considering that there were more than 300 name occurrences in each edition, these observations suggest that there is little representation of these groups within the mainstream of the text, However, we believe these results should be interpreted with caution. In the case of the Hispanic names, which were also present in the census data, there do seem to be somewhat similar trends, but fewer names. To some extent, the lack of national and historical specificity may make the NamePrism data less applicable to our research. Our methodological choices may also have played a role, as we selected only names that were most prevalent in that group. A more nuanced approach that considers others beyond the most prevalent groups for each name may provide more information.

However, we also argue that the limitations of this data on representation through names shows the importance of moving beyond more traditional extraction, coding, and counting approaches to the computational analysis of text. In analyses that follow, we consider emerging computational approaches to examining text as embedded within its semantic and relational context to uncover social meanings.

Concept Saliency

Our second approach to examining representation looks for the symbolic representation of our immigrant groups by relying upon the word embedding method Concept Movers Distance (CMDist; Stoltz and Taylor 2019). Word embeddings are an increasingly popular tool for the study of meaning in text. Word embeddings leverage the strong connection between particular words and their linguistic contexts (Ellis 2019; Firth 1957; Garvin 1962; Lenci 2018; Stoltz and Taylor 2019). The underlying premise is that text can be conceptualized as a multidimensional map of words that reflects all the contexts in which words are used. Word embeddings are trained over very large text corpora, and through that training, they develop a detailed account of the interrelationships of all the words in the text. Words used in similar contexts are presumed to have a similar meaning. Even when those words do not frequently appear together, they are assumed to have a more parallel journey through semantic space. These similarities are

quantified as cosine similarity. CMDist uses cosine similarity scores from word embeddings to estimate the distance between all words in a document and the word or words denoting a focal concept selected by the researchers. CMDist is designed for use with smaller corpora and chunks of text (Stoltz and Taylor 2019).

In terms of representation, we used CMDist to estimate the salience of the target immigrant groups. To do this, we parsed every edition into paragraph-length 150-word chunks. We specified each immigrant group as a concept. For the sake of clarity, we will use italics when referring to the way these groups (*Catholic*, *Chinese*, *Cuban*, *Irish*, *Italian*, *Jewish*, *Mexican*, and *Muslim*) are operationalized as concepts in CMDist. The word embeddings used to run CMDist were collected from FastText. FastText embeddings are pretrained on a large set of sources including Wikipedia and common crawls of the web (Bojanowski et al. 2017; Joulin et al. 2016). Prepared by Facebook’s AI research team, FastText is among the most reliable and widely-used of embeddings, and has been shown to be effective when analyzing corpora from multiple time periods (El-Ebshihy et al. 2018; Stoltz and Taylor 2019). Based on the FastText embeddings, CMDist assigned standardized scores to all chunks. These CMDist scores estimate how salient each group concept was to the text in the chunk (Stoltz and Taylor 2019). Standardized scores generally range between -3 and 3. We deemed scores exceeding positive 3 to be especially relevant to the groups in question based upon the typical words used in the chunks. These salient chunks were then aggregated by edition into a “pseudo corpus” used in subsequent analyses.

Table 4 presents the count of chunks in each edition that were salient to each immigrant group. These chunks compose the pseudo corpora used in the sentiment and mainstream inclusion analyses. Shown also in parentheses are the proportion of chunks in the corresponding edition that was selected. These proportions provide further evidence of the very limited representation of the target groups in *Etiquette*. For example, the 4 *Mexican* chunks identified in 1992 represent less than 1 percent of the total chunks in the book. This one percentile was expected given the 3 standard deviation cut-off. A less stringent cut-off would have returned more salient chunks, but at a lower level of salience. The one percent can also be used to assess the extent to which the groups become more or equally represented through their salience to the text over time. As an example, *Jewish* was salient in less than 2 chunks for all editions between 1922 and 1942. *Jewish* proceeded to be more salient in later years, having more than 5 salient chunks in all years after 1956. This change can partially be drawn to the incorporation of Jewish ceremonies, rituals, and holidays. For example, in 1956 a section on “Orthodox and Reform Weddings” was added to *Etiquette* (Post 1956: 221). *Italian*, by contrast, holds a similar level of salience over time. Of all the concepts, *Catholic* has the consistently highest salience, which is to be expected given *Etiquette*’s inclusion of many religious rituals, something that might be salient to *Catholic* even if the text generally assumes a Protestant readership in the early editions.

TABLE 4 ABOUT HERE

We found evidence for mainstream shifts indicating immigrant incorporation in our qualitative readings of the text, but the computational results show only very slight changes in the limited representation of immigrants through surnames and salience. Representation through names does appear to trend upward for most groups, there is an increase in Hispanic and Asian representation through names in particular, although the limited results must be interpreted with caution given data limitations. Representation through salience trends differently for the target groups, and there are no easily interpretable patterns merely on the basis of concept representation. Analyzing the sentiment of salient chunks can provide further insight into the position of the selected immigrant groups to the mainstream.

Sentiment

In the previous analyses we examined the salience of the selected immigrant groups to *Etiquette*, but salience scores do not indicate the nature of the text's engagement with the focal concepts. Is there presumed negativity along the same lines as Post's claim discussed above that children should not be raised to have poor (i.e. immigrant-like) table manners, or are salient texts more neutral or even positive? We get at these elements of the text through sentiment analysis. Sentiment analysis is a computational tool that uses pre-existing measures of the sentimentality of words to establish the general emotive characteristics of a text (Prabowo and Thelwall 2009). We interpret greater neutrality and positivity in texts salient to the selected immigrant groups as suggesting greater inclusion in the mainstream.

Sentiment analysis on the pseudo corpus of salient chunks was conducted using the *SocialSent* package in R (Hamilton et al. 2016). *SocialSent* contains a collection of code and data for performing domain-specific sentiment analysis. The package includes historical sentiment lexicons for the past 150 years of English. These historical English sentiment lexicons contain sentiment scores for the top-5000 words (excluding stop words) for all decades ranging from 1850-2000. Given the duration of our corpus, we used the sentiment lexicons provided by the *SocialSent* package from 1920 to 2000. To reflect historical change in the meanings of words, we aimed to utilize time period-dependent sentiment lexicons to assess change in mainstream sentimentality of various immigrant groups. Sentiment lexicons in and of themselves allow researchers to examine opinions and attitudes reflected by words used in text (Prabowo and Thelwall 2009). Lexical sentiment, however, is influenced by both domain and sociohistorical context (Hamilton et al. 2016). The sentimentality of the word *awesome*, for example, began to change in the middle of the 20th century from being synonymous with *awful* to now being more synonymous with the words *good* or *impressive*. Upon evaluation of these lexicons by decade, Hamilton et al. (2016) found that over five percent of sentiment-bearing words reversed their polarity (i.e. between positive and negative) during the 150-year time period examined, and over one quarter of all words changed their sentiment label (including switches to or from neutral). Without such considerations of changing sentimentality, analyses which use corpora that span many decades of time can be misled by sentiments assigned by modern-day understandings of

words (Hamilton et al. 2016). Developed by the Stanford Natural Language Processing Group at Stanford University, *SocialSent* is one of very few algorithms that provides automatic and comprehensive sentiment lexicons for historical data (Hamilton et al. 2016).

The sentiment analyses provided a sentiment score for each edition in the pseudo corpus consisting of the salient chunks. Since sentimentality varies across *Etiquette* editions, we present the sentiment scores for each group relative to the average sentiment of the edition. Scores from the *SocialSent* package are obtained from a state-of-the-art sentiment induction algorithm which propagates sentiment polarities from small seed sets. These propagated scores were then bootstrapped using $B = 50$ and 7 words per random subset to lessen the influence of corpus artifacts that might be present in a seed set (see Hamilton et al. 2016 for more information). Sentiment scores are finally represented as the average inferred sentiment across these bootstrap samples.

The results of the analysis are present in FIGURES 3, 4, and 5. FIGURE 3 shows trends in sentimentality for all immigrant groups across all editions of *Etiquette*. Though there is individual variation between groups, the sentimentality of immigrant groups dipped negative from around the 1930's to the 1960's before notably shifting to being more positive. These plots demonstrate sentimentality of immigrant groups across all editions relative to the general sentimentality of each edition. Sentiment shifts across editions suggests that sentimentality toward all immigrant groups in our analysis have trended toward neutrality and positivity since around the 1960's. We interpret movement toward more positive and neutral sentiment to indicate mainstream shifts that suggest immigrant integration.

FIGURE 3 ABOUT HERE

Figure 4 presents sentiment plots for a set of four immigrant groups - *Italian, Jewish, Catholic, and Irish*. While the general pattern of increasing positivity is evident, there is considerable between group variation, both in this trend and the extremity of positive/negative sentiment. The sentimentality of *Italian* and *Catholic*, for example, suffered a dip toward negativity from the 1930s until the 1960s before notably returning to positivity by 2017. *Irish*, on the other hand, remained negative from the beginning of the corpus until a sharp upswing of positivity around the 1970s that persists through the 2017 edition. The sentimentality of *Jewish*, however, demonstrates considerable movement consisting of peaks and valleys that range from moderate positivity to extreme negativity from around the 1930s to the 1960s, not reaching a discernible trend toward positivity until the 2000s.

FIGURE 4 ABOUT HERE

Figure 5 plots the sentiment scores of the next four groups of immigrants - *Cuban*, *Muslim*, *Chinese*, and *Mexican*. For some groups, the general trend of increasing positive sentiment is evident. The sentimentality of *Mexican*, for example, has followed a similar trend to that of *Italian* and *Catholic*, such that sentimentality notably dipped negative from the 1930s until the 1960s before levelling out at relative neutrality until the 2017 edition. For *Muslim*, a similar trend toward positivity is much more apparent. While the sentimentality of *Muslim* does fluctuate wildly in the 1940s, the plot indicates a fairly linear progression toward positivity. *Chinese* and *Cuban*, however, demonstrate even greater variation from the general trend of increasing positive sentiment. The sentimentality toward *Chinese*, for example, has remained in relatively unchanged neutrality since the 1922 edition, while *Cuban* has actually trended to become consistently more *negative* across time since the 1940s.

FIGURE 5 ABOUT HERE

Position between “Normal” and “Strange”

Declining negative sentiment suggests a decrease in the sense that the relevant group is a problem for the mainstream. However, as discussed in the theory section above, immigrant incorporation through assimilation is indicated when the meanings of immigrant categories remain outside of the mainstream as something foreign and strange. On the other hand, with mainstream shifts through the revision of the mainstream, we would expect that the immigrant categories would instead be seen as normal and ordinary - the kind of Americans who would need advice on how to manage invitations and thank you notes for their daughter’s quinceañera, for example. Our final analyses are designed to explore whether there are expansions of the boundary of the mainstream in relationship to the selected immigrant groups.

Using CMDist, we specified a semantic dimension anchored by a conceptual antonym pair: *normal* and *strange* (Taylor and Stoltz 2020), and we analyze the same salient pseudo corpora as in the sentiment analyses by establishing the position of each corpus in the semantic dimension. We continue with the convention of using italics when referring to *normal* and *strange* as concepts operationalized in CMDist. The premise of these analyses is that texts can be measured based upon their semantic position between *normal* and *strange*. By assuming a continuum between this antonym pair (Fellbaum 1998), we can assess how groups move across the continuum over time. Using word embeddings to map the semantic space between *normal* and *strange*, and the progression of chunks that are salient to each immigrant group over editions will reveal critical insights into mainstream shifts. Our assumption is that integration into the mainstream would lead groups to be more loosely associated with *normal* over time.

Analyses were conducted using the CMDist antonym pair function (Taylor and Stoltz 2020). FastText embeddings are used to prepare a semantic space between two concepts defined using

the words *normal* and *strange*. From here, the pseudo corpora measure how the editions move between *normal* and *strange*, giving each edition a standardized score. These scores are used to show how each group changed in relation to the American mainstream. Results of the analysis are presented in figures 6, 7, and 8.

Figure 6 presents results from a CMDist that included all editions from each of the 8 pseudo corpora for all groups. The scores are standardized and mean-centered at 0 for all of the pseudo-editions across all groups. Results depict a general movement towards *normal* from *strange*. The figure also demonstrates considerable variation among groups but the tendency toward convergence in *normality* in the twenty-first century.

FIGURE 6 ABOUT HERE

Figure 7 plots the CMDist scores for the first set of immigrant groups. These scores are mean-centered within the group, with a year's score of 0 representing the average distance between *normal* and *strange* for that group across the time period. As such, scores should not be compared between groups as they are presented. All groups follow the general trend towards integration, but there are notable historical variations. *Catholic*, for example, became *normal* during the Kennedy years of the 1960s, followed by continued integration with trends to higher levels of *normalcy* that persist through the 2017 edition. *Irish*, in a similar fashion, saw integration as *normal* begin a decade after WW2. *Italian* and *Jewish*, on the other hand, do not have such clear movements toward *normal*. Both groups crossed from *strange* to *normal* in the 1950s or 1960s, but *Jewish* was strongly *strange* before that time, has its peak *normal* in the 1960s, and a downward trajectory from the 1984 through the 2004 edition (a period corresponding with a new wave of Jewish immigrants arriving from the former Soviet Union) culminating by crossing back into *strangeness* in the most recent edition. After peak *normalcy* in the 1950s, *Italian* has generally hovered around the midway point between *normal* and *strange*, suggesting little change in terms of integration into the mainstream since the 1940s.

FIGURE 7 ABOUT HERE

Figure 8 plots the scores for the next group of immigrants. This figure also demonstrates varied patterns of historical integration. *Cuban* has a fairly clear trajectory toward *normal*, crossing the threshold to *normalcy* during the 1960s, a time of large scale migration to the United States following the 1959 Cuban Revolution. Other than one dip into *strange* territory in 1991, coinciding with the collapse of Cuba's ally the Soviet Union, *Cuban* generally moves toward *normalcy*. *Muslim*, on the other hand, fluctuates widely between the poles of *normal* and *strange* on a nearly edition-by-edition basis through the 1970s. Beginning in the 1980s, *Muslim* stabilizes a bit just on the *strange* side of the binary, before moving to the *normal* side of the binary, albeit still close to the boundary, in the twenty-first century. It is interesting to note that this transition

to *normality* begins during the period of the September 11th attacks and the trend toward hyper-visibility of Muslims in America (Maghbouleh 2017). The position of *Chinese* on the *strange-to-normal* dimension appears to have three phases. A period of persistent *strangeness* up until 1950, straddling the border between *normal* and *strange* between 1950 and 2010 with the suggestion of a trend back toward *strangeness*, in particular in beginning in the 1980s which was a period of dramatic increase in Chinese immigration to the United States, and then a big jump in *normalcy* between in the second decade of the twenty-first century and following the Beijing Olympics of 2008 and the establishment of China's reputation as an emerging superpower. *Mexican's* position between *strange* and *normal* appears to have 4 distinct phases that we interpret in light of the periods in the Mexico-US immigration system. With the exception of the 1937 edition, *Mexican* was fairly strongly *strange* between 1922 and 1945, a period coinciding with repressive and punitive treatment of Mexicans in the United States, including mass deportations and anti-Mexican sentiment resulting from the economic difficulties of the times (Massey et al. 2002). Between 1945 and 1970, a period of *normalcy* overlaps with the US government's support of labor migration from Mexico through the bracero program (Massey et al. 2002). There was another period of *strangeness* between the 1975 and 1997 editions - during a time that Massey classifies as the undocumented period characterized by large scale restrictions on legal avenues to migration and increasing attention to border control as an issue of public concern (Massey et al. 2002). *Mexican* was consistently *normal* again in the 2004, 2011, and 2017 editions, a period of increased activism on behalf of undocumented immigrants and the introduction of programs to protect undocumented young people in particular.

FIGURE 8 ABOUT HERE

The results from these analyses of the position of *Catholic, Chinese, Cuban, Irish, Italian, Jewish, Mexican, and Muslim* between *normal* and *strange* shows a general trend toward immigrant incorporation through shifts in the mainstream, but very different trajectories for the particular groups we examined. Incorporation, was not a given or a slow and steady march to inclusion, however. Very few groups experienced a straight-line trajectory from *strange* to *normal*, even if most groups did tend toward *normalcy* over time. These findings make sense in light of dynamics of migration flows, national sentiment, and national and global politics.

VI. Discussion and Conclusions

The “remaking of the American mainstream” (Alba and Nee 2009) is crucial to the trajectory of immigrant incorporation. Applying methods of computational text analysis to a unique text corpus, we examined mainstream shifts that show immigrant incorporation at the level of the social and symbolic organization of society. Past research on individuals' attitudes, preferences, and practices has established that established, non-immigrant, Americans become comfortable and familiar with the languages, foods, and cultural and religious practices that immigrants

introduce to the social world. In some cases, established individuals develop a more expansive and less ethnocentric notion of what it means to be American. Using computational sociological methods, we scaled-up consideration of immigration's impact by shifting the focus from changes in established individuals to changes in the mainstream establishment.

The text data for an analysis of the mainstream must be selected carefully. Mainstream shifts occur and are reflected within workplace relationships; food and fashion; norms of communication; familiarity with different religions; and rituals. For this reason, our data consisted of Emily Post's *Etiquette* - a book that has been guiding Americans through high moments and everyday encounters for nearly a century. The corpus is unapologetically focused on the mundane everyday world and thus very well suited to revealing the scope of immigrant integration through mainstream shifts.

Like our theoretically-indicated data, our method is also attuned to theory, in particular theories that emphasize meaning and interpretation. Given the historical scope of the corpus, we examined mainstream shifts in relationship to long-standing groups who once were and may currently still be seen as immigrants and outsiders in the United States: Catholic, Chinese, Cuban, Irish, Italian, Jewish, Mexican, and Muslim groups. Based upon qualitative insights arising through reading the texts, for each group we analyzed mainstream shifts in representation, sentiment, and meaning in terms of position between being seen as more strange or more normal.

Although our findings do generally support the conclusion that there have been mainstream shifts indicating some incorporation of all of these groups, these changes must be interpreted in relationship to the fact that the mainstream remains overwhelmingly white and Anglo-American. While we observed increased representation, positive sentiment, and a movement from being seen as strange to being seen as normal on the aggregate, there is significant variation within groups. Some acquire consistent positivity and a view of normalness over time, but many others have less consistent trajectories of incorporation, including periods of exclusion and negativity that we interpret in light of relevant social, cultural, and geo-politics facts.

It could be argued that there is little that is new in the empirical findings of this research, but the objective of this paper was to apply new data and methods to the long-standing sociological topic of immigrant integration. Etiquette books, with an implicit relationship to the topic of immigration, are not an obvious source of data and the computational methods we employed, while grounded in theoretical work on immigrant incorporation, are atypical in their focus on immigrant incorporation through changes in the mainstream. Under these circumstances, returning results that are consistent with the existing theory and literature is an asset. This research demonstrates its value precisely by corroborating both the reality of mainstream shifts

in light of immigration and the persistence of racialized exclusions of immigrants and their descendents.

This research focuses on immigrant incorporation, but the techniques we developed could be adapted to a variety of questions related to representation, social sentiment, and symbolic inclusion. Thus, we contribute to sociological methods, and computational sociological research methods more broadly, by highlighting the value of a using theoretically-selected text corpus instead of relying on a sampling logic in corpus selection. We demonstrate the potential of computational sociological methods by moving beyond counting occurrences to establish representation, and instead adopting word embedding and sentiment analyses, which are much more interpretive computational tools, for the analysis of meaning.

Finally, we want to underscore the continued centrality of qualitative logics in the practice of computational social science (Nelson nd). Both human and machine reading and text analysis were crucial to this project - a mutual give and take involving thoroughly human interpretive logics of the social scientists and fully machine neural networks of word embeddings were combined in a computational grounded theoretic approach (Nelson 2020). Qualitative content analysis served two functions in our analyses: to provide a confirmation to the validity of our embedding techniques and also to provide rich insight into the inner-workings of immigrant incorporation. Computational text analysis both guided and was guided by human content analyses. Computational approaches specified where to look in the text, but our readings of the texts both confirms and clarifies the embeddings techniques.

VII. References

- Abrutyn, Seth and Michael J. Carter. 2015. "The Decline in Shared Collective Conscience as Found in the Shifting Norms and Values of Etiquette Manuals." *Journal for the Theory of Social Behaviour* 45(3): 352-376.
- Alba, Richard D. 2005. "Bright vs. Blurred Boundaries: Second-Generation Assimilation and Exclusion in France, Germany, and the United States." *Ethnic and Racial Studies* 28:20-49.
- Alba, R. and Foner, N. 2015. *Strangers No More: Immigration and the Challenges of Integration in North America and Western Europe*. Princeton, NJ: Princeton University Press.
- Alba, R. and Nee, V. 2009. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.
- Alexander, Jeffrey C. 2001. "Theorizing the "Modes of Incorporation": Assimilation, Hyphenation, and Multiculturalism as Varieties of Civil Participation." *Sociological theory* 19(3): 237-249.
- . 2006. *The Civil Sphere*. Oxford ; New York: Oxford University Press.
- Arditi, Jorge. 1999. "Etiquette Books, Discourse and the Deployment of an Order of Things." *Theory, Culture & Society* 16(4):25-48.
- Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5: 135–146.
- Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA.: Harvard University Press.
- Brodin, Karen. 1998. *How Jews Became White Folks and What That Says About Race in America*. New Brunswick, N.J.: Rutgers University Press.
- Census of the United States. 2016. "Frequently Occurring Surnames from the 2010 Census." Retrieved February 24, 2020 (https://www.census.gov/topics/population/genealogy/data/2010_surnames.html)
- Claridge, Laura. 2009. *Emily Post: Daughter of the Gilded Age, Mistress of American Manners*. New York, NY: Random House.

Drouhot, Lucas G., and Victor Nee. 2019. "Assimilation and the Second Generation in Europe and America: Blending and Segregating Social Dynamics Between Immigrants and Natives." *Annual Review of Sociology* 45: 177-199.

Dumenil, Lynn. 1991. "The Tribal Twenties: 'Assimilated' Catholics' Response to Anti-Catholicism in the 1920s." *Journal of American Ethnic History* 11(1): 21-49.

El-Ebshihy, Alaa, Nagwa El-Makky, and Khaled Nagi. 2018. "Using Google Books Ngram in Detecting Linguistic Shifts over Time." Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management 1: 330-337.

Elias, Norbert. 2000. *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. Oxford, England: Blackwell Publishers.

Ellis, Nick C. 2019. "Essentials of a Theory of Language Cognition." *The Modern Language Journal* 103: 39-60.

Fellbaum, Christiane. 1998. "A Semantic Network of English: The Mother of all WordNets." Pp 137-148 in *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Springer.

Firth, J. R. 1957. "A Synopsis of Linguistic Theory, 1930-1955." Pp. 168-205 in *Studies in Linguistic Analysis*, edited by J. R. Firth. Oxford, England: Blackwell.

Garvin, Paul L. 1962. "Computer Participation in Linguistic Research." *Language* 38(4): 385-389.

Giuliani, Luigi, Herman Brinkman, Geert Lernout, and Marita Mathijssen, eds. 2006. *Texts in Multiple Versions: Histories of Editions* (Vol. 5). Amsterdam, Netherlands: Rodopi.

Goffman, Erving. 1963. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York, NY: Free Press of Glencoe.

Hamilton, William L., Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. "Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora." Paper presented at the Conference on Empirical Methods in Natural Language Processing, November.

Hemphill, C. Dallett. 1999. *Bowing to Necessities: A History of Manners in America, 1620-1860*. New York, NY: Oxford University Press.

Ignatiev, Noel. 2009. *How the Irish Became White*. New York: Routledge.

Jacobs, Laura. 2013. "Emily Post's Social Revolution." *Vanity Fair*, December 2001. Retrieved August 19, 2020 (<https://archive.vanityfair.com/article/2001/12/emily-posts-social-revolution>).

Jiménez, Tomás. 2017. *The Other side of Assimilation: How Immigrants are Changing American Life*. Univ of California Press.

Jiménez, Tomás R., and Adam L. Horowitz. 2013. "When White is Just Alright: How Immigrants Redefine Achievement and Reconfigure the Ethnoracial Hierarchy." *American Sociological Review* 78(5): 849-871.

Joulin, Armand, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Herve Jégou, and Tomas Mikolov. 2016. "FastText.zip: Compressing Text Classification Models." Cornell University. Retrieved August 19, 2020 from (<https://arxiv.org/abs/1612.03651>).

Kroneberg, Clemens and Andreas Wimmer. 2012. "Struggling over the Boundaries of Belonging: A Formal Model of Nation Building, Ethnic Closure, and Populism." *American Journal of Sociology* 118:176-230.

Lee, Erika. 2003. *At America's Gates : Chinese Immigration During the Exclusion Era, 1882-1943*. Chapel Hill: University of North Carolina Press.

Lees-Maffei, Grace. 2012. "Small Change? Emily Post's Etiquette (1922-2011)." Pp. 217-248 in *Must Read: Rediscovering American Bestsellers*, edited by S. Churchwell and T. R. Smith. New York, NY: Continuum International Publishing Group.

Lees-Maffei, Grace. 2014. *Design at Home: Domestic Advice Books in Britain and the USA since 1945*. London, England: Routledge.

Lenci, Alessandro. 2018. "Distributional Models of Word Meaning." *Annual Review of Linguistics* 4: 151–171.

Maghbouleh, Neda. 2017. *The Limits of Whiteness: Iranian Americans and the Everyday Politics of Race*. Stanford University Press.

Massey, Douglas S., Jorge Durand, and Nolan J. Malone. 2002. *Beyond Smoke and Mirrors : Mexican Immigration in an Era of Economic Integration*. New York: Russell Sage Foundation.

Massey, Douglas S. and Magaly Sánchez. 2010. *Brokered Boundaries : Creating Immigrant Identity in Anti-Immigrant Times*. New York: Russell Sage Foundation.

Mennell, Stephen. 2007. *The American Civilizing Process*. Cambridge, England: Polity.

NLTK Project 2020. "Natural Language Toolkit." Retrieved February 24, 2020 (<https://www.nltk.org/index.html>)

Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1): 3-42.

Nelson, Laura K. Forthcoming. "Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South." *Poetics*.

Prabowo, Rudy, and Mike Thelwall. 2009. "Sentiment Analysis: A Combined Approach." *Journal of Informetrics*, 3(2): 143-157.

Portes, Alejandro Rumbaut Ruben G. 1996. *Immigrant America: A Portrait*. Berkeley: University of California Press.

Portes, Alejandro, and Robert L. Bach. 1985. *Latin Journey: Cuban and Mexican Immigrants in the United States*. Univ of California Press.

Portes, Alejandro and Ruben G. Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley: University of California Press.

Portes, Alejandro and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and its Variants." *The ANNALS of the American Academy of Political and Social Science* 530:22.

Post, Emily. 1937. *Etiquette: The Blue Book of Social Usage, 5th edition*. New York, NY: Funk & Wagnalls Company.

Post, Emily. 1956. *Etiquette: The Blue Book of Social Usage, 9th edition*. New York, NY: Funk & Wagnalls Company.

Post, Peggy. 1997. *Emily Post's Etiquette. 16th ed..* New York: HarperCollins Publishers.

Post, Peggy, Anna Post, Lizzie Post, and Daniel Post Senning. 2011. *Emily Post's Etiquette: Manners for a New World*. New York: HarperCollins.

- Roediger, David R. 2005. *Working Toward Whiteness : How America's Immigrants Became White : the Strange Journey from Ellis Island to the Suburbs*. New York: Basic Books.
- Schachter, Ariela. 2016. "From “Different” to “Similar” an Experimental Approach to Understanding Assimilation." *American Sociological Review* 81(5): 981-1013.
- Schinkel, Willem. 2018. "Against ‘Immigrant Integration’: For an End to Neocolonial Knowledge Production." *Comparative Migration Studies* 6(1): 1-17.
- Stoltz, Dustin S. and Marshall A. Taylor. 2019. “Concept Mover’s Distance: Measuring Concept Engagement via Word Embeddings in Text.” *Journal of Computational Social Science* 2: 293-313.
- Takaki, Ronald T. 1993. *A Different Mirror : A History of Multicultural America*. Boston: Little, Brown & Co.
- Taylor, Marshall A. and Dustin S. Stoltz. 2020. “Integrating Semantic Directions with Concept Mover’s Distance to Measure Binary Concept Engagement.” *Journal of Computational Social Science*. <https://doi.org/10.1007/s42001-020-00075-8>
- Voyer, Andrea. 2013a. "Notes on a Cultural Sociology of Immigrant Incorporation." *American Journal of Cultural Sociology* 1(1): 26-41.
- Voyer, Andrea M. 2013b. *Strangers and Neighbors: Multiculturalism, Conflict, and Community in America*. New York: Cambridge University Press.
- Wimmer, A. 2008. "The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory." *American Journal of Sociology* 113:970-1022.
- Wimmer, Andreas. 2009. "Herder's Heritage and the Boundary-Making Approach in Immigrant Societies." *Sociological Theory* 27:244-270.
- Wouters, Cas. 2007. *Informalization : Manners and Emotions Since 1890*. Los Angeles, CA: SAGE Publications.
- Ye, Shuchu Han, Yifan Hu, Baris Coskun, Meizhu Liu, Hong Qin, and Steven Skiena. 2017. "Nationality Classification Using Name Embeddings." Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17). Retrieved February 24, 2020 (<https://doi.org/10.1145/3132847.3133008>)

Table 1: Document Similarity Matrix Over Time by Author

| | 1922 | 1927 | 1931 | 1934 | 1937 | 1940 | 1942 | 1945 | 1950 | 1955 | 1960 | 1965 | 1968 | 1975 | 1984 | 1992 | 1997 | 2004 | 2011 | 2017 | |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|--|
| 1922 | 1 | | | | | | | | | | | | | | | | | | | | |
| 1927 | .816 | 1 | | | | | | | | | | | | | | | | | | | |
| 1931 | .804 | .968 | 1 | | | | | | | | | | | | | | | | | | |
| 1934 | .805 | .969 | .996 | 1 | | | | | | | | | | | | | | | | | |
| 1937 | .734 | .877 | .900 | .898 | 1 | | | | | | | | | | | | | | | | |
| 1940 | .746 | .891 | .914 | .911 | .988 | 1 | | | | | | | | | | | | | | | |
| 1942 | .737 | .867 | .869 | .868 | .925 | .936 | 1 | | | | | | | | | | | | | | |
| 1945 | .727 | .842 | .866 | .864 | .905 | .919 | .914 | 1 | | | | | | | | | | | | | |
| 1950 | .714 | .825 | .852 | .849 | .890 | .905 | .891 | .958 | 1 | | | | | | | | | | | | |
| 1955 | .723 | .820 | .813 | .814 | .830 | .847 | .857 | .893 | .911 | 1 | | | | | | | | | | | |
| 1960 | .690 | .797 | .817 | .820 | .831 | .847 | .837 | .896 | .911 | .943 | 1 | | | | | | | | | | |
| 1965 | .591 | .710 | .678 | .678 | .685 | .700 | .701 | .734 | .746 | .820 | .774 | 1 | | | | | | | | | |
| 1968 | .512 | .594 | .628 | .613 | .635 | .656 | .642 | .679 | .692 | .683 | .705 | .752 | 1 | | | | | | | | |
| 1975 | .509 | .640 | .642 | .644 | .651 | .666 | .623 | .660 | .665 | .676 | .708 | .751 | .733 | 1 | | | | | | | |
| 1984 | .490 | .568 | .583 | .583 | .593 | .606 | .603 | .638 | .643 | .656 | .673 | .721 | .702 | .816 | 1 | | | | | | |
| 1992 | .462 | .552 | .530 | .531 | .541 | .552 | .556 | .584 | .587 | .646 | .619 | .778 | .650 | .760 | .837 | 1 | | | | | |
| 1997 | .446 | .517 | .537 | .535 | .552 | .563 | .560 | .598 | .603 | .604 | .634 | .658 | .683 | .753 | .823 | .877 | 1 | | | | |
| 2004 | .357 | .407 | .430 | .428 | .445 | .456 | .453 | .482 | .485 | .486 | .511 | .526 | .541 | .602 | .653 | .673 | .745 | 1 | | | |
| 2011 | .233 | .303 | .263 | .265 | .271 | .277 | .376 | .294 | .293 | .357 | .311 | .432 | .322 | .381 | .417 | .506 | .456 | .581 | 1 | | |
| 2017 | .343 | .378 | .387 | .387 | .397 | .407 | .406 | .433 | .431 | .449 | .458 | .503 | .477 | .564 | .617 | .651 | .670 | .836 | .685 | 1 | |

Key:

| | | |
|--|--|--|
| Emily Post | Peggy Post | Author Transition |
| Elizabeth L. Post | Lizzie Post | |

Table 2. Discussion of Quinceñara and Sweet-sixteen Tradition, 1997

| Quinceñara | Sweet-sixteen Parties |
|---|--|
| <p>The celebration of a Latin American girls fifteenth birthday, a quinceñara, often simply called a quince, can be either a party or a religious celebration followed by a party, depending on the practice of the Roman Catholic Church in the area...</p> <p>Not dissimilar in concept to a debutante ball, the celebration commemorates coming-of-age. or a young girl's passage into womanhood. A highlight of the party usually includes the tradition of the birthday girl waltzing with her father before general dancing begins.</p> <p>Dress Very often the party is formal, with guests wearing either black tie or white tie attire, as noted on the invitation, and the young girl wearing an elaborate, formal gown.</p> <p>Gifts Gifts are expected, either money or personal items, depending on the tradition of the area, and usually are not opened during the party, which follows the form of a ball or dance rather than a traditional birthday celebration (Post 1997: 566-567).</p> | <p>A sixteenth birthday is a big milestone—really the division between childhood and young adulthood. Therefore it is often celebrated more elaborately than other birthdays, with a sweet-sixteen party...</p> <p>(...)</p> <p>Invitations may be telephoned, but when the celebration is more formal, they are usually written on decorated, fill-in commercial invitations, which can be found, if you wish, specifically for sweet-sixteen parties...</p> <p>All invitations should have "R.s.v.p." on them, followed by a telephone number. Telephoning a response seems easier to phone-prone youngsters than writing a reply, and this will elicit more and quicker answers.</p> <p>Refreshments If you are serving a meal—luncheon or dinner—the menu should simply be the favorite food of the birthday girl, although nothing so exotic that it will not appeal to the majority of the guests... (Post 1997: 567)</p> |

Table 3. Discussion of Quinceñara, 2011

Invitations are issued by the honoree's parents or by the honoree and her parents. For an informal quinceanera a phone call, personal note, or informal card are good invitational choices. But for a formal event invitations may be printed or engraved and might include a "Black tie" or "White tie" notation.

A printed, formal invitation to the religious service would also include a reception card inviting guests to the quinceanera party afterward. Since the invitation itself is to a religious service, it does not include an RSVP, but the reception card does.

It's customary to take gifts to a quinceanera party, but presents aren't opened at a formal ball. Religious items such as rosaries or crosses are appropriate and traditional; monetary gifts and personal items for teenage girls are also popular. The young honoree must send written thank-you notes to all friends and family who gave her a gift. (Post et al. 2011: 318).

Table 4: Count and Proportion (in Parentheses) of Chunks Selected from Full Corpus that Compose Pseudo Corpus for Each Immigrant Group

| Edition | 1922 | 1927 | 1934 | 1937 | 1940 | 1942 | 1945 | 1950 | 1956 | 1960 | 1965 | 1969 | 1975 | 1984 | 1992 | 1997 | 2004 | 2011 | 2017 |
|----------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|
| Italian | 8 (.016) | 10 (.015) | 10 (.015) | 12 (.015) | 8 (.009) | 13 (.014) | 9 (.013) | 10 (.014) | 12 (.016) | 10 (.013) | 8 (.011) | 8 (.012) | 7 (.008) | 8 (.009) | 9 (.011) | 7 (.008) | 10 (.010) | 12 (.013) | 7 (.010) |
| Jewish | 0 (.000) | 1 (.002) | 1 (.001) | 2 (.002) | 1 (.001) | 2 (.002) | 4 (.006) | 2 (.003) | 5 (.007) | 6 (.008) | 4 (.006) | 5 (.007) | 7 (.008) | 5 (.005) | 8 (.009) | 7 (.008) | 5 (.005) | 9 (.010) | 7 (.010) |
| Irish | 6 (.012) | 8 (.012) | 6 (.009) | 9 (.011) | 11 (.013) | 10 (.011) | 10 (.014) | 8 (.011) | 11 (.014) | 11 (.015) | 7 (.010) | 7 (.010) | 13 (.014) | 8 (.009) | 10 (.012) | 12 (.014) | 13 (.013) | 7 (.008) | 8 (.011) |
| Catholic | 3 (.006) | 3 (.005) | 5 (.007) | 8 (.010) | 10 (.012) | 8 (.009) | 7 (.010) | 10 (.014) | 13 (.017) | 12 (.016) | 7 (.010) | 6 (.009) | 10 (.011) | 11 (.012) | 10 (.012) | 12 (.014) | 15 (.015) | 12 (.013) | 9 (.012) |
| Mexican | 4 (.008) | 5 (.008) | 2 (.003) | 4 (.005) | 4 (.005) | 8 (.009) | 5 (.007) | 5 (.007) | 6 (.008) | 6 (.008) | 6 (.008) | 8 (.012) | 7 (.008) | 9 (.010) | 7 (.008) | 5 (.006) | 10 (.010) | 8 (.009) | 7 (.010) |
| Chinese | 5 (.010) | 5 (.008) | 8 (.012) | 7 (.008) | 6 (.007) | 6 (.007) | 6 (.008) | 8 (.011) | 10 (.013) | 10 (.013) | 9 (.013) | 7 (.010) | 10 (.011) | 13 (.014) | 6 (.007) | 8 (.009) | 10 (.010) | 6 (.007) | 6 (.008) |
| Cuban | 5 (.010) | 4 (.006) | 4 (.006) | 4 (.005) | 5 (.006) | 7 (.008) | 4 (.006) | 6 (.009) | 4 (.005) | 4 (.005) | 9 (.013) | 6 (.009) | 5 (.006) | 7 (.008) | 2 (.002) | 4 (.005) | 6 (.006) | 3 (.003) | 3 (.004) |
| Muslim | 0 (.000) | 0 (.000) | 0 (.000) | 1 (.001) | 1 (.001) | 1 (.001) | 1 (.001) | 0 (.000) | 4 (.005) | 1 (.001) | 1 (.001) | 2 (.003) | 3 (.003) | 9 (.010) | 11 (.013) | 7 (.008) | 8 (.008) | 7 (.008) | 7 (.010) |

Figure 1. Representation: Proportion of Surnames Represented in *Etiquette*

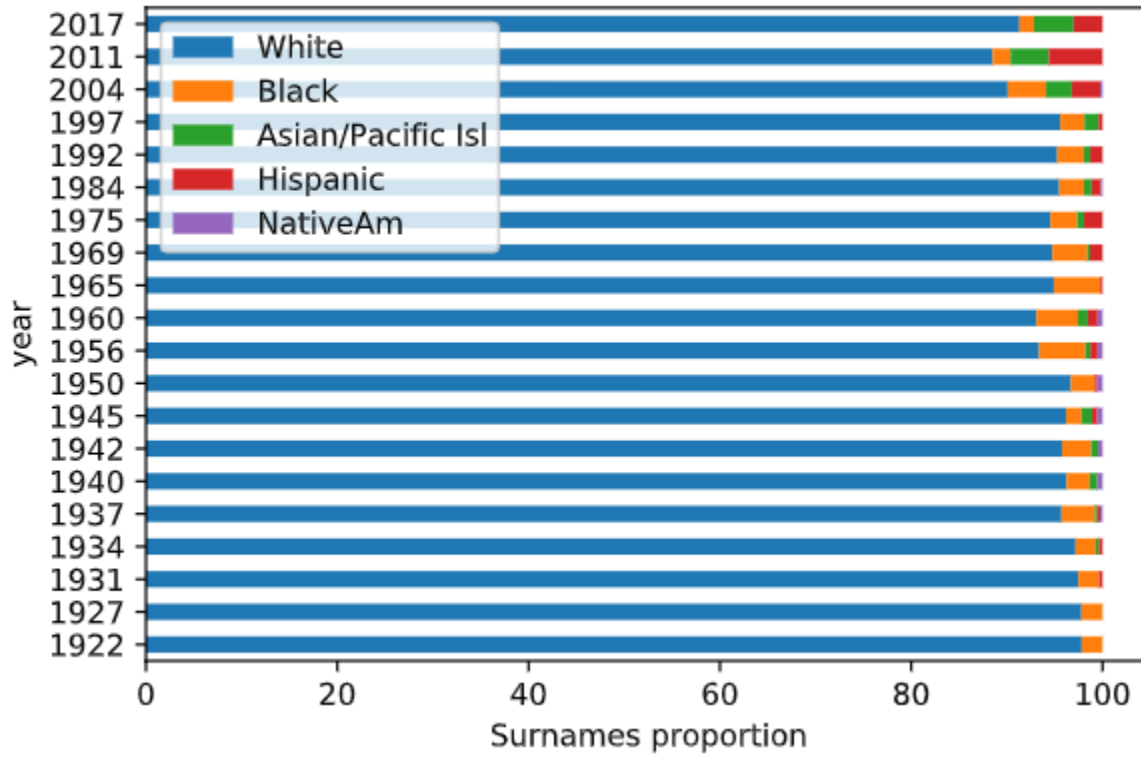


Figure 2. Representation: Count of Surnames by Group Over Time

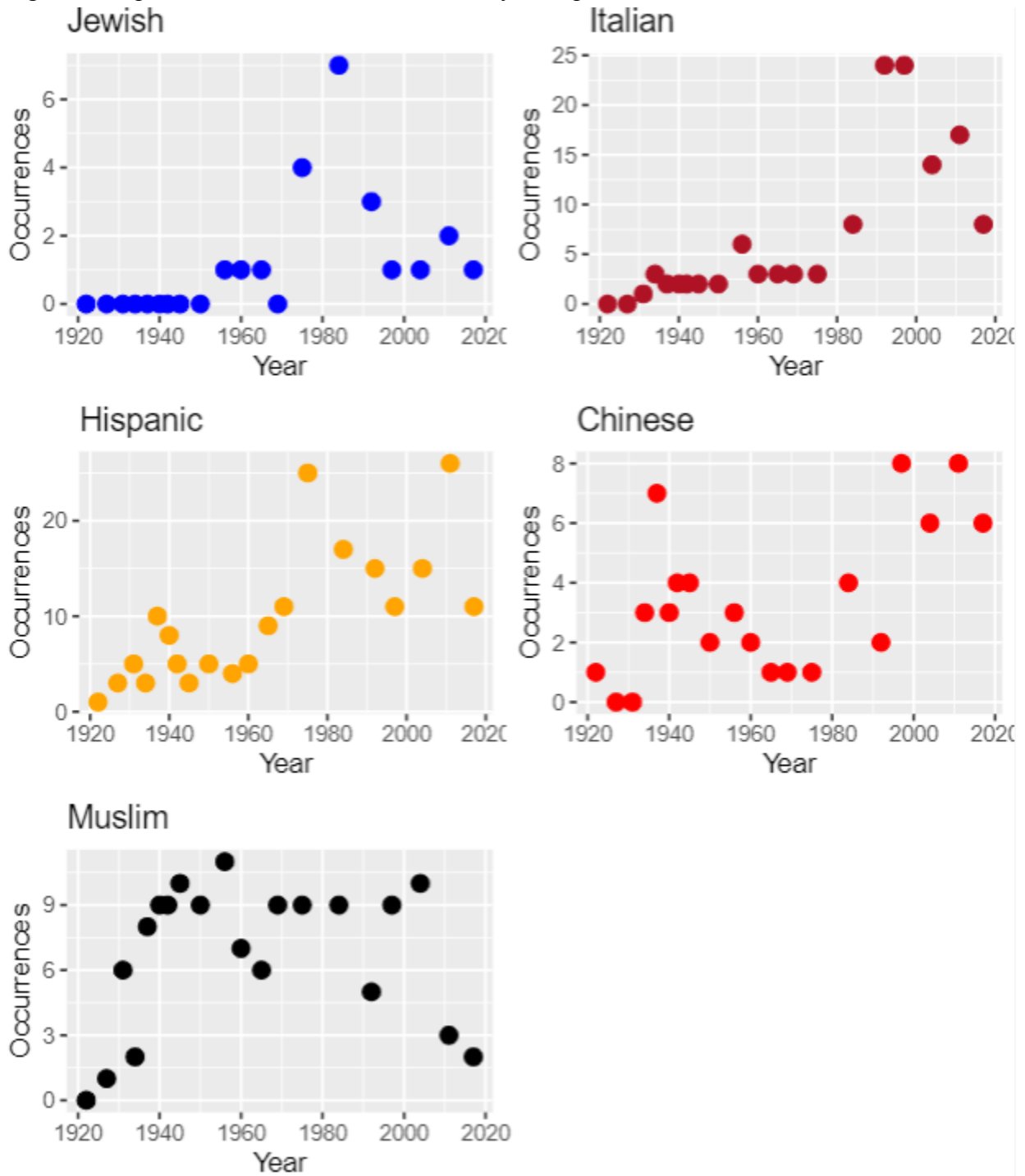


Figure 3. Sentiment: Positive versus Negative Over Time for All Pseudo Corpora

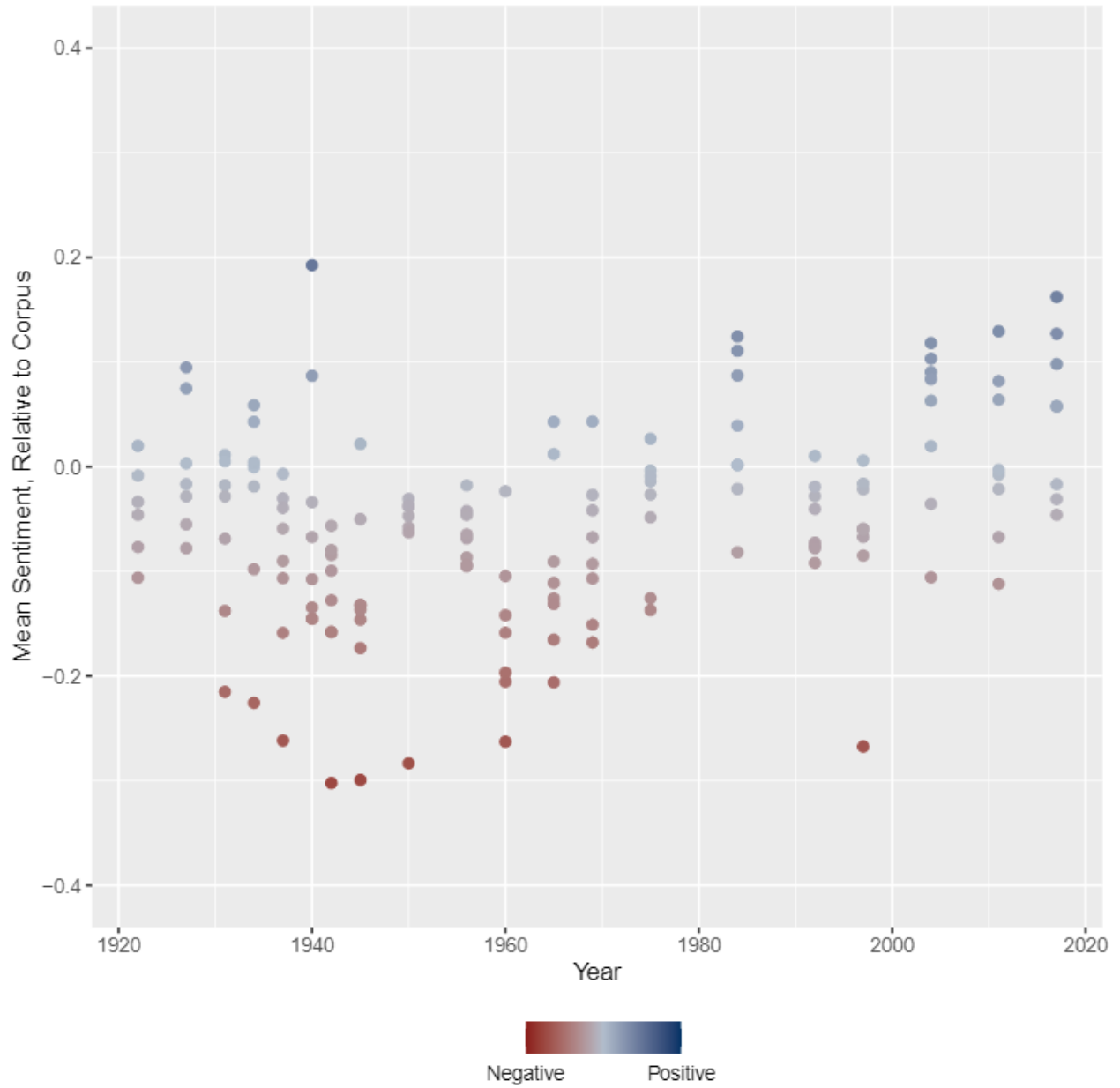


Figure 4. Sentiment: Positive versus Negative Over Time for Group 1

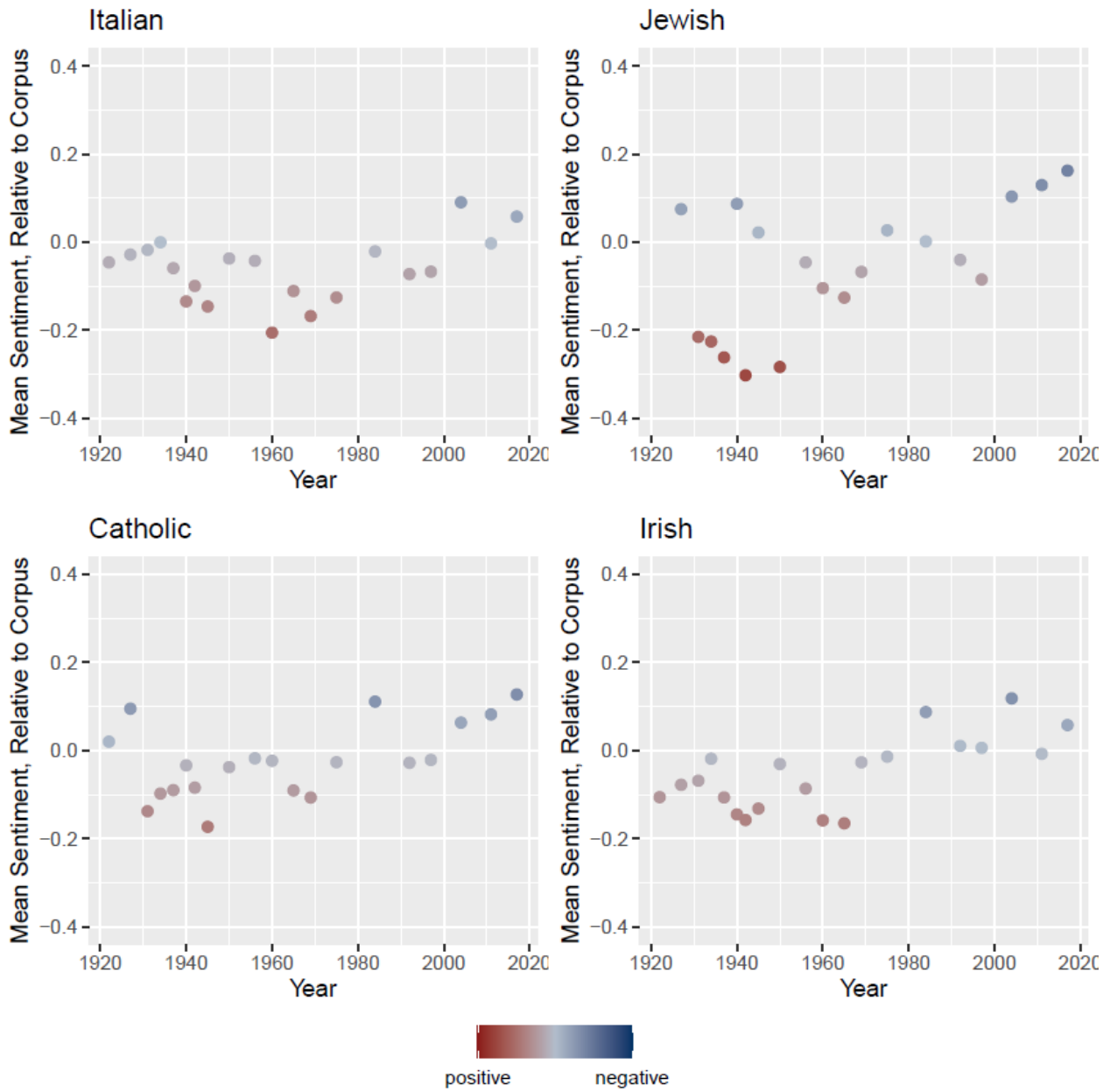


Figure 5: Sentiment: Positive versus Negative Over Time for Group 2

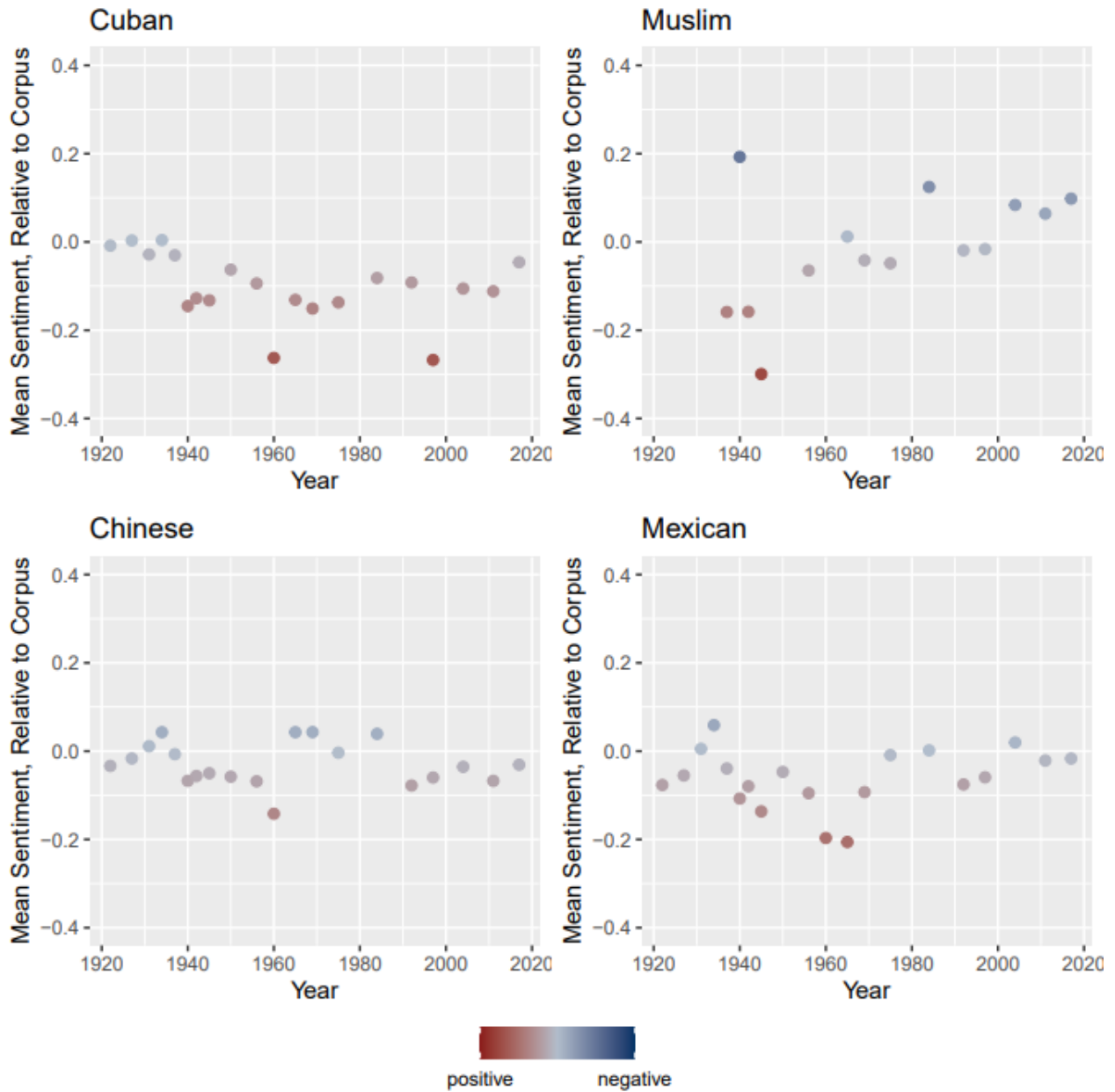


Figure 6. Meaning: Normal versus Strange Over Time for All Pseudo Corpora

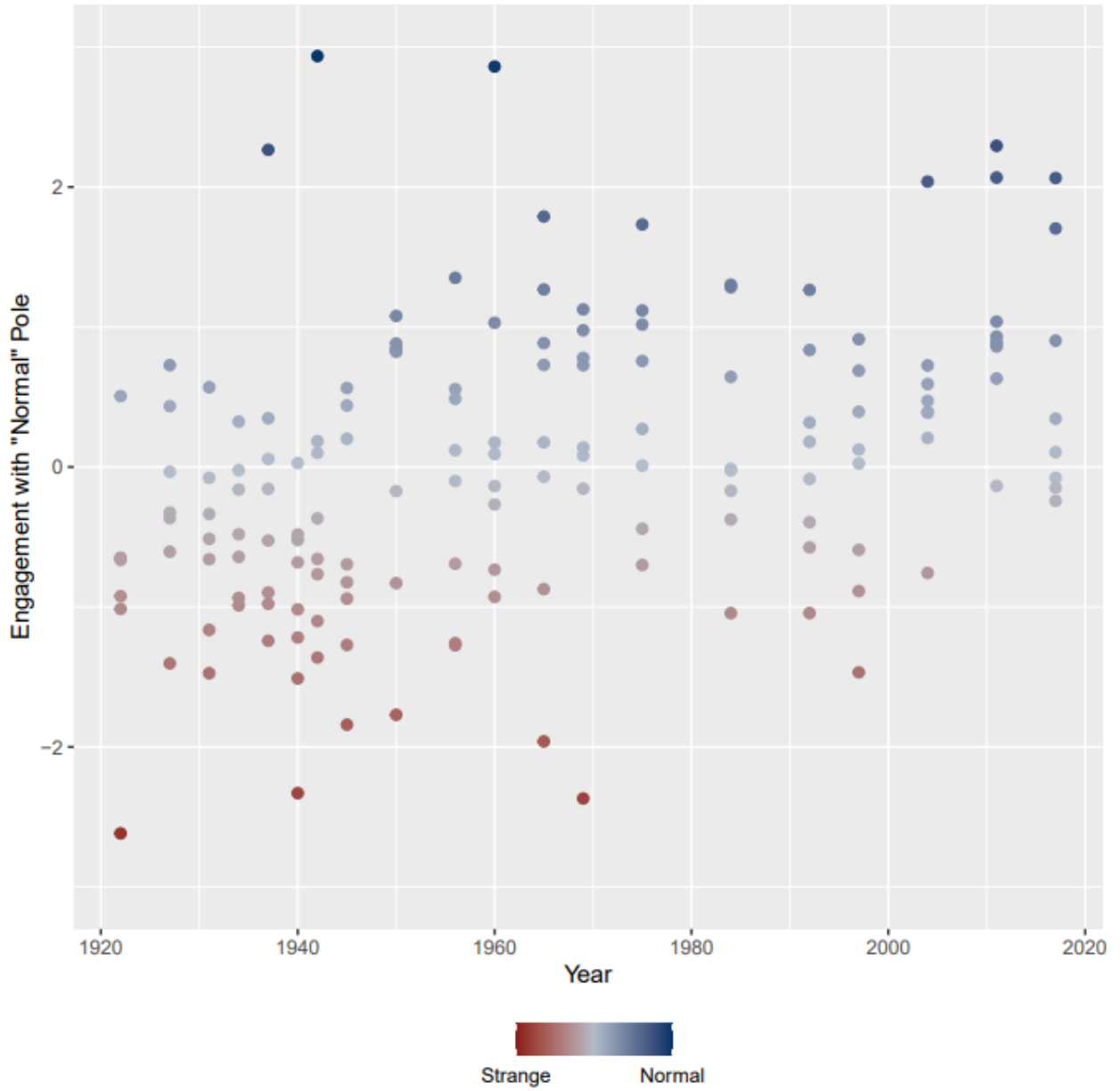


Figure 7. Meaning: Normal versus Strange Over Time for Group 1

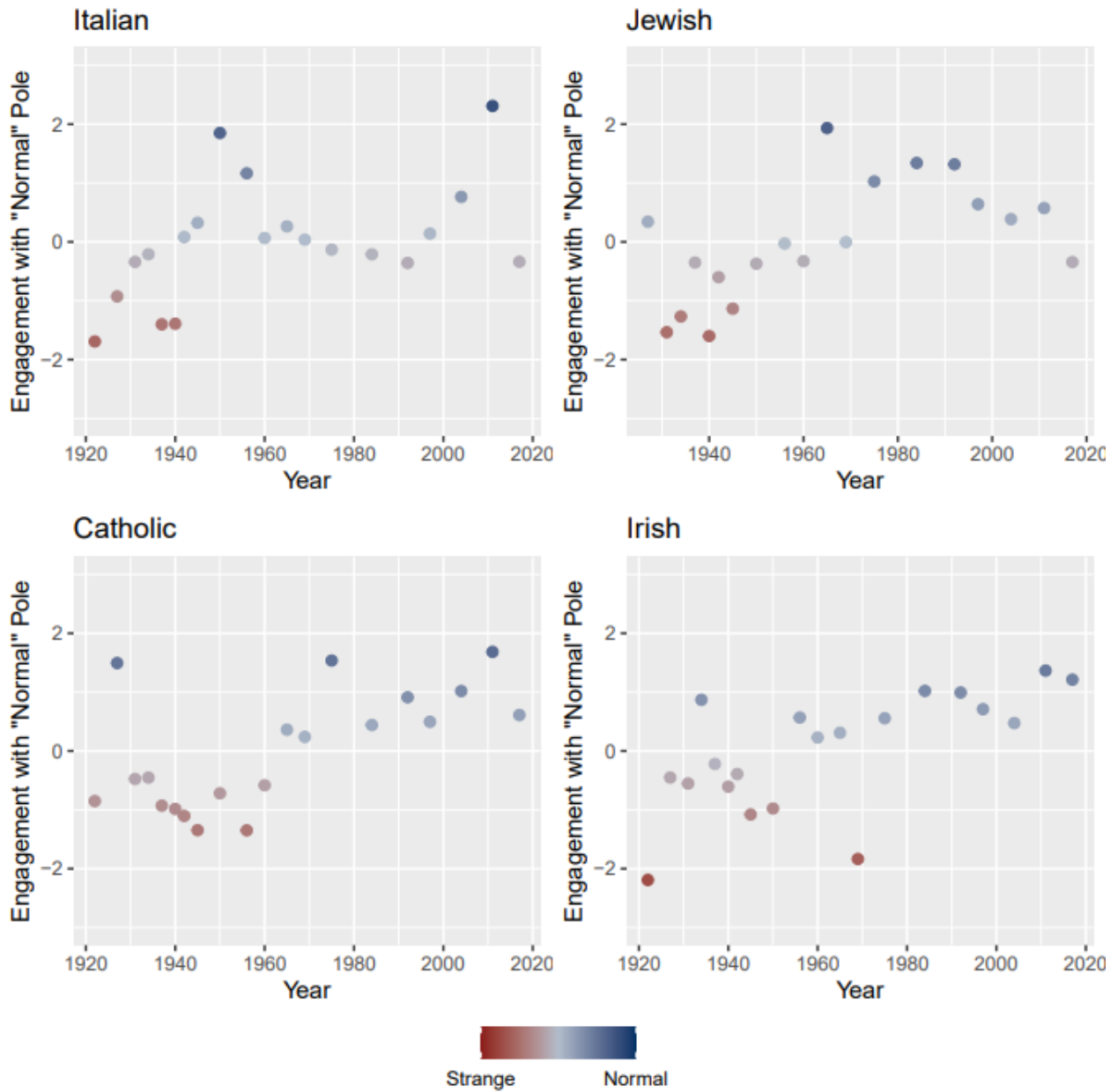


Figure 8. Meaning: Normal versus Strange Over Time for Group 2

