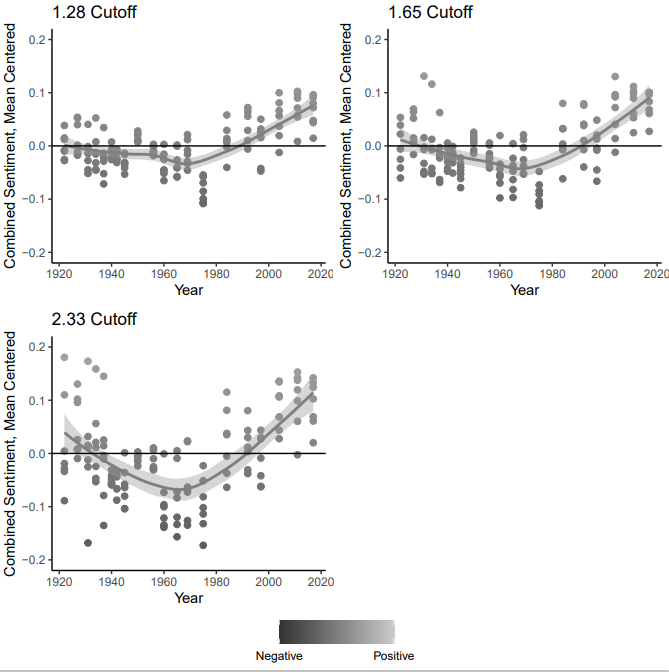
**Appendix A: Salience Cut Off**

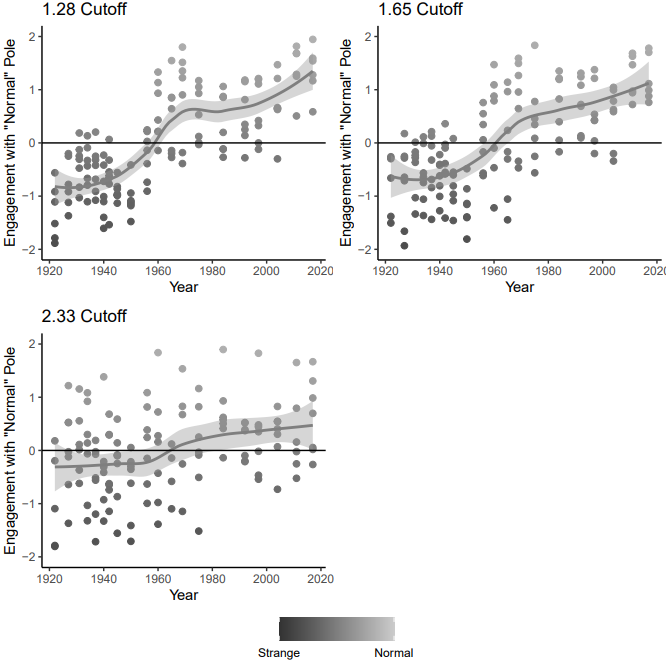
Appendix A assesses the tradeoffs associated with selecting a salient cutoff. The general premise is that a higher cutoff level will select chunks that more closely relate to the focal group concepts in question, at the expense of textual robustness. Recent research using CMDist to select “salient” texts have considered scores of “1” as “high,” (Carbone and Mijs 2022). Yet, to our knowledge, there is no standardized best-practice for assessing this potentially-important tradeoff in the context of varying text types and theoretical goals. We created the corpus presented in the paper after considering results using three salient cutoffs: 1.28, 1.65, and 2.33. These scores roughly correspond to the 90th percentile, the 95th percentile, and the 99th percentiles on a single tail of a normalized distribution. These cutoffs correspond nicely with the standardized distribution of CMDist scores.

Figure A1 and A2 demonstrate that the broad result trends are not sensitive to the salience cutoff. Figure A1 display results from the sentiment analyses for each cutoff point while A2 shows the strange to normal trend using each of the corpora. The general trend is consistent across all the three specifications. Importantly, results from the highest cutoff display a high degree of variation, with outlying scores disproportionately affecting the overall trends. These variations and influential outliers suggested to us that the 2.33 cutoff does not yield enough text to provide robust estimates across pseudoeditions. By contrast, standardized scores from the lower cutoff are more robust to the influential scores of particular pseudoeditions.

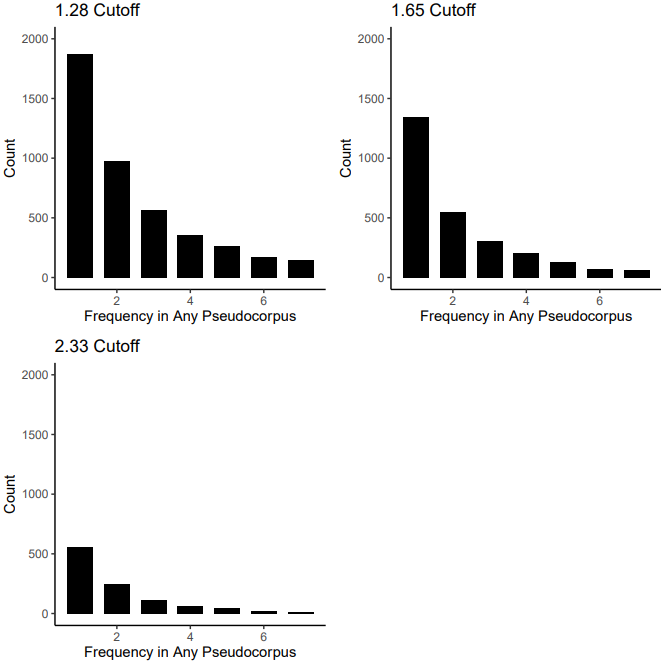
Figures A3 further demonstrate the specificity component of the tradeoff. The figures present the frequency that any particular chunk occurs in any of the pseudocorpora, varying across the three salience cutoffs. The general premise is that a smaller cutoff yield more text chunks - robust to year/edition change - at the expense of greater overlap across pseudocorpora. When using a salient cutoff of 1.28, the average chunk that is selected in any pseudocorpus is selected in 2.532 pseudocorpora. When raising our cutoff to 1.65, that mean falls to 2.163. Changing the cutoff to 2.33 yields diminishing returns, with the mean only falling to 2.057. This small return is achieved at the expense of over half of the chunks, introducing a risk that outlier chunks will unduly influence standardized estimates. To be clear, overlap is not in-and-of-itself concerning. To some degree, overlap is expected given the theory of mainstream shifts this analysis is based on. Nonetheless, the decreasing mean present in the figures nicely illustrates that increasing the cutoff for inclusion increases specificity with a diminishing return at the expense of textual robustness, a familiar tradeoff for practitioners of random sampling.



*Figure A1. Sentiment analyses of all pseudocorpora with salience cutoffs corresponding to 90th percentile (1.28), 95th percentile (1.65), and 99th percentile (2.38) on a single tail of a normalized distribution. The general trend is consistent across all the three specifications, but results from the highest cutoff display a high degree of variation, with outlying scores disproportionately affecting the overall trends.*



*Figure A2. Strange-to-normal analyses of all pseudocorpora with salience cutoffs corresponding to 90th percentile (1.28), 95th percentile (1.65), and 99th percentile (2.38) on a single tail of a normalized distribution. The general trend is consistent across all the three specifications, but results from the highest cutoff display a high degree of variation, with outlying scores disproportionately affecting the overall trends.*



*Figure A3. Frequency that any particular chunk occurs in any of the pseudocorpora, varying across salience cutoffs corresponding to 90th percentile (1.28), 95th percentile (1.65), and 99th percentile (2.38) on a single tail of a normalized distribution. A smaller cutoff yield more text chunks at the expense of greater overlap across pseudocorpora. Changing the cutoff to 2.33 yields diminishing returns at the expense of over half of the chunks, introducing a risk that outlier chunks will unduly influence standardized estimates.*

**Appendix B: Focal Group Concept Verification**

Appendix B presents a selection of text excerpts represented by the salient chunks. To locate the excerpts, we examined the word tokens in the salient chunks that were selected for the main analyses for each focal group, per edition and located the original text excerpts for 2-3 chunks per group. We provide some brief notes regarding the connection between chunks and the group concepts. However, the background “logic” of the embedding algorithm is not something a researcher is access so our explanations are informed interpretations. We believe that the value of our method is precisely in leveraging this aspect of embeddings and machine learning in general to derive new interpretive insights into the text corpus.

**Catholic**

1922 Edition, Chunk 73, Located on page 5

*Note: Explicit Catholic reference. Implicit idea that readers are more likely to make a mistake by calling a priest a minister (instead of the other way around). Also note the Irish last name for the priest.*

Very few people in polite society are introduced by their formal titles. A hostess says, Mrs. Jones, may I present the Duke of Overthere?” or “Lord Blank?” ; never “his Grace “or “his Lordship.” The Honorable is merely Mr. Lordson, or Mr. Holdoffice. A doctor, a judge, a bishop, are addressed and introduced by their titles. The Protestant clergy are usually Minister unless they hold the title of Doctor, or Dean, or Canon, in which case the surname is added: “Dean Wood,” “Doctor Starr,” “Canon Cope.” A Catholic priest is “Father Kelly.” To call him Minister is an inexcusable offense. An introduction to a Catholic archbishop, “Your Grace, may I present-----------?”

1984 Edition, Chunk 417, Located on page 71

*Note: Explicit reference but no explicit mention of “Catholic.” Connection to foreign travel.*

An Audience with the Pope

Any American tourist visiting Rome can be granted an audience with the Pope, for although there are often hundreds of people in a day who wish an audience no one with a proper introduction is denied. Obviously only relatively few can be granted one of the three types of audience that are considered to be personal; group or collective audiences are arranged for the great majority.

Requests by Americans for these group audiences as well as for the personal ones are cleared by the North American College and then sent to the Office of the Master of the Chamber

2004 Edition, Chunk 3315, Located on page 508

*Note: Implicit reference. Rosaries are Catholic.*

perfume; or a scarf, belt, or other fashion accessory.

For quinceanera parties

For quinceanera—a girl's fifteenth-birthday celebration in Latin cultures—religious items such as rosaries or crosses are appropriate. Monetary gifts and personal items for teenage girls are also popular gifts. (See also Chapter 16. page 214:"Quinceanera.‘‘)

**Chinese**

1922 Edition, Chunk 16, Located in introduction (page i)

*Note: Explicit reference to Chinese culture/thought*

The Chinese sage, Confucius, could not tolerate the suggestion that virtue is in itself enough without politeness, for he viewed them as inseparable and "saw courtesies as coming from the heart," maintaining that "when they are practised with all the heart, a moral elevation ensues."

People who ridicule etiquette as a mass of trivial and arbitrary conventions, "extremely troublesome to those who practise them and insupportable to everybody else," seem to forget the long, slow progress of social intercourse in the upward climb of man

1965 Edition, Chunk 1567, Located on page 253

*Note: Explicit reference. Suggestion that Chinese food is not a common meal, but part of a “themed” dinner.*

enamelware also can be purchased for individual servings— little casseroles with covers, perfect for hot chowder, stew, meat pies, etc.

Types of centerpieces and decorations are unlimited, but you will receive more compliments if they are appropriate to the meal or the season. For example, Chinese lanterns strung above the table and a centerpiece of gay paper parasols might accompany a Chinese dinner. Or a fishnet tablecloth with colorful felt cutouts of fish is decorative on the table set for a lobster dinner.

2017 Edition, Chunk 4041, Located on page 570

*Note: Explicit reference. Appropriation of a “foreign” wedding tradition.*

elements from faiths other than their own. One example is the Jewish tradition of having mothers and fathers in the processional. Some couples have adapted the Greek Orthodox crowning, or wreath, ceremony (see “Wedding Customs in Many Faiths," page 672) as a symbol of their unity. Another gracious addition to some American weddings is the Chinese tea ceremony, during which the couple offers cups of sweet tea to each other's families. Native American ceremonial sand painting has inspired sand-blending rituals to signify the mingling of two individuals and their families

**Irish**

1927 edition, chunk 109, Located on page 5

*Note: Implicit reference. Irish last name of Catholic priest.*

A Catholic priest is “Father Kelly.” To call him Minister is an inexcusable offense. An

introduction to a Catholic archbishop, “Your Grace, may I present-----------?” A senator is always introduced as “Senator Davis,” whether he is still in office or not. But the

President of the United States, once he is out of office, is merely “Mr.” and not “Ex-President.”

1945 edition, chunk 228, Located on pages 33-34

*Note: Explicit reference. Irish accent is conveyed as unproblematic.*

pretentiousness, however, is an annihilating blight that destroys root and vine.

PRONUNCIATION

Traits of pronunciation which are typical of whole sections of the country, or accents inherited from European parents, must not be confused with crude pronunciations that have their origin in illiteracy. A gentleman of Irish blood may have a brogue as rich as plum cake, or another's accent be soft Southern, or flat New England, or rolling Western; and to each of these the utterance of the others may sound too flat, too soft, too harsh, too refined, or drawled, or clipped short, but not uncultivated.

1956 edition, chunk 4169, Located on page 607

*Note: Explicit reference. Cultural similarity and difference compared with American-ness.*

Sensitive perception of the feelings of others is something that few of us possess instinctively. It is necessary, therefore, that we train ourselves to perceive the point of

view of the people of each country we visit, remembering that it is always the stranger who must adapt himself, just as the visitor does to the ways of the house in which he is a guest.

The English, the Scots, the Welsh, and the Irish are like us in so many ways we can easily be thrown out of balance by a few differences. The best type of Britisher, for example, has just as keen an appreciation of his own ability as we have

**Italian**

1934 Edition, Chunk 1804, Located on page 338

*Note: Implicit reference. “Neapolitans” (note use of quotations) is a pejorative for Italians.*

“man’s dinner,” the details depending upon the extravagance or the frugality of the host, and upon whether his particular friends are staid citizens of sober years or mere boys full of the exuberance of youth. Usually there is music of some sort, or “Neapolitans” or “darkies” who sing; or there are two or three instrumental pieces, and the dinner party itself does the singing. Often the dinner is short and all go to the theater.

GIFTS PRESENTED TO USHERS

The groom’s presents to his ushers are always put at their places at the bachelor dinner. Cuff links are the most popular gift; scarf-pins in localities where they are still fashionable.

1975 Edition, Chunk 254, Located on page 33

*Note: Explicit reference to Italian people and Italian food. A nod to the cultural impact of immigrants in the United States.*

There are many foreign words that have become an accepted part of our language, and we should be familiar with their meaning and pronunciation. We cross paths so often with the many foreigners living in the United States that it is easy to increase one's knowledge of their languages. We have Italian grocers, German bakers, French and Spanish waiters—to name only a few—all of whom are delighted if you show an interest in their language, how it sounds, and what it means. In fact, if you ask the waiter how to pronounce “parmigiana",

he very well may tell you not only how to say it, but how to cook it. Everyone loves to be an expert, and foreigners as well as Americans like to have you express interest in their language and their customs.

2011 Edition, Chunk 4968, Located on pages 671-672

*Note: Explicit reference to Italian particularities in Catholicism.*

ROMAN CATHOLIC. Traditional Roman Catholic weddings may include a nuptial Mass with communion; this precedes the final prayer and nuptial blessing. Generally, non-Catholics do not take communion. In Italian tradition, the priest may greet the couple and their families at the church door.

ORTHODOX (Eastern Orthodox), The service consists of two parts: the betrothal and

blessing of the rings followed by the ceremony at the altar. During the latter, crowns or garlands joined by a white ribbon are exchanged three times over the couple. There are no vows said during the service.

**Jewish**

1927 edition, Chunk 2188, Located on pages 382-383

*Note: Explicit reference. In-text acknowledgement conveying author’s lack of knowledge.*

REFORM AND ORTHODOX JEWISH WEDDINGS\*

The Orthodox wedding ceremony differs somewhat from the Reform Jewish ceremony. In the Orthodox ceremony, the bride

\* The Author is indebted for the following material to Rabbi Nathan Erase of Temple Emanu-El, New York.

is veiled and is escorted under a cloth canopy supported by four poles, usually held by hand, by the father and mother.

1965 edition, Chunk 1887, Located on page 311

*Note: Unclear connection.*

On occasion, divorced parents may remain good friends and their daughter’s time may be divided equally between them. If this is true, they may both wish to announce the engagement.

Mr. Gordon Smythe of Philadelphia, and Mrs. Howard Zabriskie of 12 East 72nd Street, New York City, announce the engagement of their daughter, Miss Carla Farr Smythe...

1984 edition, chunk 2189, Located on page 388

*Note: Explicit reference. Bar Mitzvah on the list of religious traditions for youth.*

knitted, baked, constructed—in short, made—yourself!

For a Religious Confirmation, First Communion, or Bar Mitzvah

Any of the following are acceptable: prayer book, religious charm or pendant, a gift of money, jewelry, a fine book, or a pen and pencil set. Select your gift based on your closeness to the youngster.

**Mexican**

1927 edition, chunk 1071, Located on page 188

*Note: Implicit reference. Enchiladas are Mexican in origin. Described here as foreign food.*

choose well-balanced dishes; an especially rich dish balanced by a simple one. Fish timbale with a thick creamed sauce might perhaps be followed by spring

lamb, other plain roast meat or a filet mignon. A broiled fish might be followed by boned capon or another elaborate meat dish. It is equally bad to select peculiar food except as a secondary course. Some people love highly flavored Spanish or Indian dishes, but they are not appropriate for a formal dinner.

At an informal dinner an Indian curry or Spanish enchilada for one dish is delicious for those who like it, and if there is another substantial dish, such as a plain roast, which practically everyone is able to eat

1965 edition, chunk 411, Located on page 61

*Note: Explicit reference. Diplomacy and use of last name to flag ethnicity.*

Representatives of other countries who are living in the United States present no particular problems as the ways of addressing them are firmly fixed by governmental protocol. But whether their wives are addressed as Mrs., Madame. Senora, or some other title depends upon the usage of her particular country. In many instances, she uses whatever she

would in her own country (that is, the wife of the Mexican ambassador is Senora Ortega), but sometimes, especially when hers is a difficult or little known language, she uses Mrs. or Madame. In this age of international travel, we may very well find ourselves in need of information about the important personages of countries other than our own.

2004 edition, chunk 3116, Located on page 409

*Note: Explicit reference to Mexican cooking as part of an international cooking repertoire.*

don't leave the impression that your interpretation of the author's work is the correct one and somebody else's is wrong.

Cooking Groups

Serious cooks who meet regularly get the chance to try out new gourmet dishes, taste a variety of wines, and have a great time in the bargain. A certain cuisine—say, Italian. Mexican, or Thai—is often chosen, and club members, who usually number six to twelve—are assigned certain dishes to prepare.

**Muslim**

1997 edition, chunk 3562, Located on page 587

*Note: Explicit reference. Describing religious practices for people encountering them as guests.*

Mecca is faced, and worshippers bow, prostrate themselves, and recite fixed prayers. When a congregational prayer is recited at a mosque, men and women form separate lines for prayer. Guests are expected to sit separately from Muslims, since they are not permitted to join the prayer line. Guests sit on a prayer rug on the floor, as do the rest of the worshippers.

2004 edition, chunk 3645, Located on page 552

*Note: Explicit reference. Basic information for non-Muslims.*

Islamic Holy Days

Ramadan. A monthlong time for reflection and spiritual discipline, including fasting between dawn and dusk

Greeting: Ramadan Mubarak ("May God give you a blessed month")

Laylat al-Qadr. The last ten days of Ramadan, during which special prayers are

offered

‘Eid al-Fitr. The Feast of the Breaking of the Fast, celebrated at the end of

Ramadan

Greeting: Id Mubarak ("May God make it a blessed feast")

'Eid al-Adha. Commemorating Abraham’s obedience to God

Greeting: Mubarak

2017 edition, chunk 3553, Located on page 680

*Note: Implicit reference. Muslim religious leader included with in list of religious leaders.*

Special titles

it’s not uncommon for someone who is called by one title to refer to himself by another—for example, a pastor who introduces himself as Reverend Thompson. You might find that some clergy prefer to be called by their title and first name, such as Pastor John. A female Episcopal priest might ask to be called Reverend Weatherly or Mother Anne. Catholic priests are called Father and most nuns are addressed as Sister. A rabbi is addressed as Rabbi and an imam is addressed as Imam. Generally, a Buddhist priest is addressed as His Holiness or His Eminence, a male Hindu priest as Swami, and a female priest as Swamini,

**Appendix C: Embedding Sensitivity**

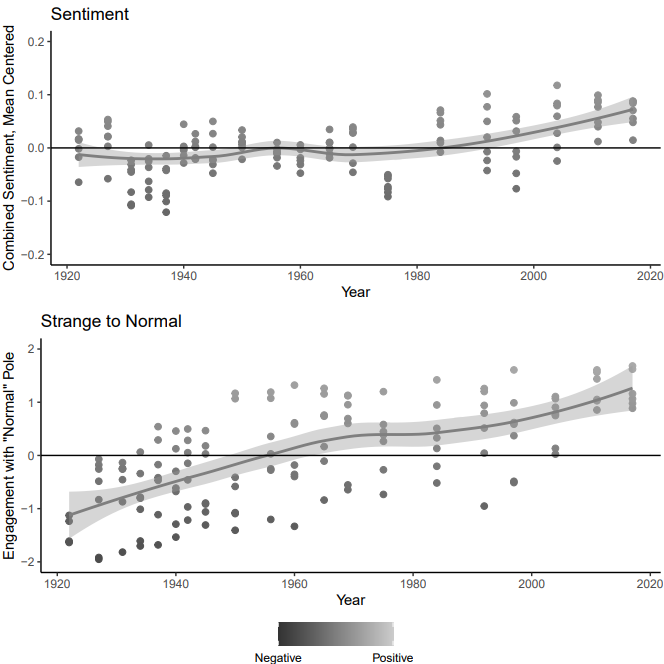
Appendix C assesses whether pseudocorpus construction is sensitive to the embedding strategy. CMDist takes embeddings as its foundational building block, using the embeddings to estimate the “distance” between documents. CMDist results are therefore only as strong as the embeddings used to create the model, and the researcher is responsible for justifying the applicability of their embeddings in their particular context (Stoltz and Taylor 2018; Stoltz and Taylor 2020). We should also make clear that CMDist tracks how varying documents engage with a *concept* that is identified using the semantic meanings in word embeddings as a map, rather than tracking the changing meaning of a *word* over time in different word embeddings (i.e., Words Movers Distance; e.g., Hamilton, Leskovec, and Jurafsky 2016).

To verify that the results are not sensitive to the embeddings used, we recreated all of the main analyses of the paper using 5 different word embeddings. These sensitivity results are not without their own limitations, as “test” embeddings should pass through the same sensitivity and confirmation analyses. For example, each new set of analyses should revisit appendixes A to select an appropriate cut off given new parameters in the tradeoff between specificity and robustness. Appendix B should also be revisited to verify the presence of groupness.

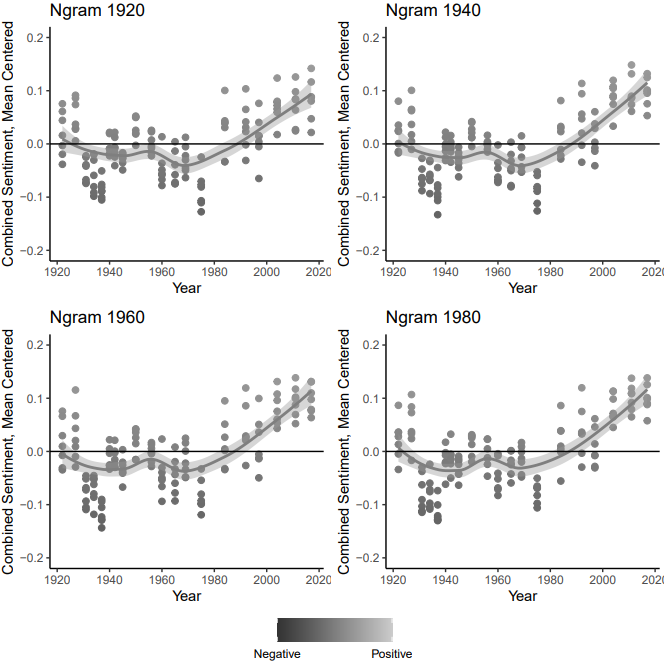
In consideration of the limitations above and the results from the sensitivity analyses below, we are confident that our current embedding strategy for pseudocorpus creation best represents both our theoretical goals - which are historical in nature - and our analytic goals. Overall, the results are similar regardless of the embedding strategy used to create the pseudocorpus. The greatest deviation occurs in the sentiment analysis, which in particular circumstances yields muted results. Yet, for most focal group concepts, trends in sentiment and meaning are highly consistent.

Figure C1 present results using FastText embeddings. 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 embeddings. FastText embeddings have been applied when examining contemporary texts, texts spanning the 20th century, and even texts published in the 17th century (Stoltz and Taylor 2019). Results yield similar overall trends to the main results, although perhaps the sentiment results more muted. As may be expected given the large size and consistent use, FastText embeddings provide tighter estimates, suggesting that a higher cutoff may be warranted should those embeddings be used in the main analysis.

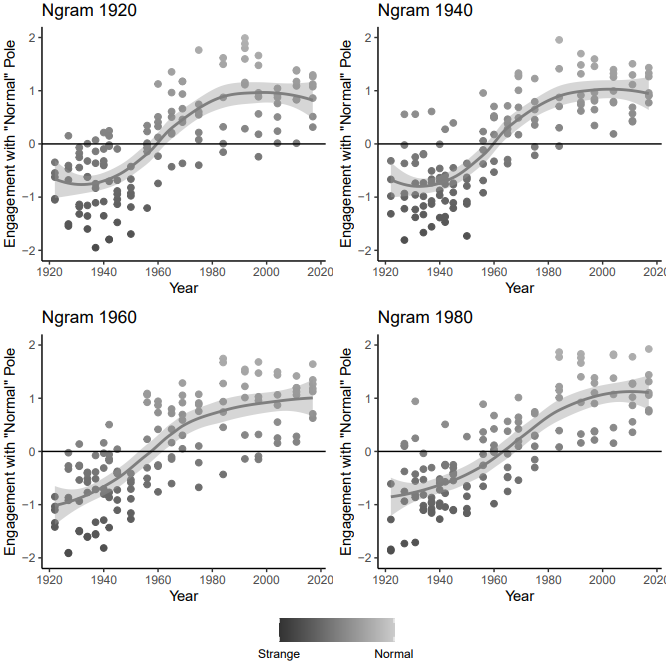
Figures C2 and C3 use embeddings from four decade-specific embeddings (1920; 1940; 1960; 1980). With similar caveats to the FastText embeddings, these results suggest consistent patterns. These analyses suggest that variations across time in semantic meanings – the meanings of words – do not affect the conceptual location identified by CMDist for the core groups or meanings of normal and strange.



*Figure C1. Results using FastText embeddings. Results show similar overall trends compared to the main results, although sentiment results are muted. FastText embeddings also provide tighter estimates, suggesting that a higher cutoff may be warranted with the use of FastText.*



*Figure C2. Sentiment analyses using four decade-specific n-gram embeddings (1920; 1940; 1960; 1980). Results are robust across these historical embeddings.*

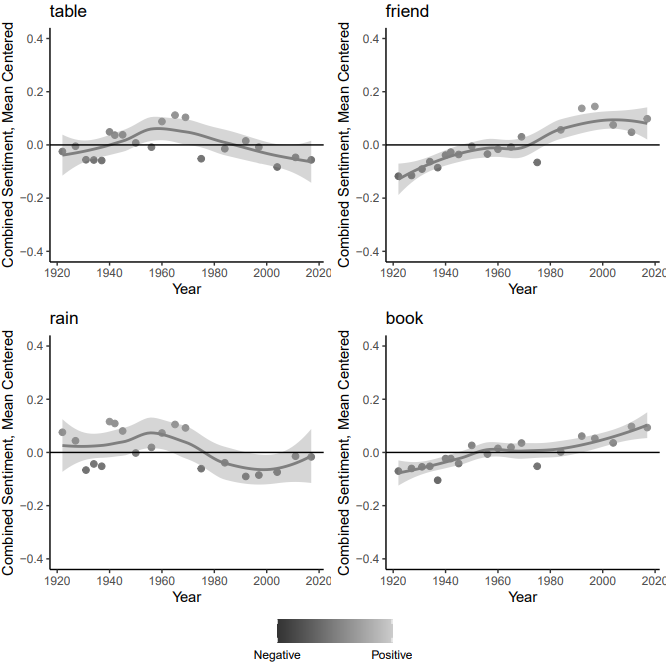


*Figure C3. Strange-to-Normal analyses using four decade-specific n-gram embeddings (1920; 1940; 1960; 1980). General trends are fairly consistent across these historical embeddings.*

Appendix D: Verifying sentiment measurement

We assess our sentiment measurement process by producing pseudocorpora and analyses using eight words that we consider “neutral” concepts, by which we mean they are analytically distinct from our focal group concepts. The goal of this analysis is to confirm that our results are not merely an artifact of the method and data that would be observed no matter what focal concepts were selected. We created pseudocorpora composed of chunks in the text most salient to *table, friend, book, rain, morality, immigrant,* and *women.* Apart from immigrant, we have few reasons to expect that the sentiment trends for these concepts would resemble those for our focal group concepts. In the case of the concept *immigrant*, this analysis also serves the purpose of showing whether that concept does have a similar sentiment profile in *Etiquette*.

Figures D1 and D2 present the mean-centered scores from these analyses. As expected, results follow different patterns. Some concepts (e.g., *friend, book*) correspond to chunks that become more positive over time while rain perhaps corresponds to less positivity. The rest of the terms follow no clear trend and display more or less variability over time. *Immigrant* does display more consistent positivity after 1980, a trend that is consistent with the sentiment analysis of several of the focal group concepts as well.



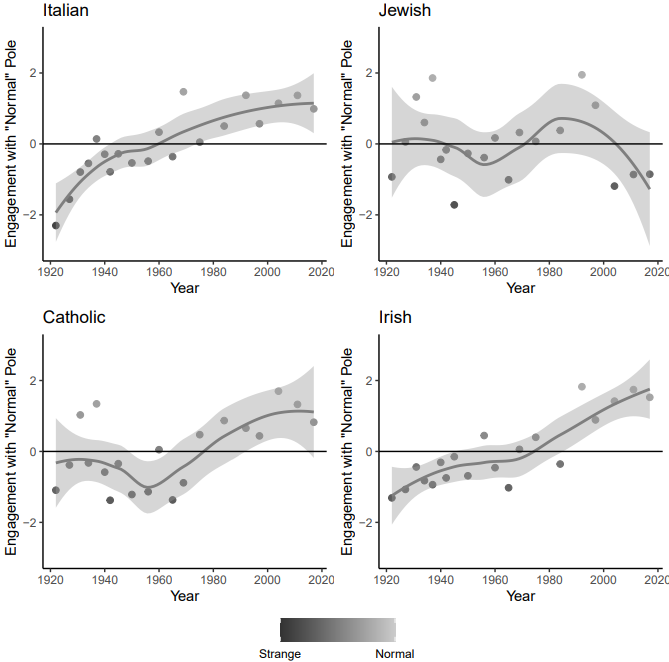
*Figure D1. Confirmatory pseudocorpora creation and sentiment analyses with “neutral” concepts. The rest of the terms follow no clear trend and display more or less variability over time, confirming that the main results are not artifacts of the method and data that would be observed no matter what focal concepts were selected.*



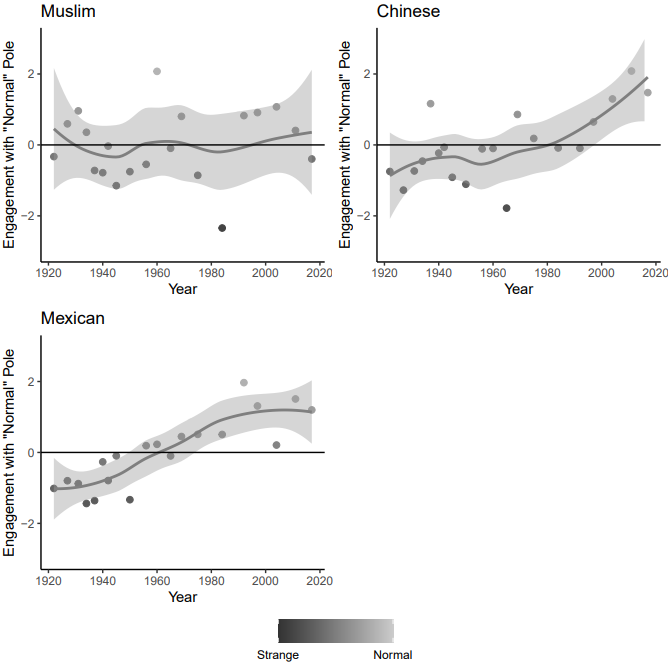
*Figure D2. Confirmatory pseudocorpora creation and sentiment analyses with “neutral” concepts. The rest of the terms follow no clear trend and display more or less variability over time, confirming that the main results are not artifacts of the method and data that would be observed no matter what focal concepts were selected. Immigrant does display more consistent positivity after 1980, a trend that is consistent with the sentiment analysis of several of the focal group concepts.*

**Appendix E. Strange-to-Normal Sensitivity**

Figures E1 and E2 present results from analyses where the strange-to-normal distinction is specified using an alternative, combined-concept specification. In these analyses, we specify two concept centroids: “immigrant, foreign, strange” vs “citizen, american [sic], normal.” Results follow a similar pattern, although there is far more variability in estimates. Importantly, this broader specification introduces especially influential outliers to a few particular pseudocorpus years. These outliers are particularly influential regarding the trend for *Jewish.* These wide variations are unsurprising considering there is less theoretical reason to expect a relationship between these alternative concepts and immigrant incorporation *in the context of the mainstream shifts we are looking for in Etiquette*. Stoltz and Taylor (2021) have used similar techniques to consider the topic of immigration in the news and press releases. For these reasons, we prefer the simple normal to strange dimension for our analyses. It is possible, and even likely, that a focal group concept becomes more or less associated with immigrant-ness over time, for example, as part of the process of immigrant incorporation or as a result of new waves of immigration. In theory, these shifts in the immigrant-ness of our focal group concepts have no determined relationship to whether or not a focal group concept becomes more or less normal. Likewise, a focal group concept could maintain a similar level of association with immigrant-ness and become more normal through mainstream shifts that normalize immigration.



*Figure E1. Results from analyses where the strange-to-normal dimension is specified using an alternative, combined-concept specification: “immigrant, foreign, strange” vs “citizen, american [sic], normal.” Results follow a similar pattern, but with more variability in estimates. This specification introduces influential outliers, in particular for Jewish.*



*Figure E2. Results from analyses where the strange-to-normal dimension is specified using an alternative, combined-concept specification: “immigrant, foreign, strange” vs “citizen, american [sic], normal.” Results follow a similar pattern, but with more variability in estimates. A focal group concept may become more or less associated with immigrant-ness over time, for example, as part of the process of immigrant incorporation or as a result of new waves of immigration. In theory, these shifts in the immigrant-ness have no determined relationship to whether or not a focal group concept becomes more or less normal. Likewise, a focal group concept could maintain a similar level of association with immigrant-ness and become more normal through mainstream shifts that normalize immigration.*