Proceedings of the Eighteenth International AAAI Conference on Web and Social Media (ICWSM 2024)

Othering and Low Status Framing of Immigrant Cuisines in US Restaurant Reviews and Large Language Models

Yiwei Luo[1], **Kristina Gligorić**[2], **Dan Jurafsky**[1,2]

[1]Department of Linguistics, Stanford University
[1]Department of Computer Science, Stanford University
{yiweil, gligoric, jurafsky}@stanford.edu

Abstract

Identifying implicit attitudes toward food can mitigate social prejudice due to food's salience as a marker of ethnic identity. Stereotypes about food are representational harms that may contribute to racialized discourse and negatively impact economic outcomes for restaurants. Understanding the presence of representational harms in online corpora in particular is important, given the increasing use of large language models (LLMs) for text generation and their tendency to reproduce attitudes in their training data. Through careful linguistic analyses, we evaluate social theories about attitudes toward immigrant cuisine in a large-scale study of framing differences in 2.1M English language Yelp reviews. Controlling for factors such as restaurant price and neighborhood racial diversity, we find that immigrant cuisines are more likely to be othered using socially constructed frames of authenticity (e.g., *authentic, traditional*), and that non-European cuisines (e.g., Indian, Mexican) in particular are described as more exotic compared to European ones (e.g., French). We also find that non-European cuisines are more likely to be described as cheap and dirty, even after controlling for price, and even among the most expensive restaurants. Finally, we show that reviews generated by LLMs reproduce similar framing tendencies, pointing to the downstream retention of these representational harms. Our results corroborate social theories of gastronomic stereotyping, revealing racialized evaluative processes and linguistic strategies through which they manifest.

Introduction

Identifying and understanding implicit attitudes toward food is important for efforts to mitigate social prejudice due to food's pervasive role as a marker of cultural and ethnic identity. Stereotypes about cuisines are a form of representational harm that contribute to racialized public discourse and may in turn perpetuate prejudice toward ethnic groups and negatively impact restaurants' economic outcomes (Luca 2016; Muchnik, Aral, and Taylor 2013). For example, in the United States, Chinese food is commonly associated with dirty kitchens and cheap takeout, while French food is associated with upscale dining (Hirose and Pih 2011; Ray 2007). Understanding the presence of representational harms in online corpora is especially important, given that these corpora are used to train large language models (LLMs) that are increasingly used to generate texts of all kinds.

Prior work on immigrant food attitudes has used coarser-grained signals such as numerical restaurant ratings (Ray 2007) or studied review language within smaller scale datasets (Johnston and Baumann 2007; Gottlieb 2015; Gualtieri 2022; Hammelman, Carr Salas, and Tornabene 2023). Meanwhile, larger-scale studies have ignored restaurant and neighborhood-related factors (e.g., price point, neighborhood racial diversity) that may be confounded with review sentiment (Boch, Jiménez, and Roesler 2021; Yu and Margolin 2022). Prior studies have moreover relied on naive word-count-based methods that do not distinguish between contextually different uses, e.g., *a **stinky** restaurant* vs. *the **stinky** tofu* (the latter being the name of a Chinese dish).

We build upon previous work through a study of framing in 2.1M English language Yelp reviews, leveraging NLP parsing techniques to conduct careful linguistic analysis of how customers evaluate restaurants while controlling for restaurant quality and neighborhood demographics (Figure 1). Further, since LLMs are increasingly used to generate reviews and texts of all kinds, we investigate the persistence of human stereotypes in LLMs through a controlled review-generation task. Our research questions are the following:

- **Q1** *Framing of immigrant vs. non-immigrant food:* How are immigrant restaurants (i.e. those identified with a cuisine based outside the US)[1] framed compared to non-immigrant restaurants?
- **Q2** *Framing within immigrant foods:* Are cuisines of more assimilated immigrant groups framed differently from those of less assimilated groups?
- **Q3** *Framing in synthetic reviews:* Do LLMs transmit the same framing disparities as Yelp reviewers?

[1]We recognize that questions of immigrant identity are deeply nuanced and individual; we thus rely solely on restaurants' self-reported cuisine categories and their associated geographic regions to designate restaurants as immigrant or not. We similarly acknowledge that "US" cuisine is itself the result of multiple immigrant influences, though these immigrant origins may not be as salient for contemporary restaurants with US-related labels (e.g., *New American, Southern*) compared to cuisines of regions with high levels of ongoing immigration. Nevertheless, we encourage future work to adopt a more nuanced treatment of individual cuisines.

(**Chinese**) Their menu includes the usual American Chinese flairs like sweet and sour (fill in the blank), and (fill in a meat) with black/white or sweet and sour sauce. But they also have a lot of actual authentic Chinese dishes [...] dim sum is the usual stuff, I'm not sure if they push it around in a cart at lunch since I only go there for dinner. The service is pretty good, but because I been going there for a while the owner is even nicer to my party if she happens to be our server. The place is very nice and clean [...]

(**Italian**) [...] currently my favorite local Italian-American restaurant by far. Their food is made with care and focus on presentation. The prices are more than reasonable for the breathtaking feasts sent to your table with each entree. Wonderful classic dishes and surely the most impressive chicken-parm you will find in the state.

Figure 1: Example reviews showing frames detected in our analysis (blue: luxury; green: prototypicality; pink: hygiene; gray: authenticity). Both customers gave 4 stars and both restaurants are designated as $ (on the 4-point scale $ to $$$$) with the same mean rating.

These questions are informed by work from the sociology of taste. Regarding **Q1**, the theory of cultural omnivorousness posits that immigrant food is climbing in social status, as part of a broader contemporary shift occurring in other forms of cultural consumption (e.g., music) from a small number of highbrow genres to a more democratic variety (Peterson 1997). Crucially, this shift is enabled by the perception and framing of immigrant food as exotic and authentic: by valorizing immigrant food as novel, as well as authentically so, consumers can legitimize their choice of a more omnivorous palate beyond traditional haute cuisine, while still maintaining a pretense of discerning taste over an indiscriminate appetite (Johnston and Baumann 2007). At the same time, the frames of exoticism and authenticity represent harmful forms of "othering" (Said 1978) that reinforce a mode of outsider cultural consumption that objectifies food in reductive and essentializing ways.

Regarding **Q2**, the theory of the ethnic succession of taste argues there are status differences in cuisines due to migration patterns and resulting socioeconomic gaps. E.g., as Italian immigration to the US slowed in the 20th century and Italian Americans moved up the socioeconomic ladder, so too did the status of Italian food; conversely, Asian and Hispanic food remain lower status, since ongoing Asian and Hispanic immigration continues to populate low-wage jobs, especially in the restaurant industry (Ray 2007, 2017). These status differences are reflected similarly to class distinctions in other domains of cultural consumption: whereas highbrow genres tend to receive aesthetic and emotional judgments, lowbrow genres are evaluated on functional and material criteria (Bourdieu 1987; Beagan, Power, and Chapman 2015; Domański et al. 2017). In a culinary context, the latter translates to a concern for hygiene and cost (Zukin, Lindeman, and Hurson 2017; Yu and Margolin 2022; Williamson et al. 2009; Beagan, Power, and Chapman 2015; Hammelman, Carr Salas, and Tornabene 2023). Together, these theories paint a portrait of US food review discourse that others immigrant cuisines and devalues those of less assimilated ethnicities. Finally, we expect LLMs trained on online discourse data to show similar tendencies, in line with work on representational harms, e.g., Crawford (2017); Cheng, Durmus, and Jurafsky (2023). Thus, our hypotheses are:

H1 Immigrant restaurants are othered (i.e. framed as more exotic **H1a**, prototypical **H1b**, and authentic **H1c**) compared to non-immigrant restaurants.

H2 Restaurants associated with more assimilated immigrant groups are framed in high status terms of luxury **H2a**; conversely, restaurants of less assimilated groups are framed in low status terms of cost **H2b** and hygiene **H2c**.

H3 LLMs reproduce the same framing differences as Yelp reviewers.

We test our hypotheses using 2.1M reviews of restaurants in 14 US states, controlling for factors such as restaurant price and star rating. We focus on the three largest sources of immigrant food in the US: European cuisine (EUR), Asian cuisine (AS), and Latin American cuisine (LAT).

We find evidence for all 3 sets of hypotheses: immigrant restaurants are overwhelmingly more likely to be described as authentic, though only AS and LAT are more likely to be described as exotic. We also find AS and LAT are framed as more authentic in zipcodes with fewer Asian and Hispanic residents, respectively, suggesting that othering may be driven by cultural outsiders, rather than the members of an ethnic group themselves.

Further, reviewers are more likely to describe AS and LAT as cheap and dirty compared to EUR. These disparities persist even for the most expensive $$$-$$$$ restaurants (Yelp categorizes restaurants into 4 tiers based on menu prices: $, $$, $$$, $$$$), supporting the idea of a culinary glass ceiling preventing the cuisines of less assimilated non-European immigrants from attaining the same status as their European counterparts (Ray 2017).

Finally, we find that GPT-3.5 shows similar framing disparities to human consumers, with immigrant cuisines subject to more othering and non-European immigrant cuisines receiving more low status framing.

Together, our results corroborate social theories about racialized evaluative processes and the systematic devaluation of non-white associated cuisines, and demonstrate downstream harms present in LLMs. Our dictionaries and methods[2] can further be applied to other domains of cultural consumption to study perceptions of immigrant cultures in our increasingly globalized society.

Data and Methods

Real Consumer Reviews. We use the Yelp open dataset,[3] containing 5.2M reviews from 2005-2022 of 64K US-based businesses (after subsetting to restaurants only). We additionally exclude chain restaurants, cafes, and fast food. We

[2]https://github.com/yiweiluo/immigrant-food-framing/

[3]www.yelp.com/dataset, retrieved January 2023, ©Yelp Inc.

use businesses' self-declared category tags to obtain cuisine labels for each business (e.g., *Italian*, *Mexican*) and perform analyses on the top 25 cuisines, excluding *Asian fusion, ethnic food, Caribbean, Middle Eastern,* and *Tex-Mex* as these are difficult to associate with a single region. We then map cuisine labels to broader regions (e.g., *Italian* $\rightarrow$ *Europe*, *Mexican* $\rightarrow$ *Latin America*). Since some restaurants belong to multiple regions (e.g., *Mexican* & *Spanish*), we subset to restaurants whose cuisines all map to a single region. The above filtering steps yield 2.1M reviews of 16K restaurants based primarily in 14 states, mapping onto the 3 primary sources of immigrant food in the US: European food (EUR), Asian food (AS), and Latin American food (LAT), in addition to non-immigrant food (i.e. associated with a US-based cuisine). Since cuisine distribution within regions is non-uniform, we replicate analyses with the most frequent cuisine per region removed and at the level of individual cuisines (see Appendix). Table 1 summarizes dataset totals and Figure 2 shows distribution over cuisines per state. We note that Yelp's dataset is highly skewed towards restaurants in Pennsylvania and Florida, and not representative of restaurant concentration in the US overall.

LLM Reviews. For synthetically generated reviews, we prompt the `gpt-3.5-turbo-0613` and `gpt-3.5-turbo-1106` models via Chat Completion API. For all prompts, we varied price point ($ – $$$$), cuisine (Table 1), and sentiment. To collect a variety of review content comparable to Yelp reviews, we also varied the focus of the review for two of our prompts (e.g., entrées, staff, ambiance). We also matched the actual distribution of sentiment ratings in Yelp reviews, which skews highly positive (45% 5-stars). Occasionally, generated reviews contain disclaimer text (e.g., "As an AI language model, I can say that this customer seems happy with their experience at a French restaurant. They specifically mention that the prices are affordable [...]"). We removed such disclosure text from synthetic reviews so as to analyze only instances of framing from the first person perspective rather than the LLM's meta-commentary (i.e. in the above example, we did not count "prices are affordable" toward cost framing). After subsampling to stratify evenly across cuisine regions, our dataset contained 58K total synthetic reviews. See Appendix for further details on our prompting parameters.

Extracting Linguistic Features. We focus on the framing of three basic restaurant attributes: food (e.g., *chicken, noodles*), staff (e.g., *waiter, server*), and the venue (e.g., *place, atmosphere*). We measure framing via the adjectives modifying tokens from each of the 3 anchor sets. By targeting the framing of specific attributes, we obtain higher precision and more interpretable results compared to aggregate measures of sentiment over the entire text of reviews. For instance, we distinguish true evaluative uses (e.g., *a* ***regular*** *Mexican place*; *the restaurant was* ***stinky***) from false positives (e.g., *I am a* ***regular***, *I had the* ***stinky*** *tofu*). We curate anchor sets

[4] Yelp collapses Cajun and Creole into a single category. Since both cuisines have immigrant origins (West African, French), we re-do analyses without this category and the main results still hold.

	Region	Cuisine
Non-imm.	US	american traditional (3.6K, 546K), american new (3.1K, 561K), cajun/creole[4] (0.5K, 161K), southern (0.5K, 141K), soul food (0.3K, 43.7K)
Immigrant	LAT	mexican (1.7K, 184K), latin american (0.4K, 42.9K), cuban (0.1K, 14.4K)
	EUR	italian (2.2K, 228K), mediterranean (0.5K, 63.4K), greek (0.3K, 32.9K), french (0.2K, 26.2K), irish (0.1K, 10.1K), spanish (60, 11.6K)
	AS	chinese (1.6K, 122K), japanese (1.1K, 146K) thai (663, 81.6K), vietnamese (527, 57.3K) indian (442, 46.1K), korean (306, 36.4K)

Table 1: Summary of Yelp cuisine categories, associated geographic regions, and (#restaurants, #reviews) in our dataset.

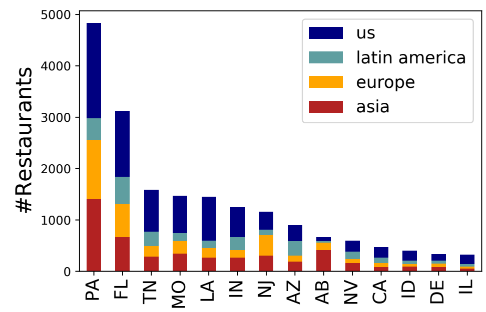

Figure 2: Distribution of restaurants over cuisine, state.

with WordNet and augment the food anchors with a dataset of menu items (Jurafsky et al. 2016). We parse all reviews using spaCy with the Coreferee add-on and leverage dependency parse relations to retrieve adjectival features, excluding those under the scope of negation (e.g., ***not*** *clean*). We then aggregate adjectival features over all anchor sets.[5]

Quantifying Framing. We quantify framing along broader dimensions of interest with hand-crafted dictionaries. We measure raw counts of linguistic features belonging to each dictionary within review texts to obtain numerical framing scores along each dimension. We next aggregate framing dimensions into broader theoretical constructs as follows: we measure othering as **exoticism**, which asserts difference and unfamiliarity; **authenticity**, which asserts faithfulness to something's ethnic origins and thus implicitly situates it outside the mainstream (Boch, Jiménez, and Roesler 2021), and **prototypicality**, a related form of the "outsider gaze" that asserts all X are alike (Rhodes, Leslie, and Tworek 2012; Golash-Boza 2016). We measure perceived high and low status as **luxury** and fine dining, **hygiene** (a marker of low perceived status (Zukin, Lindeman, and Hurson 2017; Yu and Margolin 2022)), and **cost/value** (another marker of low perceived status (Williamson et al. 2009)). We acknowledge that these dimensions do not exhaustively measure

[5] In early analyses, we examined framing along each attribute individually but found no notable differences.

status/class distinction as a broader construct, though they represent particularly salient dimensions given longstanding theoretical work showing that lowbrow genres are evaluated on functional and material criteria (Bourdieu 1987; Beagan, Power, and Chapman 2015; Domański et al. 2017).

Our lexicons for exoticism and authenticity are based on social science work (Johnston and Baumann 2007; Yu and Margolin 2022; Kovács, Carroll, and Lehman 2014) and augmented with a thesaurus. For the other framing dimensions, we use the Empath lexicon induction tool (Fast, Chen, and Bernstein 2016). We manually filter false positives from all lexicons. Example lemmas per framing dimension/construct are shown in Table 2. See Appendix for full lexicons.

Construct	Frame	Example features
Othering	Exoticism	different, distinctive, exotic, foreign, odd
	Prototypicality	archetypal, classic, stereotypical, usual
	Authenticity	authentic, handmade, legit, traditional
Status (high)	Luxury	alluring, classy, elegant, posh, refined
Status (low)	Cost	affordable, budget, cheap, overpriced
	Hygiene	clean, dirty, grimy, nasty, sanitary, stinky

Table 2: Example lemmas for each framing dimension.

Controlling for Confounds in Real World Reviews. In real world reviews, consumer sentiment is confounded with a number of factors beyond cuisine region, such as review length, a restaurant's price point and average star rating, and attributes of the restaurant's neighborhood, such as median income and racial diversity. We control for these confounds by including them as covariates in regression analyses to predict framing from cuisine region. We check for multicollinearity among features and find none (variance inflation factors all below 2.0). Figure 3 shows the distribution of restaurants used in our analyses along restaurant- and neighborhood-attributes. Restaurant price point and mean star rating data is supplied by the Yelp Academic dataset; neighborhood income and racial diversity figures were extracted from 2020 census data after linking each restaurant's geographic coordinates to individual census tracts. Yelp does not release user-specific information, so we were not able to control for user attributes (e.g., age, gender). Instead, we replicate analyses on the reviews of 89 high-volume reviewers and find similar framing disparities across cuisines, within a single user (see Appendix).

Qualitatively Measuring Framing. Separate from our regressions, which as we will see later in the Results section, reveal disparities in aggregate framing scores, we are also interested in qualitative differences in how cuisines are reviewed. We use the Fightin' Words method (Monroe, Colaresi, and Quinn 2008), which measures the strength of association between individual features with a given sample as the weighted log-odds ratio between a feature occurring in a given sample over a reference sample. In our case, we compare a feature's odds of occurring in a review of cuisine C to its odds of occurring in a review of a non-C cuisine. We also include an informative prior in the form of all review texts, and compute the z-score to measure the statistical significance of the association after controlling for variance in feature frequency. Formally, we compute the association strength δ for a word w with a cuisine C as:

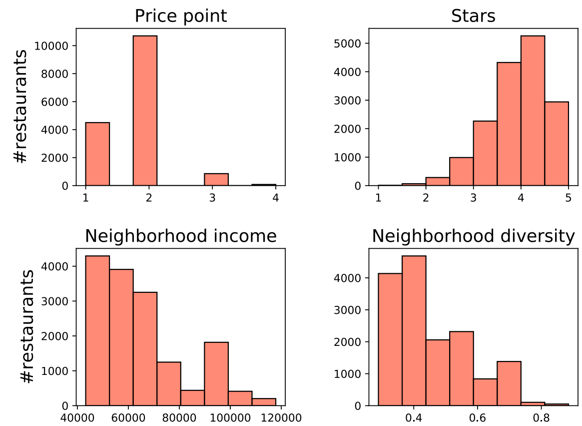

Figure 3: Distribution of restaurants along price point (corresponding to Yelp-designated \$), mean star rating, neighborhood income (2020 USD), and neighborhood racial diversity from 2020 (computed as the Simpson Diversity Index).

$$\delta_{w,C} = \frac{log\left(\frac{L_{w,C}}{L_{w,C'}}\right)}{\sqrt{\frac{1}{N_{w,C}+N_{w,P}} + \frac{1}{N_{w,C'}+N_{w,P}}}}, \qquad (1)$$

where $N_{w,C}$ is the count of w in reviews of C, $N_{w,C'}$ is the count of w in reviews of non-C, and $N_{w,P}$ is the count of w in the prior. $L_{w,C}$ and $L_{w,C'}$ are defined as follows:

$$L_{w,C} = \frac{N_{w,C}+N_{w,P}}{\sum_{x\in C} N_{x,C} - N_{w,C} + \sum_{x\in P} N_{x,P} - N_{w,P}}$$

$$L_{w,C'} = \frac{N_{w,C'}+N_{w,P}}{\sum_{x\in C'} N_{x,C'} - N_{w,C'} + \sum_{x\in P} N_{x,P} - N_{w,P}}$$

The drawback of Fightin' Words is that we cannot control for confounds like restaurant price as in regressions, but we employ the method to gain insights regarding, e.g., qualitative variation in exoticism framing across cuisines.

Results

Study 1A: Othering of Immigrant Cuisines in Yelp

To test our first set of hypotheses concerning othering of immigrant restaurants, namely that immigrant restaurants are

Frame	Most Eur	Most Lat	Most As
Exoticism	—	different (3.9)	exotic (5.0) distinct (2.7) unfamiliar (2.6)
Protoypicality	classic (6.4) regular (3.0) exemplary (2.0)	typical (4.4) usual (2.6)	usual (11.7) typical (5.9) standard (5.5) common (4.4) essential (2.2) stereotypical (2.0)
Authenticity	homemade (17.5) quaint (5.9) true (2.6) rustic (2.1)	authentic (51.1) handmade (4.6)	authentic (42.8) traditional (14.6) legit (4.6) unassuming (3.6) modest (2.1)

Table 3: Features within each othering frame most associated with each immigrant cuisine. Association strengths measured as z-scores of the weighted log odds ratio between a feature occurring in a review of a cuisine over all other cuisines (see Eq. 1).

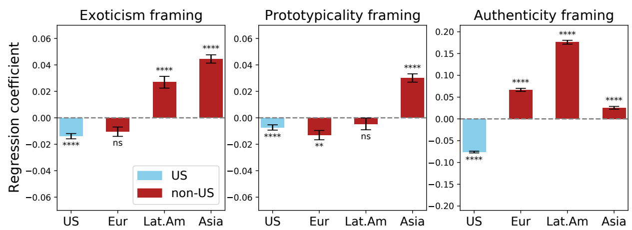

Figure 4: Othering of immigrant cuisine: Linear regression coefficients predicting othering in review text from cuisine region, showing immigrant cuisines receive more othering. Note the different y-scale for Authenticity framing. Error bars are 95% CIs. Significance values for US (the reference level) indicates whether being US has a significant effect on the outcome variable, other values indicate whether a cuisine is significantly different from US with respect to its effect on the outcome. ns=p>0.05, **=p<0.01, ***=p<0.001.

framed as more exotic **H1a**, prototypical **H1b**, and authentic **H1c** compared to non-immigrant restaurants, we fit separate linear regression models using the following equation:

$$Y = \beta_0 \cdot C + \beta_1 \cdot l + \beta_2 \cdot p + \beta_2 \cdot s + \beta_3 \cdot i + \beta_4 \cdot d + \alpha, \quad (2)$$

where Y is a continuous variable measuring the extent of a single frame (e.g., exoticism), the β_i terms are coefficients, C is categorical cuisine (with US as the reference level), l is review length, p is categorical restaurant price (with $$ as the reference level), s is restaurant mean star rating, i is neighborhood income, d is neighborhood diversity, and α is an intercept term capturing the framing score of a review at the reference level for all categorical variables (i.e. a review of a $$ US restaurant). This model predicts each form of othering from cuisine region, while controlling for review length, restaurant mean star rating, restaurant price point, neighborhood median income, and neighborhood racial diversity. Additionally, we qualitatively explore how each cuisine is othered using the Fightin' Words method (see Eq. 1; also Monroe, Colaresi, and Quinn (2008)).

We find from our regression analyses that overall, **H1a-c** are all borne out: othering along all 3 dimensions is quantitatively more prevalent in reviews of immigrant cuisine (Figure 4). However, we also observe intra-cuisine variance: othering of immigrant cuisine is driven primarily by As and Lat, with both cuisines consistently framed as more exotic, and As framed as more prototypical. In contrast, Eur is not framed as significantly more exotic or prototypical than non-immigrant cuisine (which is consistently associated with a decrease in othering). The negative coefficients for non-immigrant and Eur exoticism in particular support Janer's (2005) suggestion that US dining culture remains strongly rooted in a Western culinary perspective,[6] with non-European immigrant cuisines perceived as different with respect to this implicit reference point.

Qualitatively, we find from applying the Fightin' Words method (Eq. 1) that there are also notable differences in how Eur vs. As and Lat are othered (Table 3). Although all cuisines are described with neutral prototypicality features (e.g., *regular*, *standard*, *typical*), only Eur is described with positive features like *classic* and *exemplary*. Both the neutral and positive features represent essentializing language in that they reduce a cuisine to a prototype, but the positive features do so by placing favorable emphasis on certain representative elements of a cuisine. Our results also corroborate Gualtieri (2022), who found that non-white restaurants in the *Michelin Guide* tend to be described in terms of authenticity, and European restaurants according to what they call the "logic of technique," or terms like *exemplary* that assert the existence of a prized culinary canon.

Further, we notice that authenticity is afforded to Eur through adjectives of simplicity, such as *homemade, rustic,* and *quaint*, whereas Lat and As are afforded authenticity by being *authentic* or *legit*. In particular, Eur authenticity features evoke the current prestige fare of simple, artisanal, farm-to-table cooking championed by restaurateurs such as Alice Waters of Chez Panisse. As Ray (2007, p.100) writes: "The craftsmanship of bourgeois home-cooking was the new posture, contrasted with the mannered style of French haute cuisine [...] Rusticity replaced elegance." In other words, the framing of Eur authenticity via simplicity may reflect high cultural capital rather than authenticity per se.

Study 1B: Othering by Cultural Outsiders. Since we posit that the above framing dimensions of exoticism, prototypicality, and authenticity are forms of gastronomic othering, in which a cuisine is objectified by outsiders, we expect othering to interact with the racial background of reviewers. For instance, Boch, Jiménez, and Roesler (2021) found that reviews of Mexican restaurants mentioned authenticity less in areas with larger Mexican populations. Demographic information for Yelp users is not available in the dataset, so

[6]Or perhaps, a Global North perspective. The Global North vs. South division could also explain why Japan patterns more like Europe than Asia for certain kinds of framing, though we do not see this for South Korea (see per-cuisine results in the Appendix).

we similarly examine the influence of the % population of self-identifying Asian and Hispanic residents in the neighborhood as a proxy, using 2020 census data linked to each restaurant's postal code.[7] In cases where users self-divulge that they are not locals (e.g., "I'm from out of state"), we detect and exclude those users' reviews with a regex filter.

We fit two separate linear regression models[8] to examine the effect of local self-identifying Asian and Hispanic % population on othering of As and Lat, respectively (i.e. the effect of cultural outsiders). We find that a higher % population of cultural outsiders significantly increases authenticity framing for As and Lat (Figure 5). Interestingly, a higher % population of outsiders did not have significant effects on exoticism framing and had inconsistent effects on prototypicality framing: while areas with fewer Hispanic residents had more prototypicality framing, those with more Asians had less. However, these results may be skewed by the narrow range of geographic regions represented in our dataset, especially in the case of Asian neighborhoods: 83% of the neighborhoods with a high Asian % population are located in Pennsylvania, and 36% of high Hispanic % population neighborhoods are located in Florida (despite PA having only 4% of the overall US Asian population and FL having only 26% of the overall US Hispanic population).

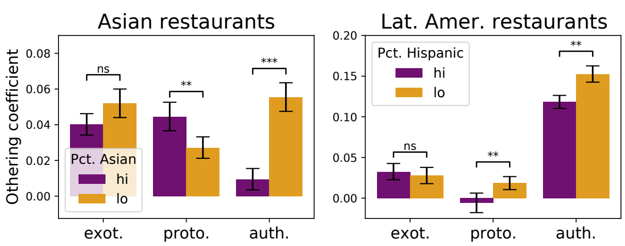

Figure 5: Othering by outsiders: Linear regression coefficients predicting othering of As/ Lat within neighborhoods with a high/low % (i.e. above/below the median) of Asian and Hispanic residents. Error bars are 95% CIs. Significance determined by a Wald test comparing coefficients within the same model. ns=p>0.05, **=p<0.01, ***=p<0.001.

[7]https://data.census.gov/

[8]Regressions are fit using the following equation:

$$Y_j = \beta_0 \cdot P_j + \beta_1 \cdot l + \beta_2 \cdot p + \beta_2 \cdot s + \beta_3 \cdot i + \beta_4 \cdot d + \alpha, \quad (3)$$

where Y_j is a continuous variable measuring the extent of a single frame within either As or Lat (e.g., exoticism of As), the β_i terms are coefficients, P_j is a categorical coding of the % self-identifying population of the associated race (`hi` if % pop. $\geq$ the median, else `lo`, with `hi` as the reference level), l is review length, p is categorical restaurant price (with $$ as the reference), s is restaurant mean star rating, i is neighborhood income, d is neighborhood diversity, and α is an intercept term capturing the framing score of a review at the reference level for all categorical variables (i.e. a review of a $$ restaurant in a neighborhood with a high % population of the associated race). Since regressions are fit within the sample of As or Lat restaurants, we omit the cuisine type variable.

Study 2: Low Status Framing of Non-European Cuisines

To test our second set of hypotheses concerning status framing, i.e. restaurants of less assimilated immigrant groups are reviewed less in high status terms of luxury **H2a** and more in low status terms of cost **H2b** and hygiene **H2c**, we fit linear regression models (Equation 2) to predict status framing from cuisine region, while controlling for the same confounds as in Study 1A (restaurant price point, stars, neighborhood income and diversity). Since Study 2 does not make explicit predictions about othering, we did not examine the effect of reviewer racial background as in Study 1B. To qualitatively explore how different cuisines are framed as high vs. low status, we used the Fightin' Words method (Eq. 1) to retrieve the features most associated with each cuisine.

We find from our regression analyses support for **H2**: Lat and As (associated with less assimilated Latin Americans and Asians) receive less luxury framing, and more cost and hygiene framing, compared to Eur (Figure 6).

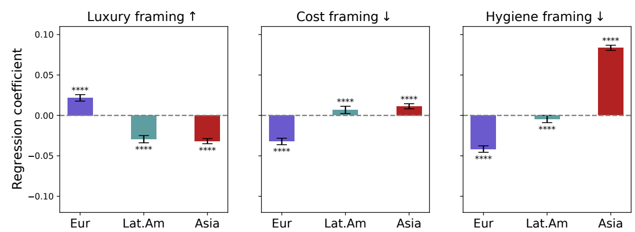

Figure 6: Europe afforded higher status: Linear regression coefficients predicting high ↑ (luxury) and low status ↓ (cost; hygiene) framing from cuisine region, controlling for review length, restaurant mean star rating, restaurant price point (reference level: $$), neighborhood median income and racial diversity. Error bars are 95% CIs. Significance value for Eur (the reference level) indicates whether being Eur has a significant effect on the outcome variable, other values indicate whether a cuisine is significantly different from Eur with respect to its effect on the outcome. ns=p>0.05, **=p<0.01, ***=p<0.001.

Qualitatively, we find from the Fightin' Words analysis that the cost features associated with Eur also tend to connote luxury (*expensive, pricey*) compared to those associated with Lat and As (e.g., *cheap, affordable*; Table 4).

Disaggregating hygiene framing into clean and dirty, we find from regression analyses that As continues to receive the most of both framing types (Figure 7). Interestingly, Eur is the only cuisine to receive more dirty than clean framing, suggesting that its cleanliness may be taken for granted, and hygiene conditions are only noteworthy when dirty. Evaluating a restaurant in terms of hygiene, regardless of whether the framing is positive (clean) or negative (dirty), presupposes that cleanliness is a relevant and important dimension of discussion, reflecting and potentially reinforcing negative assumptions about sanitary conditions of As and Lat.

Glass Ceiling Effect. Since status is confounded with restaurant price point, we additionally study framing within

Frame	Most EUR	Most LAT	Most AS
Luxury	delicate (3.2) elegant (3.0) exquisite (2.8)	outstanding (4.2)	delicate (8.7) sleek (5.9) pleasing (4.6) tasteful (4.5) ornate (3.0) posh (2.1) stylish (2.0)
Cost	expensive (3.7) pricey (2.5)	cheap (5.7) inexpensive (4.0) affordable (2.2)	cheap (13.2) affordable (8.5) inexpensive (6.2)
Hygiene	—	clean (5.4)	clean (34.1) stinky (4.9) unhygienic (2.5)

Table 4: Lemmas most associated with high, low status frames of each immigrant cuisine. Association strength measured as the z-score of the weighted log odds ratio between a feature occurring in a review of a region over all other regions (see Eq. 1).

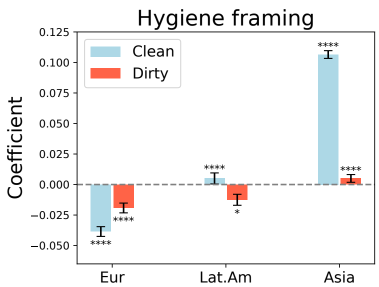

Figure 7: AS and LAT evaluated more on hygiene: Linear regression coefficients predicting clean, dirty framing from cuisine region, controlling for review length, restaurant mean star rating, restaurant price point (reference level: $$), neighborhood median income and racial diversity. Significance value for EUR (the reference level) indicates whether being EUR has a significant effect on the outcome variable, other values indicate whether a cuisine is significantly different from EUR with respect to its effect on the outcome. ns=p>0.05, **=p<0.01, ***=p<0.001.

reviews of $$$ and $$$$ restaurants (N=166K). We group $$$ with $$$$ restaurants due to the small number of the former (N=326). Although we control for price point in all analyses, we wished to more carefully examine the extent of framing disparities: even within the most upscale tier, do non-EUR cuisines continue to receive more low status and less high status framing compared to EUR?

We find that framing disparities persist within reviews of high price point restaurants, pointing to a glass ceiling effect by which AS and LAT are viewed as lower status than their EUR counterparts (Figure 8).[9] This effect is most pronounced for AS, which continues to be evaluated as less luxurious, and more in terms of cost and hygiene, and slightly less pronounced for LAT, which continues to be evaluated as less luxurious and more in terms of cost.

Study 3: Reporting Bias in LLMs

LLMs are being used increasingly for content creation in all sorts of domains—marketing, academic research, legal contexts, creative writing—and by users who may be unaware of the social biases and stereotypes they contain. As a result, the damage from these representational harms can quickly compound as they are reproduced automatically on large scales. Despite debiasing efforts, simple prompts suggest that stereotypes in human reviews may propagate to LLMs. For instance, when asked to compare French and Korean culture, one of the "key distinctions" that ChatGPT identifies is "Cuisine: French cuisine is renowned for its sophistication, emphasis on quality ingredients, and intricate preparation methods. Korean cuisine is characterized by bold flavors, a variety of side dishes (banchan), and a reliance on staples like rice and kimchi" (Figure 9).

[9]We also examined framing within James Beard award-winning restaurants but found no significant effects.

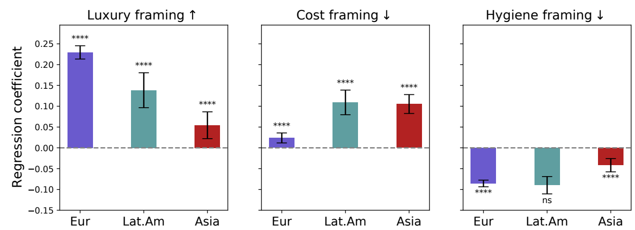

Figure 8: Glass ceiling: Linear regression coefficients predicting high ↑, low ↓ status framing of $$$-$$$$ restaurants from cuisine region, controlling for review length, restaurant mean rating, neighborhood income, racial diversity. Error bars are 95% CIs. ns=p>0.05, **=p<0.01, ***=p<0.001.

French and Korean cultures exhibit significant differences shaped by their unique histories, societal norms, and traditions. Here are some key distinctions: 1. Cultural Orientation: French culture tends to be more individualistic [...] 5. Cuisine: French cuisine is renowned for its sophistication, emphasis on quality ingredients, and intricate preparation methods. Korean cuisine is characterized by bold flavors, a variety of side dishes (banchan), and a reliance on staples like rice and kimchi [...]

Figure 9: ChatGPT's response to the prompt: "What are some differences between French and Korean culture?" Retrieved December 26, 2023.

Here we test our final hypothesis **H3**, whether framing disparities in online text corpora indeed propagate to the LLMs trained on them. To do this, we conduct a controlled review-generation task in which we prompt GPT-3.5 to write reviews evaluating hypothetical restaurants that differ in price point and cuisine using a variety of prompts and models (Appendix Table 8). We then quantitatively test **H3** using regression models[10] to predict othering and status framing from cuisine region. We also qualitatively study LLM stereotypes using the Fightin' Words method (Eq. 1).

Regardless of the specific prompt and model used,[11] we find from our regression analyses (N=58K) that GPT-3.5 exhibits many of the same tendencies as Yelp consumers, such as othering AS and LAT through frames of authenticity and exoticism (Table 5). However, we also see that immigrant restaurants are framed as less prototypical compared to non-immigrant restaurants. Nevertheless, closer inspection reveals that the main overlap between synthetic reviews and our prototypicality lexicon is due to the word *classic* (rather than words like *stereotypical* or *standard*, as we found in human reviews), suggesting that our measurement of prototypicality captures a construct more akin to "iconicity" in the context of synthetic reviews. Moreover, similar to human reviews, we found that LLMs frame EUR with higher status compared to AS and LAT (Table 6).

Finally, qualitative results of the features most strongly associated with each region (Table 7) exemplify additional stereotypes, such as the stereotype of tropicalism for LAT (e.g., *vibrant, lively, festive*) (Martynuska 2016), suggesting that gastronomic and cultural stereotypes in LLMs are not limited to the ones we study in the present work.

	EUR	LAT	AS	Immigrant
Exot.	-0.01	**0.09****	**0.07***	0.05
Proto.	**-0.09****	**-0.14*****	**-0.23*****	**-0.15*****
Auth.	**0.32*****	**0.33*****	**0.30*****	**0.32*****

Table 5: Linear regression coefficients predicting othering in GPT-3.5 reviews from cuisine region, controlling for review sentiment (reference level: neutral), showing that, as with Yelp reviews, immigrant restaurants are othered with frames of authenticity, and AS and LAT with frames of exoticism. Unlike Yelp reviews, immigrant restaurants are framed as less prototypical. Each row represents a separate regression. *=p<0.05, **=p<0.01, ***=p<0.001.

[10]Regressions are fit using the following equation:

$$Y = \beta_0 \cdot C + \beta_1 \cdot s + \alpha, \quad (4)$$

where Y is a continuous variable measuring the extent of a single frame, the β_i terms are coefficients, C is categorical cuisine (with US as the reference), s is review sentiment (with neutral as the reference), and α is an intercept term capturing the framing score of a review at the reference level for all categorical variables (i.e. a neutral review of a US restaurant). We do not control for review length or price point as we balanced synthetic reviews on these parameters for all cuisine and sentiment combinations.

[11]Results presented in this section are based on reviews generated by `gpt-3.5-turbo-0613` using a structured role-playing prompt. We obtained similar results using other prompts and model combinations (see Appendix).

	EUR	LAT	AS
Luxury	0.02	**-0.13*****	0.06
Hygiene (aggregate)	**-0.18*****	-0.14	**0.04*****
Hygiene (clean)	**-0.15*****	-0.09	**0.09*****
Hygiene (dirty)	**-0.10***	-0.11	**-0.03***
Cost (aggregate)	**-0.10***	-0.11	-0.11
Cost (expensive)	-0.06	-0.09	**0.15****
Cost (cheap)	-0.08	-0.06	**0.02****

Table 6: Linear regression coefficients predicting high and low status framing in GPT-3.5-generated reviews from cuisine region, showing LAT is framed as less luxurious and AS more in terms of cost and hygiene. Each row represents a separate regression predicting framing score from cuisine region, with EUR as the intercept and an additional factor of sentiment (reference level: neutral). *= p<0.05, ** = p<0.01, *** = p<0.001.

Discussion

In this work, we examined the framing of immigrant and non-immigrant cuisines within 2.1M Yelp reviews of restaurants located in 14 US states. Despite the reputation of the US as a culinary melting pot, we find that immigrant restaurants are systematically described as more authentic, and Asian and Latin American restaurants in particular are described as more exotic. Further, we find that restaurants associated with less assimilated Asian and Latin American immigrants are more likely to be reviewed as cheap and dirty compared to European restaurants. Our results are robust to differences in restaurant price, average star rating, and neighborhood income and racial diversity. We further show that LLMs replicate similar framing tendencies.

Beyond creating representational harms, such as reinforcing negative perceptions of immigrant cultures, framing differences in online reviews may contribute to economic inequality, since reviews can influence the decisions of millions of consumers (Luca 2016). Future work could seek to better understand and quantify the link between differences in review language and restaurant revenue.

There are a number of limitations in our current studies that future work could address. We found an inconsistent effect of reviewer race on othering, potentially due to the coarse estimate of race (based on restaurant zipcode) we used. Future research can explore datasets in which racial background at the level of individual users is available. We also made a number of simplifying assumptions in locating individual cuisines within broader regions, and neglected entire regions due to sparsity. Future work could explore within-region and within-cuisine variance more carefully, as well as study a wider range of cuisines. In addition, future work could study a more representative sample of US cities, as well as review language outside the US.

Finally, our work points to a need for strategies to combat representational harms in online food discourse. Controlled psychological experiments could reveal the deeper mechanisms underlying observed framing differences, and AI-mediated tools could intervene during the review-writing process. For instance, web plug-ins could alert reviewers to

Region	Dominant features
US	southern (18.5) fried (9.8) soulful (5.1) rustic (4.9) crispy (4.4) cheesy (4.4) slow (2.9) classic (2.5) gooey (2.5) dry (2.5) mediocre (2.4) terrible (2.2) local (2.2) creamy (2.1) overpriced (2.1)
EUR	romantic (5.8) charming (4.9) traditional (4.8) hearty (4.1) mashed (3.5) creamy (3.2) perfect (3.1) cozy (2.9) warm (2.6) red (2.4) stuffy (2.1) soft (2.1) exquisite (2.0) stuffed (2.0) rich (2.0) thin (2.0)
LAT	vibrant (11.2) lively (11.2) colorful (6.7) authentic (4.7) black (3.6) energetic (3.4) fun (3.4) shredded (3.2) festive (3.1) moist (2.6) upbeat (2.4) homemade (2.3) seasoned (2.2) mixed (2.1) tender (2.1) generous (2.0) friendly (2.0)
AS	wide (6.0) fresh (5.0) aromatic (4.9) clean (4.1) modern (4.1) sticky (3.9) iced (3.7) fragrant (3.5) spicy (3.5) beautiful (3.2) authentic (3.2) elegant (3.2) steamed (3.2) serene (3.1) hot (2.9) soothing (2.6) helpful (2.6) peaceful (2.5) balanced (2.4) comfortable (2.3) marinated (2.3) pickled (2.2) light (2.1) quick (2.1) traditional (2.1) stunning (2.1)

Table 7: Features most associated with each cuisine within GPT-3.5-generated reviews. Association strength measured as the z-score of the weighted log odds ratio between a feature occurring in a review of a given region over all other regions.

harmful linguistic choices that they may not have been aware of and provide suggestions for rephrasing.

Broader Perspective

All review data provided by the Yelp Academic Dataset is posted publicly and anonymized. We perform all our analyses on aggregate and avoid targeting any specific users or restaurants. We rely on restaurants' self-declared cuisine tags rather than inferring cuisine labels from other attributes, which may encourage biased profiling practices. We also use purely geographic criteria to locate individual cuisines within broader regions, so as to limit similar profiling biases.

Due to data sparsity, we regrettably omitted cuisines from Africa, the Middle East, and Oceania, in addition to many individual countries and culinary traditions. We also excluded intersectional cuisines with connections to multiple regions, such as Caribbean cuisine and Tex-Mex. In next steps, we aim to re-include omitted cuisines to resist tendencies of marginalizing minorities. We also recognize that the Yelp user base skews toward college-educated people with high incomes.[12] Our findings should therefore not be taken as representative of US consumers as a whole.

Additionally, our experiments with LLMs may engender risks associated with adversarial use (e.g., abuse of our prompting procedure to "review bomb" businesses with negative reviews). Finally, by studying stereotypes and frames of cultural prejudice, we risk further reifying these constructs. At the same time, we hope this work stimulates reflection, conversation, and additional research to interrogate and dismantle the preconceived attitudes that each of us as a consumer and restaurant-goer may hold.

Further Related Work

Our work extends theoretical and empirical work on the sociology of taste (Bourdieu 1987; Peterson 1997; Johnston and Baumann 2007; Williamson et al. 2009), especially as it relates to racialized criticism and the devaluation of marginalized producers (Grazian 2005; Childress and Nault 2019; Chong 2011; Janer 2005; Liu 2009; Ray 2007, 2017; Gualtieri 2022; Hammelman, Carr Salas, and Tornabene 2023). We also draw from the broader literatures on othering, authenticity, and stereotyping (Said 1978; Martynuska 2016) and work examining these attitudes in Yelp reviews (Hirose and Pih 2011; Kovács, Carroll, and Lehman 2014; Gottlieb 2015; Boch, Jiménez, and Roesler 2021) and food contexts more broadly (Freedman and Jurafsky 2011; Lee et al. 2012; Oleschuk 2017). We are also inspired by work on representational harms in discourses of race and ethnicity (Card et al. 2022) and in machines (Crawford 2017; Barocas, Hardt, and Narayanan 2017; Blodgett et al. 2020; Cheng, Durmus, and Jurafsky 2023).

Analyzing broader social aspects of food through online text has been an active area of research. Previous work leveraged various data sources, e.g., menus (Jurafsky et al. 2016; Turnwald et al. 2020), search engine logs (West, White, and Horvitz 2013; Gligorić et al. 2022), recipes (Wagner, Singer, and Strohmaier 2014; Trattner, Moesslang, and Elsweiler 2018), and review platforms (Jurafsky et al. 2014; Chorley et al. 2016). ICWSM research, in particular, has focused on studying diets through social media, such as Instagram (Garimella, Alfayad, and Weber 2016; Ofli et al. 2017) and Twitter (Mejova, Abbar, and Haddadi 2016; Gligorić, Djordjević, and West 2022). Our work is also related to studies of the impact of reviews on restaurant performance (Luca 2016; Kim, Li, and Brymer 2016; Wang, Kim, and Kim 2021) and suggests directions for finer-grained impacts that may affect ethnic groups differentially.

[12] https://www.yelp-press.com/company/fast-facts/default.aspx

Acknowledgments

We thank the reviewers and the Jurafsky Lab for their feedback. Kristina Gligorić is supported by the Swiss National Science Foundation (Grant P500PT-211127). This work is also funded by the Hoffman–Yee Research Grants Program and the Stanford Institute for Human-Centered Artificial Intelligence.

References

Barocas, S.; Hardt, M.; and Narayanan, A. 2017. Fairness in machine learning. *NIPS tutorial*, 1: 2017.

Beagan, B. L.; Power, E. M.; and Chapman, G. E. 2015. "Eating isn't just swallowing food": Food practices in the context of social class trajectory. *Canadian Food Studies*, 2(1): 75–98.

Blodgett, S. L.; Barocas, S.; Daumé III, H.; and Wallach, H. 2020. Language (Technology) is Power: A Critical Survey of “Bias” in NLP. In *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, 5454–5476.

Boch, A.; Jiménez, T.; and Roesler, K. 2021. Mainstream flavor: Ethnic cuisine and assimilation in the United States. *Social Currents*, 8(1): 64–85.

Bourdieu, P. 1987. *Distinction: A social critique of the judgement of taste*. Harvard University Press.

Card, D.; Chang, S.; Becker, C.; Mendelsohn, J.; Voigt, R.; Boustan, L.; Abramitzky, R.; and Jurafsky, D. 2022. Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proc. of the National Academy of Sciences*, 119(31): e2120510119.

Cheng, M.; Durmus, E.; and Jurafsky, D. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. In *Proc. of the 61st Annual Meeting of the ACL*, 1504–1532. Toronto, Canada.

Childress, C.; and Nault, J.-F. 2019. Encultured biases: The role of products in pathways to inequality. *American Sociological Review*, 84(1): 115–141.

Chong, P. 2011. Reading difference: How race and ethnicity function as tools for critical appraisal. *Poetics*, 39(1): 64–84.

Chorley, M. J.; Rossi, L.; Tyson, G.; and Williams, M. J. 2016. Pub crawling at scale: Tapping untappd to explore social drinking. In *Proc. of the 10th International AAAI Conference on Web and Social Media (ICWSM)*.

Crawford, K. 2017. The Trouble with Bias. NIPS.

Domański, H.; et al. 2017. Omnivorism of eating and ’highbrow-lowbrow’ distinction: cultural stratification in Poland. *Polish Sociological Review*, 199(3): 299–314.

Fast, E.; Chen, B.; and Bernstein, M. S. 2016. Empath: Understanding topic signals in large-scale text. In *Proc. of the 2016 CHI conference on human factors in computing systems*, 4647–4657.

FORCE11. 2020. The FAIR Data principles. https://force11.org/info/the-fair-data-principles/.

Freedman, J.; and Jurafsky, D. 2011. Authenticity in America: Class distinctions in potato chip advertising. *Gastronomica: The journal of food and culture*, 11(4).

Garimella, V. R. K.; Alfayad, A.; and Weber, I. 2016. Social media image analysis for public health. In *Proc. of the 2016 Conference on Human Factors in Computing Systems (CHI)*.

Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Iii, H. D.; and Crawford, K. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12): 86–92.

Gligorić, K.; Chiolero, A.; Kıcıman, E.; White, R. W.; and West, R. 2022. Population-scale dietary interests during the COVID-19 pandemic. *Nature Communications*, 13(1).

Gligorić, K.; Djordjević, I.; and West, R. 2022. Biased Bytes: On the Validity of Estimating Food Consumption from Digital Traces. *Proc. of the ACM Conference on Computer Supported Cooperative Work and Social Computing*.

Golash-Boza, T. 2016. A critical and comprehensive sociological theory of race and racism. *Sociology of race and ethnicity*, 2(2): 129–141.

Gottlieb, D. 2015. “Dirty, Authentic… Delicious” Yelp, Mexican Restaurants, and the Appetites of Philadelphia’s New Middle Class. *Gastronomica: The Journal of Food and Culture*, 15(2): 39–48.

Grazian, D. 2005. *Blue Chicago: The search for authenticity in urban blues clubs*. University of Chicago Press.

Gualtieri, G. 2022. Discriminating palates: Evaluation and ethnoracial inequality in American fine dining. *Social Problems*, 69(4): 903–927.

Hammelman, C.; Carr Salas, C.; and Tornabene, S. 2023. Paving over taquerias to put up condos: constructing urban imaginaries of migrant foodscapes via digital food narratives. *Food, Culture & Society*, 1–20.

Hirose, A.; and Pih, K. K.-H. 2011. ‘No Asians working here’: racialized otherness and authenticity in gastronomical Orientalism. *Ethnic and Racial studies*, 34(9): 1482–1501.

Janer, Z. 2005. Cooking Eurocentrism: Towards a Critical History of French Cuisine. In *Joint meetings of the Agriculture, Food and Human Values Society and Association for the Study of Food and Society, Portland, June*.

Johnston, J.; and Baumann, S. 2007. Democracy versus distinction: A study of omnivorousness in gourmet food writing. *American Journal of Sociology*, 113(1): 165–204.

Jurafsky, D.; Chahuneau, V.; Routledge, B.; and Smith, N. 2016. Linguistic markers of status in food culture: Bourdieu’s distinction in a menu corpus. *Journal of Cultural Analytics*, 1(1).

Jurafsky, D.; Chahuneau, V.; Routledge, B. R.; and Smith, N. A. 2014. Narrative framing of consumer sentiment in online restaurant reviews. *First Monday*.

Kim, W. G.; Li, J. J.; and Brymer, R. A. 2016. The impact of social media reviews on restaurant performance: The moderating role of excellence certificate. *International Journal of Hospitality Management*, 55: 41–51.

Kovács, B.; Carroll, G. R.; and Lehman, D. W. 2014. Authenticity and consumer value ratings: Empirical tests from the restaurant domain. *Organization Sci*, 25(2): 458–478.

Lee, L. E.; Niode, O.; Simonne, A. H.; and Bruhn, C. M. 2012. Consumer perceptions on food safety in Asian and Mexican restaurants. *Food Control*, 26(2): 531–538.

Liu, H. 2009. Chop Suey as imagined authentic Chinese food: the culinary identity of Chinese restaurants in the United States. *Journal of Transnational Amer. Studies*, 1(1).

Luca, M. 2016. Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School NOM Unit Working Paper*, (12-016).

Martynuska, M. 2016. The exotic other: Representations of Latina tropicalism in US popular culture. *Journal of Language and Cultural Education*, 4(2): 73–81.

Mejova, Y.; Abbar, S.; and Haddadi, H. 2016. Fetishizing food in digital age:# foodporn around the world. In *Proc. of the 10th International AAAI Conference on Web and Social Media (ICWSM)*.

Monroe, B. L.; Colaresi, M. P.; and Quinn, K. M. 2008. Fightin’words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4): 372–403.

Muchnik, L.; Aral, S.; and Taylor, S. J. 2013. Social influence bias: A randomized experiment. *Science*, 341(6146): 647–651.

Ofli, F.; Aytar, Y.; Weber, I.; Al Hammouri, R.; and Torralba, A. 2017. Is saki# delicious?: The food perception gap on Instagram and its relation to health. In *Proc. of the 26th International Conference on World Wide Web (TheWebConf)*.

Oleschuk, M. 2017. Foodies of color: Authenticity and exoticism in omnivorous food culture. *Cultural Sociology*, 11(2): 217–233.

Peterson, R. A. 1997. The rise and fall of highbrow snobbery as a status marker. *Poetics*, 25(2-3): 75–92.

Ray, K. 2007. Ethnic succession and the new American restaurant cuisine. *The restaurants book: Ethnographies of where we eat*, 97: 113.

Ray, K. 2017. Bringing the immigrant back into the sociology of taste. *Appetite*, 119: 41–47.

Rhodes, M.; Leslie, S.-J.; and Tworek, C. M. 2012. Cultural transmission of social essentialism. *Proc. of the National Academy of Sciences*, 109(34): 13526–13531.

Said, E. 1978. *Orientalism*. Pantheon Books.

Trattner, C.; Moesslang, D.; and Elsweiler, D. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science*, 7(1).

Turnwald, B. P.; Anderson, K. G.; Jurafsky, D.; and Crum, A. J. 2020. Five-star prices, appealing healthy item descriptions? Expensive restaurants' descriptive menu language. *Health Psychology*.

Wagner, C.; Singer, P.; and Strohmaier, M. 2014. The nature and evolution of online food preferences. *EPJ Data Sci.*, 3.

Wang, Y.; Kim, J.; and Kim, J. 2021. The financial impact of online customer reviews in the restaurant industry: A moderating effect of brand equity. *International Journal of Hospitality Management*, 95: 102895.

West, R.; White, R. W.; and Horvitz, E. 2013. From Cookies to Cooks: Insights on Dietary Patterns via Analysis of Web Usage Logs. In *Proc. of the 22nd International Conference on World Wide Web (TheWebConf)*.

Williamson, D.; Tregidga, H.; Harris, C.; and Keen, C. 2009. The working engines of distinction: Discourse for main course in restaurant reviews. *Journal of Hospitality and Tourism Management*, 16(1): 55–61.

Yu, C.; and Margolin, D. 2022. Heightened scrutiny: The unequal impact of online hygiene scores on restaurant reviews. *New Media & Society*, 14614448221127674.

Zukin, S.; Lindeman, S.; and Hurson, L. 2017. The omnivore's neighborhood? Online restaurant reviews, race & gentrification. *Journal of Consumer Culture*, 17(3): 459–79.

Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes, we do not violate any privacy norms and actively seek to mitigate unfair profiling and disrespect to cultures through this work.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes, our main claims are direct summaries of our main results given in Figures 4-8 and Tables 5-6.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes, see the Data and Methods Section.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes, we acknowledge possible artifacts in our convenience sample in the Discussion and Broader Perspective Sections.
   (e) Did you describe the limitations of your work? Yes, see the Discussion and Broader Perspective Sections.
   (f) Did you discuss any potential negative societal impacts of your work? Yes, see Broader Perspective.
   (g) Did you discuss any potential misuse of your work? Yes, see Broader Perspective.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes, see Broader Perspective. We will also release all code and data for the de-anonymized version.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes, and we tested for collinearity and considered unobservable covariates. We also discuss simplifying assumptions we made in the Discussion.
   (b) Have you provided justifications for all theoretical results? Yes, we present and explain empirical evidence supporting all results.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes, we rule out the competing hypotheses of price point, restaurant quality, neighborhood income, neighborhood racial diversity, or reviewer differences accounting for framing disparities by controlling for these variables to the best of our ability (see main paper Data and Methods & Results Sections; Appendix User-Controlled Analysis Section). We also rule out the possibility that our results are driven by the majority categories in our dataset by replicating analyses with the most frequent categories removed (see Appendix).

(d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes, and we perform regressions to control for confounds that may affect our primary variables of interest.

(e) Did you address potential biases or limitations in your theoretical framework? Yes, see the Discussion and Broader Perspective Sections.

(f) Have you related your theoretical results to the existing literature in social science? Yes, our hypotheses are formulated directly from theories in the social science literature.

(g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes, see the Discussion.

3. Additionally, if you are including theoretical proofs...

(a) Did you state the full set of assumptions of all theoretical results? NA.

(b) Did you include complete proofs of all theoretical results? NA.

4. Additionally, if you ran machine learning experiments...

(a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? Answer.

(b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA.

(e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Answer.

(f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA.

5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? Yes, see Data and Methods.

(b) Did you mention the license of the assets? Yes, see Data and Methods.

(c) Did you include any new assets in the supplemental material or as a URL? NA.

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes, see Broader Perspective.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes, see Broader Perspective.

(f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see FORCE11 (2020))? NA.

(g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset (see Gebru et al. (2021))? NA.

6. Additionally, if you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots? NA.

(b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA.

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA.

(d) Did you discuss how data is stored, shared, and deidentified? NA.

Appendix

LLM Prompting Details. We share our prompt templates and model parameters for reproducibility: For the first 2 of

*1. A customer posted the following restaurant review to an online restaurant review website:<span class='head-'line' title="[**sentiment**] review about a [**price point**] [**cuisine**] restaurant, focused on the [**focus**]">*
*2. Write a [**sentiment**] review of a [**price point**] [**cuisine**] restaurant, focusing on the [**focus**]*
*3. Give an example of a [**sentiment**] review of a [**price point**] [**cuisine**] restaurant*

Table 8: LLM prompt templates

the templates in Table 8, we varied the **focus**: staff, waitstaff, employees, waiter, waitress, food, drinks, main courses, appetizers, desserts, place, spot, atmosphere, experience, ambiance; for all templates we varied: **sentiment**: very positive, positive, neutral, negative, very negative; **price point**: $ ($10 and under), $$ ($10-$25), $$$ ($25-$45), $$$$ ($50 and up); **cuisine**: see Table 1. We use the following parameters: temperature=1, max_tokens=256, top_p=1, frequency_penalty=0, presence_penalty=0.

Abridged LLM Replication Results. For robustness, we use various standard prompt types (structured role-playing; instructing to write; instructing to give an example) to prompt `gpt-3.5-turbo-0613` and the newer `gpt-3.5-turbo-1106`, which has an expanded context window of 16K tokens (compared to 4K). We find that results with different prompts and models replicate our main results from Study 3 (see Tables 9-10; we display a subset of model/prompt combinations due to space limitations).

Lexicons. ***Exoticism*** abnormal, bizarre, different, distinct, distinctive, exotic, fascinating, foreign, intriguing, odd, peculiar, strange, unfamiliar, unnatural, unsettling, unusual, weird ***Protoypicality*** archetypal, archetype, average, basic, characteristic, classic, classical, common, commonplace, definitive, emblematic, essential, everyday, exemplary, generic, habitual, mundane, norm, normal, ordinary,

	EUR	LAT	AS	Immigrant
Exot.	-0.01	**0.12****	**0.08***	0.06
Proto.	0.02	**-0.09***	**-0.21*****	**-0.09****
Auth.	**0.29*****	**0.29*****	**0.28*****	**0.29*****

Table 9: Linear regression coefficients predicting othering in reviews generated by `gpt-3.5-turbo-1106` using prompt 1, from cuisine region, controlling for review sentiment (reference level: neutral). Each row represents a separate regression. *=p<0.05, **=p<0.01, ***=p<0.001.

	EUR	LAT	AS
Luxury	0.00	**-0.21*****	0.00
Hygiene (aggregate)	**-0.14****	0.01	**0.11****
Hygiene (clean)	-0.08	0.00	**0.13****
Hygiene (dirty)	-0.12	0.02	0.02
Cost (aggregate)	-0.04	0.04	0.01
Cost (expensive)	**-0.15****	0.00	-0.05
Cost (cheap)	0.07	0.05	0.06

Table 10: Linear regression coefficients predicting status framing in reviews generated by `gpt-3.5-turbo-0613` and prompt 2, from cuisine region. Each row represents a separate regression predicting framing score from cuisine region, with EUR as the intercept and an additional factor of sentiment (reference level: neutral). *=p<0.05, **=p<0.01, ***=p<0.001.

predictable, quintessential, regular, standard, stereotypical, typical, unremarkable, usual ***Authenticity*** accurate, authentic, hand-made, handmade, home-made, homemade, homey, humble, idyllic, laid-back, laidback, legit, legitimate, modest, original, pastoral, proper, quaint, real, real deal, rural, rustic, simple, traditional, true, unassuming, uncomplicated, unfussy, unpretentious ***Luxury*** alluring, astonishing, breathtaking, classy, dazzling, delicate, dignified, elaborate, elegant, enchanting, enticing, exquisite, extraordinary, extravagant, fashionable, glamorous, glorious, graceful, grand, lavish, lush, luxurious, magnificent, majestic, marvelous, ornate, outstanding, picturesque, pleasing, polished, posh, refined, regal, remarkable, sleek, sophisticated, spectacular, stylish, tasteful, voluptuous ***Cost*** affordable, bargain, budget, cheap, costly, economical, exorbitant, expensive, inexpensive, low-cost, low-priced, low cost, low priced, overpriced, pricey, uncostly, unexpensive ***Hygiene*** clean, dirty, disgusting, filthy, grimy, gross, (un)hygienic, messy, nasty, (un)sanitary, smelly, spotless, stinking, stinky, tidy

User-Controlled Analysis. We replicate analyses on 17K reviews of 89 high-volume users contributing ≥100 reviews and ≥10 per region. We fit linear mixed effects models with random user effects and fixed cuisine effects. Results, when significant, corroborate main results (Figures 14-15).

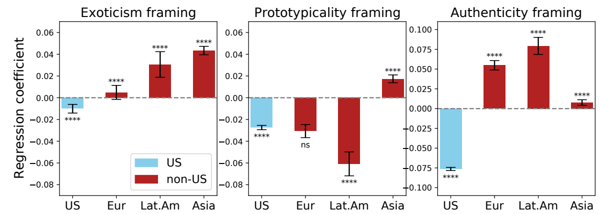

Figure 10: Coeff. from linear regressions estimating othering from cuisine, with most frequent cuisine per region removed (*US: Trad.Am.; Eur: Ital.; Lat: Mex.; As: Chinese*).

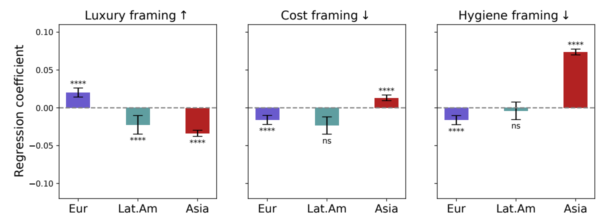

Figure 11: Coefficients from linear regressions estimating high ↑ vs. low status ↓ framing from cuisine, with most frequent cuisines per region removed.

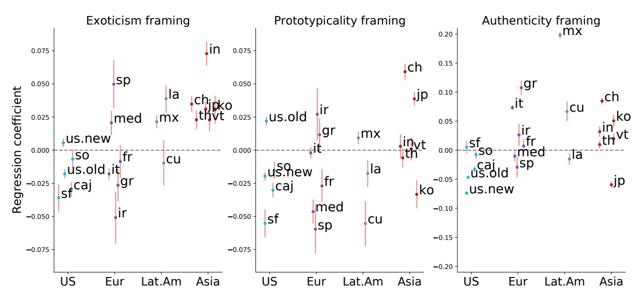

Figure 12: Per cuisine coefficients predicting othering.

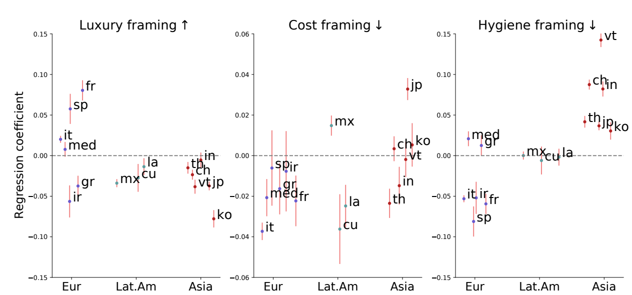

Figure 13: Per cuisine coefficients predicting high ↑ & low status ↓ framing.

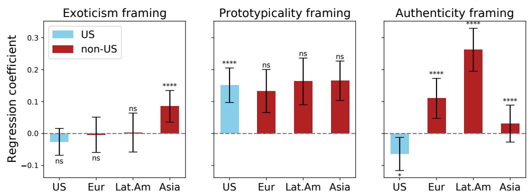

Figure 14: Coefficients from linear mixed effects models estimating othering from cuisine type, with random effects per user, showing that AS is framed as more exotic and immigrant cuisines are framed as more authentic.

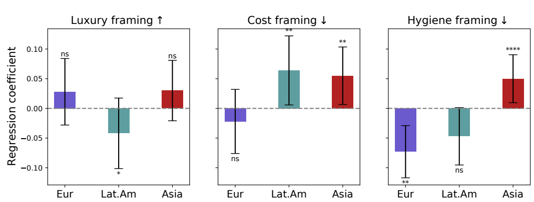

Figure 15: Coefficients from linear mixed effects models estimating high ↑ vs. low status ↓ framing from cuisine type, with random effects per user.