

Supplementary Information

The Deliciousness Efficiency Index: quantifying the hedonic–environmental decoupling across 334 dishes

F. Liu

Contents

1	Data and sample description	3
1.1	Yelp Open Dataset summary	3
1.2	Dish extraction and mention counts	3
2	Hedonic scoring methodology	4
2.1	BERT finetuning details	4
2.2	H score compression analysis	4
3	Yelp proxy bias control	6
4	Nutritional dimension (NDI)	6
4.1	Nutrient Density Index construction	6
4.2	DEI-N: nutrition-adjusted index	7
4.3	FNDDS external nutritional validation	8
4.4	NHANES/WWEIA dietary coverage	8
4.5	Nutrition-constrained substitutions	8
5	Within-category and meal-level analysis	9
5.1	Recipe-based functional categories	9
5.2	Within-meal-role variance decomposition	10
5.3	Meal-level DEI	10
6	Survivorship bias simulation	10
7	Multi-dimensional hedonic assessment	11
8	Geographic stability	11
9	Refinement curve measurement resolution	12
9.1	Minimum detectable difference	12
9.2	Cross-platform refinement validation	12
9.3	Price-tier analysis	12
10	Additional validation	15
10.1	Cross-platform hedonic validation	15
10.2	Temporal stability	15
10.3	Monte Carlo rank stability	15
10.4	Policy weight sensitivity	15
10.5	Sample expansion	15

1 Data and sample description

1.1 Yelp Open Dataset summary

The Yelp Open Dataset (2024 release) contains 6.9 million reviews across 150,000 businesses in 14 US states/provinces. After filtering for restaurant/food categories (SIC categories “Restaurants”, “Food”, “Bars”, “Cafes”, etc.), our working corpus comprises 5.3 million reviews from 70,617 restaurants. Reviews span 2005–2023, with a median length of 127 words.

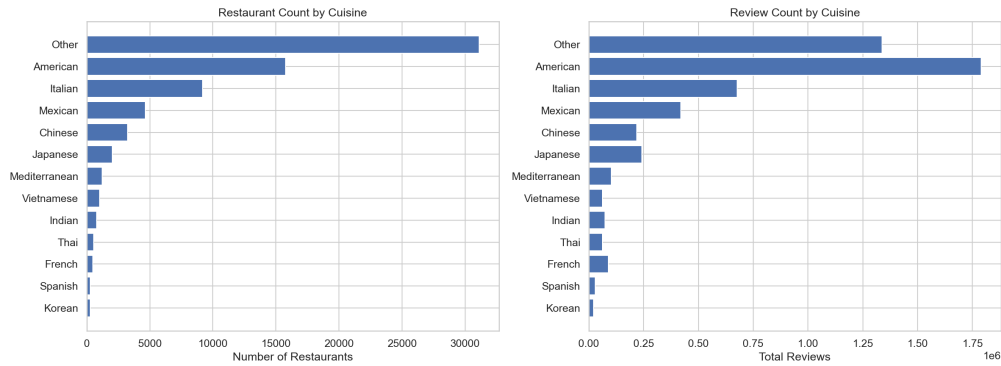


Figure 1: Supplementary Figure S1 | Cuisine distribution in the Yelp dataset. Distribution of restaurants across cuisine categories.



Figure 2: Supplementary Figure S2 | Star rating distribution. Distribution of review star ratings across all 5.3M reviews.

1.2 Dish extraction and mention counts

We extracted 158 canonical dishes using a keyword-based approach with manual curation, expanded to 334 dishes (29 cuisines) via additional recipe development for underrepresented cuisines (African, Caribbean, South American, Indonesian, Turkish, Greek, Lebanese, Filipino, Central Asian, Persian). The original 158 dishes generated 76,927 review-level mentions; the expanded 176 dishes provide recipe-level estimates without review data.

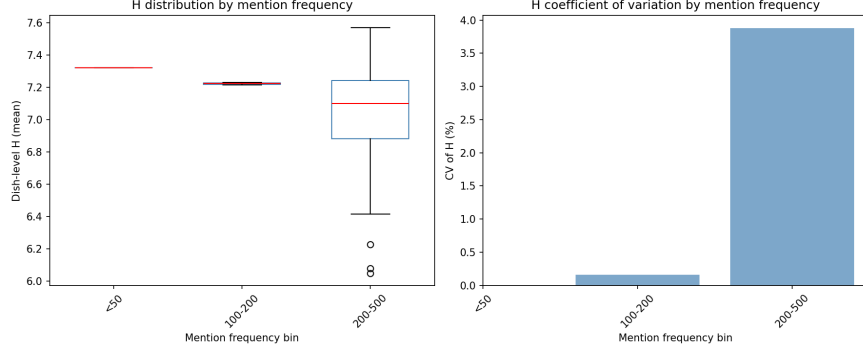


Figure 3: Supplementary Figure S3 | Mention frequency versus hedonic score precision. Dishes with more review mentions have smaller H standard errors, but all dishes exceed 50 mentions (our minimum threshold).

2 Hedonic scoring methodology

2.1 BERT finetuning details

The base model (distilbert-base-uncased) was finetuned on 1,096 valid LLM annotations (DeepSeek v3.2) using MSE loss, learning rate 2×10^{-5} , batch size 16, and 5 epochs with early stopping. The finetuned model achieves:

- Mean Absolute Error: 1.08 (on 1–10 scale)
- Dish-level Pearson r : 0.844
- H score range: [6.05, 7.57], CV = 3.9%

2.2 H score compression analysis

The observed CV of H (3.9%) is substantially smaller than the CV of E (64.3%), raising the question of whether H compression is an artefact of the measurement instrument. We investigate this through four analyses.

Compression quantification. The Yelp star-rating CV at the dish level is 12.3%, roughly $3\times$ the BERT H CV, suggesting that the BERT model compresses sentiment onto a narrower range than the raw rating data.

Beta-distribution decompression. We map the empirical CDF of H to a Beta quantile function, stretching the distribution to target CVs of 5%, 10%, 15%, 20%, 25%, 30%, and 50% while preserving rank order.

Table 1: Supplementary Table S1 | Decompression sensitivity. H contribution to $\text{Var}(\log(\text{DEI}))$ as a function of assumed H dispersion.

Target CV (%)	H contribution (%)	E contribution (%)	Rank ρ vs original
3.9 (observed)	0.3	99.7	1.000
5.0	0.4	99.6	1.000
10.0	1.8	98.2	0.991
15.0	4.3	95.7	0.973
20.0	8.5	91.5	0.950
25.0	21.8	78.2	0.922
30.0	29.6	70.4	0.890
50.0	56.2	43.8	0.782

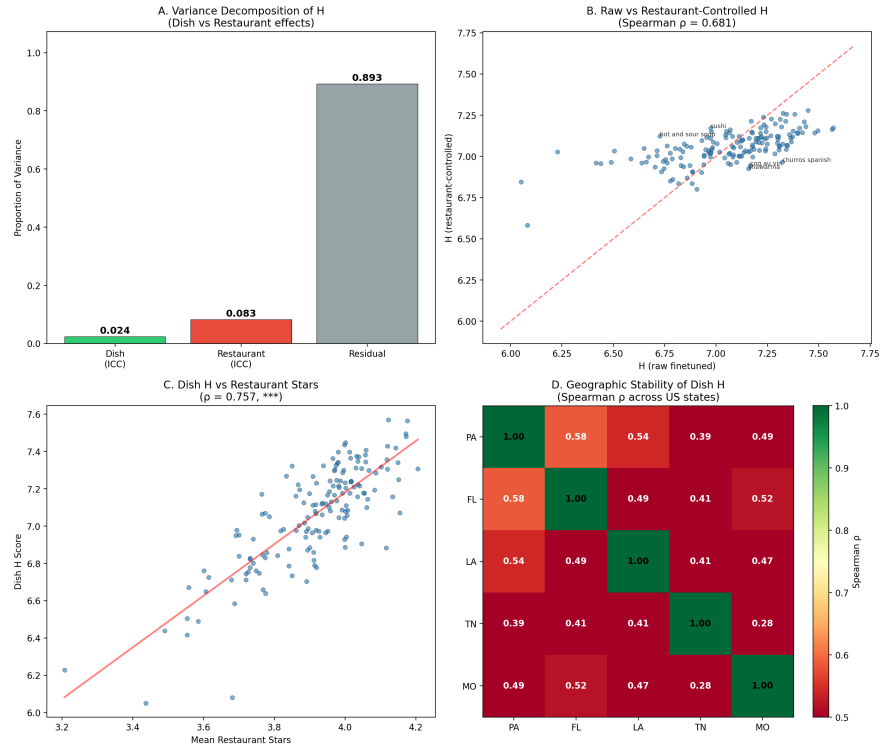


Figure 4: Supplementary Figure S4 | Hedonic score validity diagnostics. ICC decomposition, convergence analysis, and distribution of dish-level H scores.

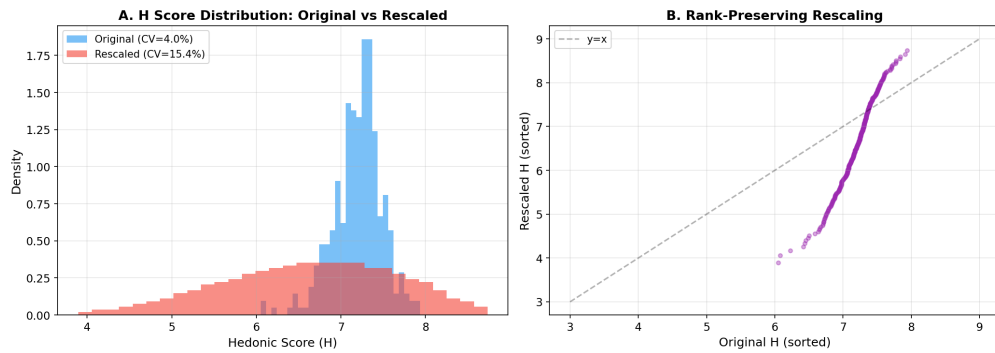


Figure 5: Supplementary Figure S5 | Rescaled H distributions at target CVs. The decompressed distributions preserve rank order while expanding the effective range.

Crossover analysis. H would need $CV > 20\%$ to contribute $> 8.5\%$ of DEI variance, and $CV > 30\%$ to surpass 25%. Since the largest reported CV for dish-level hedonic ratings in controlled settings is approximately 15–20% [1], E -dominance is robust across the plausible range.

Rank stability. Even at $CV = 20\%$, Spearman ρ with the original ranking remains 0.950, confirming that the core ranking structure is preserved.

3 Yelp proxy bias control

Restaurant-level confounders (star rating, price range, review count) could bias dish-level H estimates. We implement a two-stage residualization:

1. **Stage 1:** Regress review-level H on restaurant covariates:

$$H_{ij} = \alpha + \beta_1 \cdot \text{stars}_j + \beta_2 \cdot \text{price}_j + \beta_3 \cdot \log(\text{review_count}_j) + \varepsilon_{ij}$$

Restaurant-level $R^2 = 0.077$, confirming that confounders explain only 7.7% of H variance.

2. **Stage 2:** Extract residuals $\hat{\varepsilon}_{ij}$ as “controlled H ”, aggregate to dish level.

Table 2: Supplementary Table S2 | Proxy-bias control summary.

Metric	Value
Restaurant-level R^2	0.077
H CV (original)	3.9%
H CV (controlled)	2.5%
H % of $\text{Var}(\log(\text{DEI}))$, original	0.26%
H % of $\text{Var}(\log(\text{DEI}))$, controlled	0.10%
DEI rank ρ (original vs controlled)	0.999
Tier agreement	94.9%
Mean rank shift	1.6

The controlled H has *smaller* CV (2.5% vs 3.9%), which *strengthens* E -dominance: H ’s contribution to $\text{Var}(\log(\text{DEI}))$ drops from 0.26% to 0.10%. The DEI rank correlation between original and controlled is 0.999, confirming that proxy bias is negligible.

4 Nutritional dimension (NDI)

4.1 Nutrient Density Index construction

We compute a Nutrient Density Index (NDI) based on the NRF-7 framework [2], scoring seven encouraged nutrients per 100 kcal against Daily Reference Values (DRVs):

$$\text{NDI} = \sum_{i=1}^7 \min \left(\frac{\text{nutrient}_i / 100\text{kcal}}{\text{DRV}_i} \times 100, 100 \right)$$

The seven nutrients and their DRVs: protein (50 g), dietary fibre (25 g), iron (18 mg), zinc (15 mg), vitamin B₁₂ (2.4 µg), calcium (1,000 mg), vitamin C (60 mg).

Nutrient amounts per dish are estimated from ingredient-level USDA FoodData Central composition data, matched to the 101 ingredients in our recipe database.

4.2 DEI-N: nutrition-adjusted index

The nutrition-adjusted DEI is defined as:

$$\log(\text{DEI-N}) = \log H + \alpha \cdot \log(\text{NDI}) - \log E$$

with $\alpha = 0.5$ as the default weight (sensitivity analysis over $\alpha \in [0, 1]$).

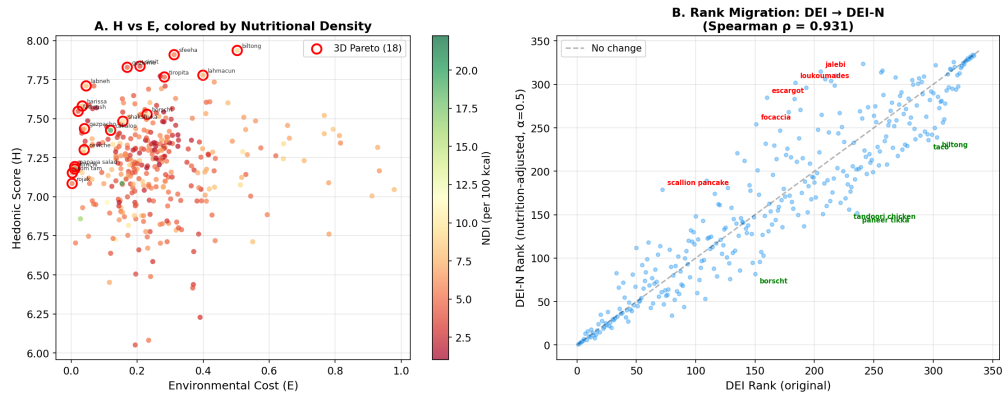


Figure 6: Supplementary Figure S6 | DEI versus DEI-N rankings. High-protein dishes (brisket, steak, ribs) gain 20–40 positions under DEI-N due to high nutrient density.

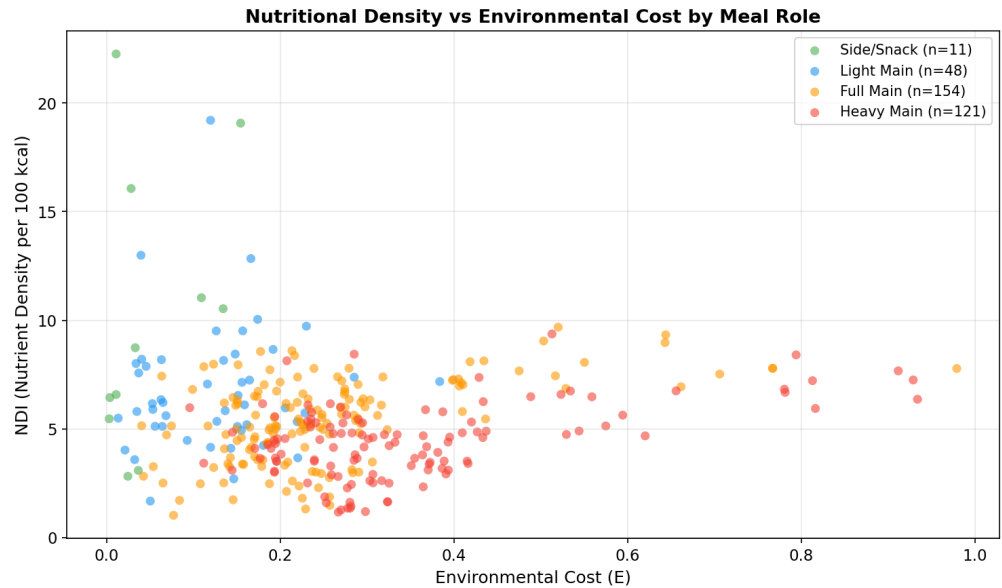


Figure 7: Supplementary Figure S7 | Nutritional profiles by recipe-based category. Mean macronutrient and micronutrient values across functional food categories.

4.3 FNDDS external nutritional validation

We cross-validated our recipe-based nutrient estimates against the USDA Food and Nutrient Database for Dietary Studies (FNDDS 2021–2023), matching 113 of our 334 dishes to FNDDS food items.

Table 3: Supplementary Table S3 | FNDDS nutritional cross-validation.

Metric	Value	p -value
Dishes matched to FNDDS	113/334 (33.8%)	—
Protein Pearson r	0.814	$< 10^{-27}$
Protein mean ratio (ours/FNDDS)	1.10	—
NDI Spearman ρ	0.707	$< 10^{-18}$
DEI-N ranking stability ρ	0.949	—
Original DEI vs FNDDS DEI-N ρ	0.928	—

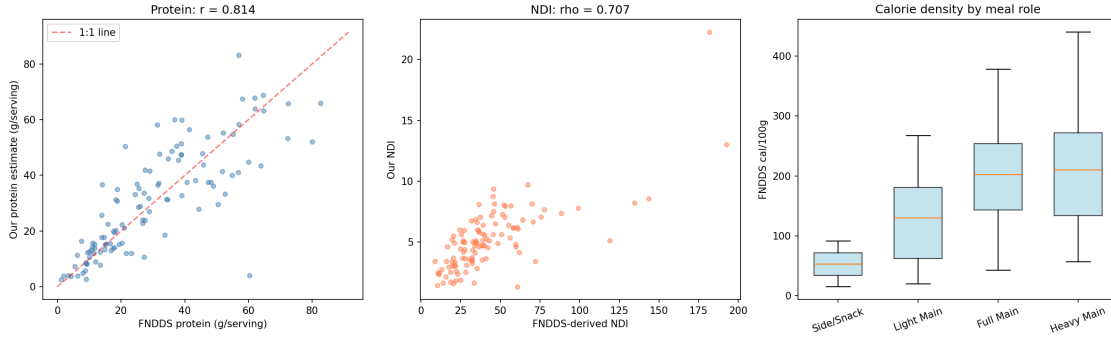


Figure 8: Supplementary Figure S8 | FNDDS nutritional cross-validation. Left: protein correlation ($r = 0.814$). Centre: NDI correlation ($\rho = 0.707$). Right: calorie density by meal role.

4.4 NHANES/WWEIA dietary coverage

Text-based matching of our 334 dishes against 5,432 FNDDS food items yields 933 matched items (17.2%) corresponding to 119 dishes (35.6%), covering 73 of 172 WWEIA categories (42.4%). Coverage is highest for prepared dishes and mixed meals (pizza 98%, burritos 100%, soups 100%) and lowest for individual ingredients (fruits, vegetables, milk, cereals). This is expected: our dataset samples *restaurant dishes*, not the comprehensive food inventory tracked by NHANES.

4.5 Nutrition-constrained substitutions

We identify substitution pairs satisfying four simultaneous constraints:

- E reduction $> 30\%$
- H loss < 1 point (on 1–10 scale)
- Protein $\geq 50\%$ of original
- Calories $\pm 50\%$ of original

Table 4: Supplementary Table S4 | Nutrition-constrained substitutions (top 10).

From	To	E reduction (%)	ΔH	Protein ratio
Onigiri	Rojak	94.9	<0.01	2.51
Guacamole	Som tam	89.7	+0.22	1.48
Brisket	Ceviche	96.0	+0.25	0.66
Osso buco	Tandoori chicken	72.1	+0.15	0.89
Steak	Chicken tikka	68.5	+0.44	0.84

5 Within-category and meal-level analysis

5.1 Recipe-based functional categories

We assign each dish to one of 13 functional categories based on primary protein source and calorie role: Red Meat Main, Poultry Main, Pork Main, Seafood Main, Dairy Main, Egg Dish, Plant Protein, Starch/Carb, Noodle/Pasta, Soup/Stew, Salad/Cold, Dessert, Mixed/Other.

Within-Category H vs E: Is H More Discriminating Within Categories?

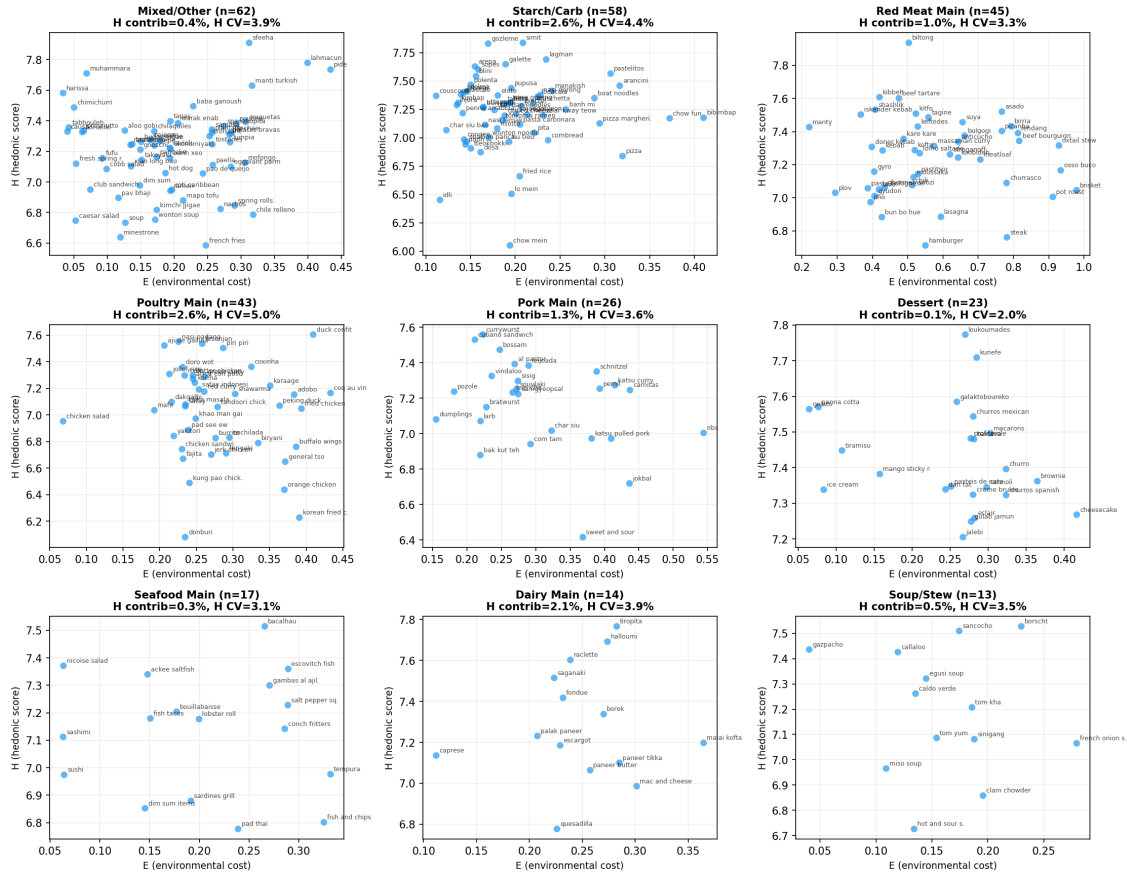


Figure 9: Supplementary Figure S9 | Within-category DEI distributions. DEI rankings within each of 13 functional categories.

Table 5: Supplementary Table S5 | Within-meal-role variance decomposition. E -dominance holds within all four functional meal roles.

Meal role	n	H CV (%)	E CV (%)	H % of Var	DEI range (\times)
Full Main	154	3.8	57.7	0.5	25
Heavy Main	121	4.2	49.0	1.0	10
Light Main	48	4.1	61.7	0.3	27
Side/Snack	11	2.7	111.0	0.04	54

5.2 Within-meal-role variance decomposition

5.3 Meal-level DEI

We compose 15,073 meal combinations (main + side) and compute calorie-weighted H_{meal} and summed E_{meal} . At the meal level, H contributes only 0.6% of $\text{Var}(\log(\text{DEI})_{\text{meal}})$, confirming that E -dominance persists. Among 6,611 calorie-equivalent swaps ($\pm 15\%$ calories), the mean E reduction is 45%.

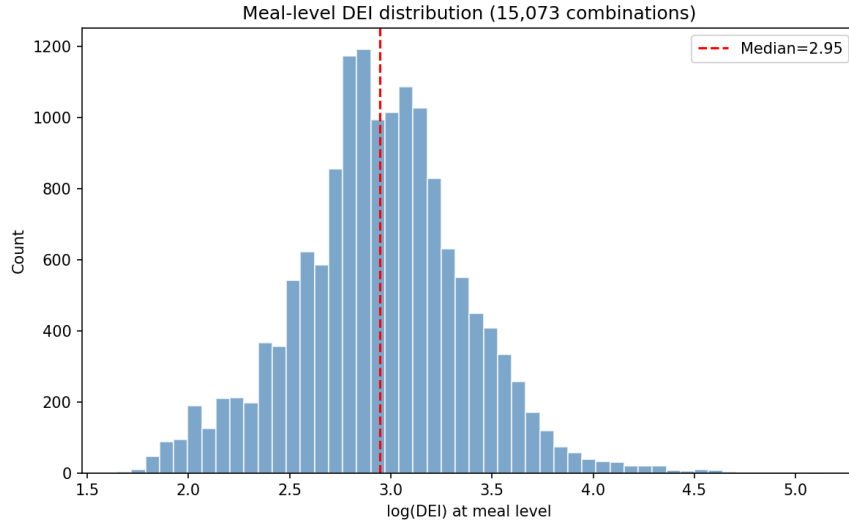


Figure 10: Supplementary Figure S10 | Meal-level DEI distribution. $\log(\text{DEI})$ across 15,073 main+side combinations.

6 Survivorship bias simulation

Our dataset only includes dishes popular enough to appear on Yelp menus. “Failed” dishes with low hedonic appeal are unobserved, potentially underestimating H variance.

We simulate this by adding K “ghost” dishes with $H_{\text{ghost}} = \bar{H} - \Delta$ and E drawn from the observed distribution. We sweep $K \in \{10, 50, 100, 167, 334\}$ and $\Delta \in \{0.5, 1.0, 1.5, 2.0, 3.0\}$.

Even in the most extreme scenario ($K = 334$, $\Delta = 2$), H contributes only 12.8% and E remains dominant at 87.2%.

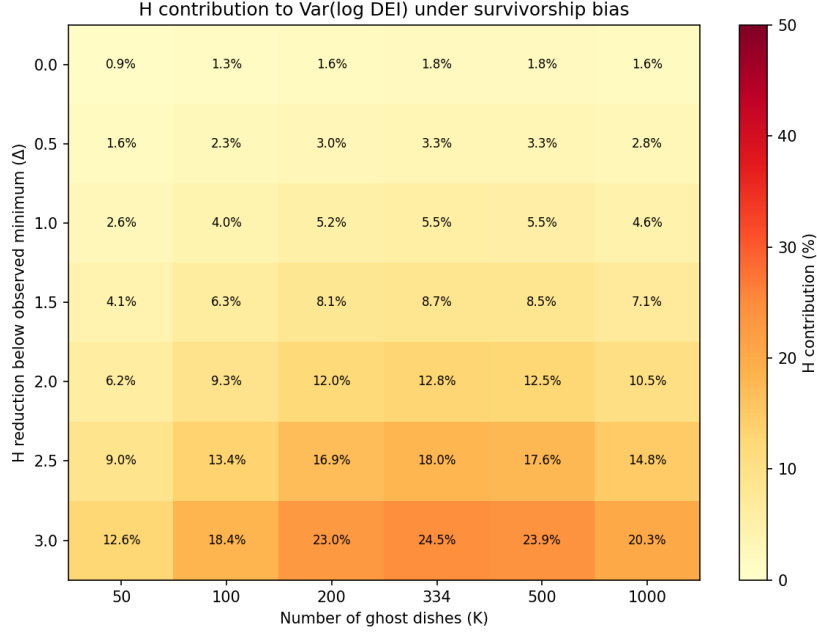


Figure 11: Supplementary Figure S11 | Survivorship bias bounds. H ’s contribution to $\text{Var}(\log(\text{DEI}))$ under various ghost-dish scenarios.

7 Multi-dimensional hedonic assessment

Hedonic experience encompasses more than taste; satiety and comfort contribute to dining satisfaction. We extract satiety signals from review text using keyword detection (“filling”, “hearty”, “satisfying”, “light”, “heavy” etc.) and construct a composite:

$$H_{\text{composite}} = (1 - w) \cdot H_{\text{taste}} + w \cdot H_{\text{satiety}}$$

Table 6: Supplementary Table S6 | Satiety sensitivity analysis.

Satiety weight w	H % of Var	Rank ρ vs original	Top rank gainer
0.0	0.3	1.000	—
0.1	0.3	1.000	dolma +3
0.3	0.5	0.998	dolma +11
0.5	1.5	0.993	coq au vin +13
1.0	11.3	0.946	dolma +11

8 Geographic stability

Our data spans 14 US states, but is geographically concentrated (state Gini = 0.461, city Gini = 0.897). We assess whether DEI rankings are stable across geography.

State-level DEI stability. For the five states with $> 5,000$ mentions each (PA, FL, LA, TN, MO), we compute state-specific H and DEI rankings. State DEI Spearman ρ with the overall ranking exceeds 0.993 for all five states, despite H CV ranging from 4.7% to 6.0%.

City-level H consistency. Among 19 cities with ≥ 500 mentions and ≥ 30 dishes, pairwise H correlations average

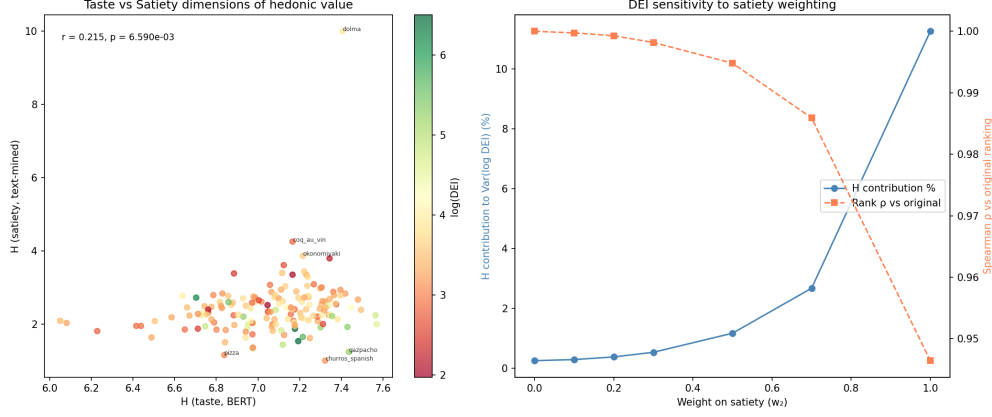


Figure 12: Supplementary Figure S12 | Taste hedonic vs satiety index. Scatter plot of H_{taste} versus H_{satiety} across 334 dishes. The satiety dimension has much higher CV (33% vs 4%) but low weight in composite.

$\rho = 0.232$ (range: -0.209 to 0.609), with 58/105 pairs significant at $p < 0.05$. The modest city-level H correlation reflects sampling noise rather than true geographic heterogeneity, since DEI rankings (dominated by E) remain highly stable.

9 Refinement curve measurement resolution

9.1 Minimum detectable difference

The within-dish standard deviation of H is 1.664, yielding a median Minimum Detectable Difference (MDD) of 0.208 at 95% confidence. Among 165 within-family pairwise comparisons across 15 dish families:

- 91 (55%) are significant at $p < 0.05$
- 54 (33%) survive Bonferroni correction
- Effect sizes: 112 negligible (68%), 50 small (30%), 3 medium (2%), 0 large
- Mean Cohen’s d : 0.164

9.2 Cross-platform refinement validation

We estimate family-level hedonic elasticity α on Google Local and TripAdvisor reviews (in addition to Yelp) for 13 families. Results are noisy: Google shows mean $\alpha = 0.35$, Yelp $\alpha = 0.12$, and TripAdvisor $\alpha = -0.02$. The inconsistency across platforms, combined with small within-family effect sizes, places the refinement curve finding in the category of suggestive rather than definitive evidence.

9.3 Price-tier analysis

Due to sparse price-range metadata in the Yelp Open Dataset (0 of 70,617 restaurants with usable price data in our filtered sample), we cannot stratify by price tier. This limitation is acknowledged in the main text.

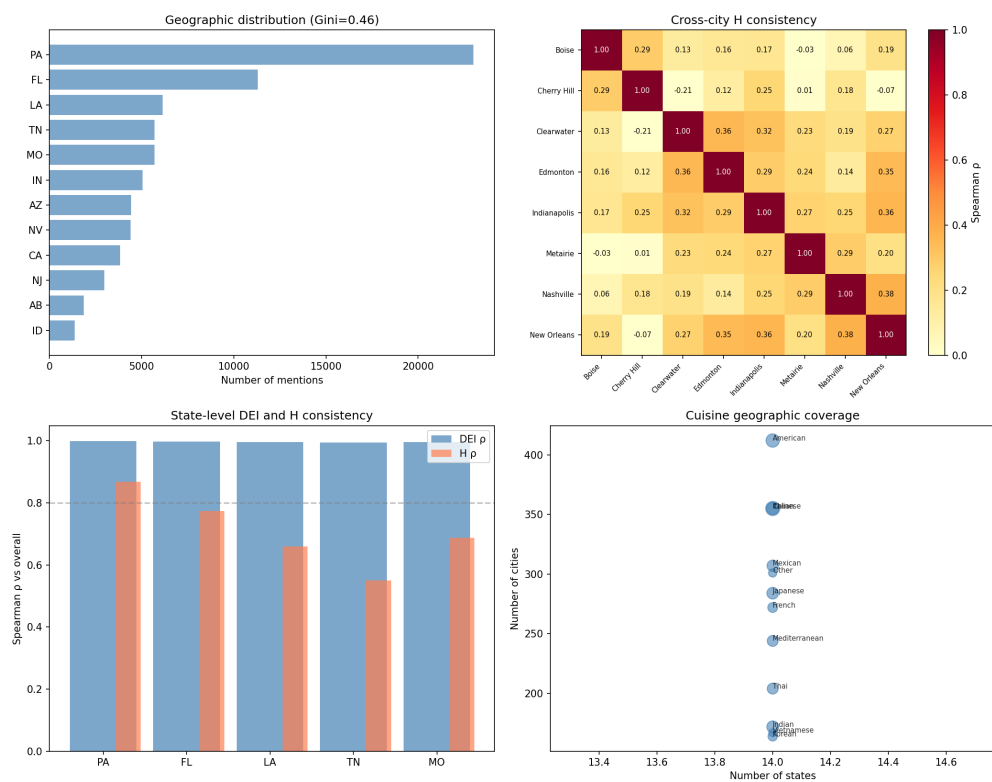


Figure 13: Supplementary Figure S13 | Geographic distribution of review data. Heatmap of mention counts across states and cities.

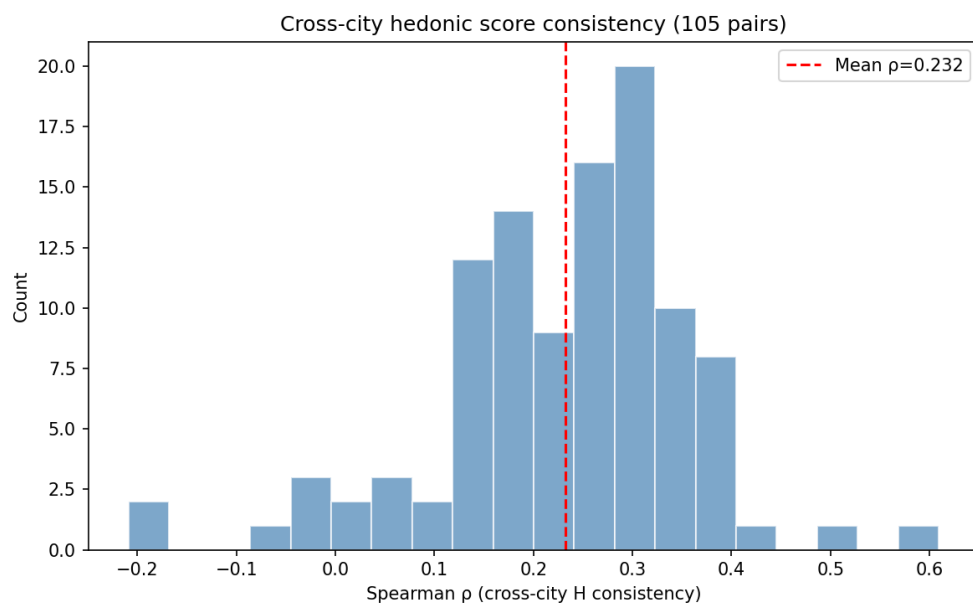


Figure 14: Supplementary Figure S14 | City-level hedonic score stability. Pairwise city H correlations showing moderate consistency across urban markets.

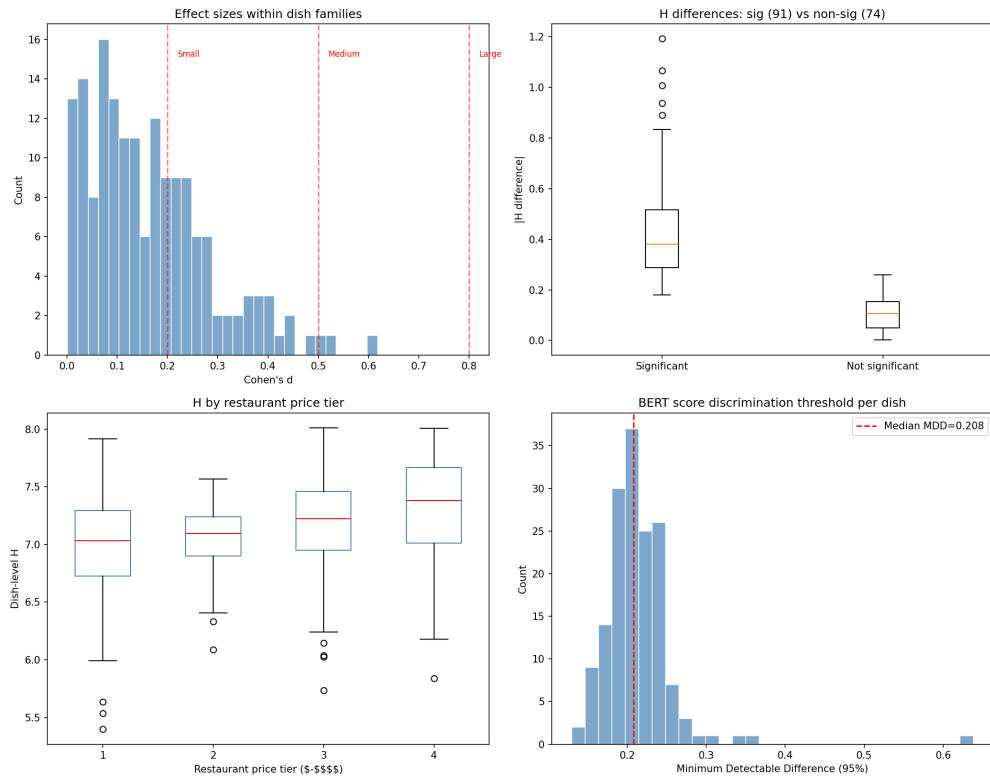


Figure 15: Supplementary Figure S15 | Refinement resolution diagnostic. Within-family pairwise effect sizes. 68% of pairs have negligible H Cohen's d (< 0.2).

10 Additional validation

10.1 Cross-platform hedonic validation

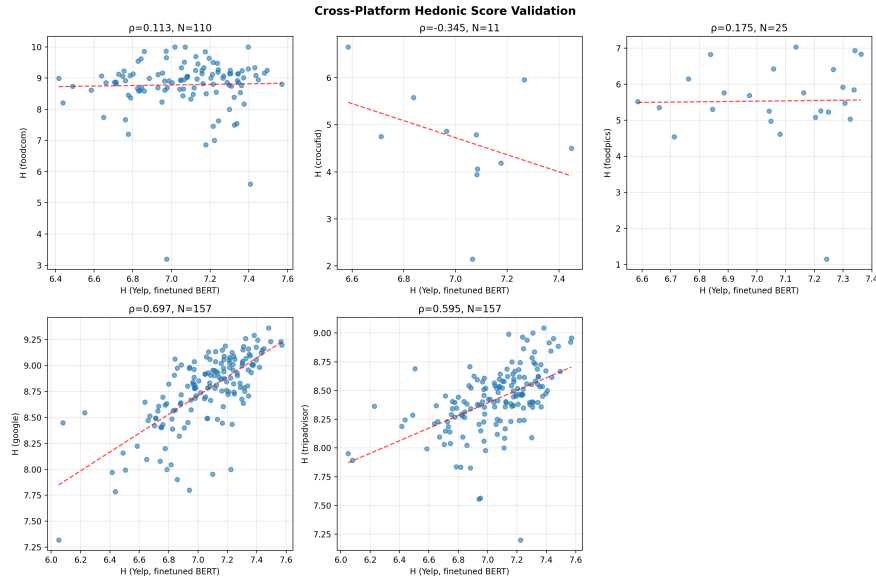


Figure 16: Supplementary Figure S16 | Cross-platform hedonic validation. Yelp H versus Google Local and TripAdvisor H .

10.2 Temporal stability

10.3 Monte Carlo rank stability

10.4 Policy weight sensitivity

10.5 Sample expansion

The 176 expanded dishes show a $\log(\text{DEI})$ distribution indistinguishable from the original 158 (KS $D = 0.073$, $p = 0.73$).

11 Additional figures

References

- [1] Lawless, H. T. & Heymann, H. *Sensory Evaluation of Food: Principles and Practices* (Springer, 2010).
- [2] Drewnowski, A. Defining nutrient density: development and validation of the nutrient rich foods index. *J. Am. Coll. Nutr.* **28**, 421S–426S (2009).

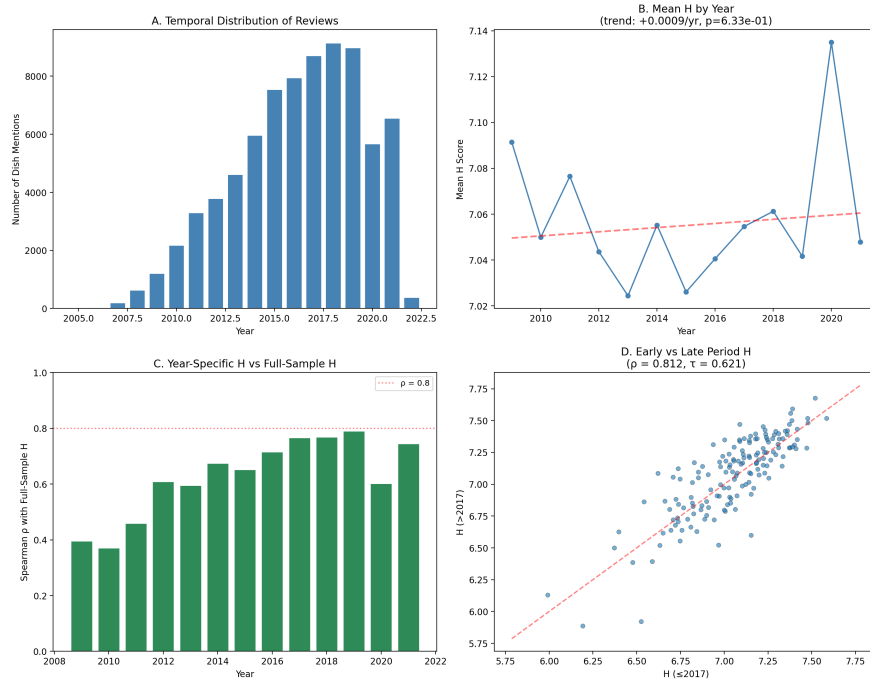


Figure 17: Supplementary Figure S17 | Temporal stability of hedonic scores. H drift: +0.00091 per year ($p = 0.63$), split-period reliability $\rho = 0.812$.

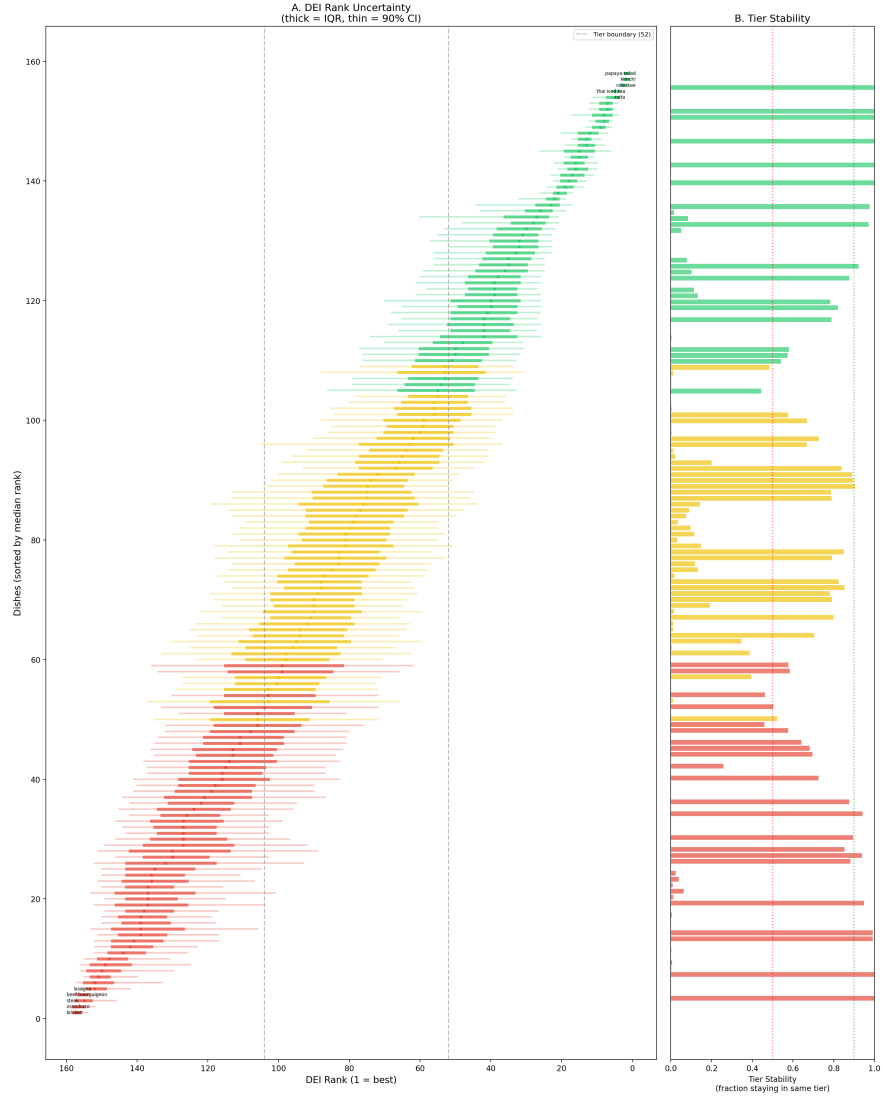


Figure 18: Supplementary Figure S18 | Monte Carlo rank uncertainty. Bootstrap-simulated rank distributions for top and bottom dishes. Top-5 and bottom-5 rankings are stable ($\text{IQR} \leq 3$).

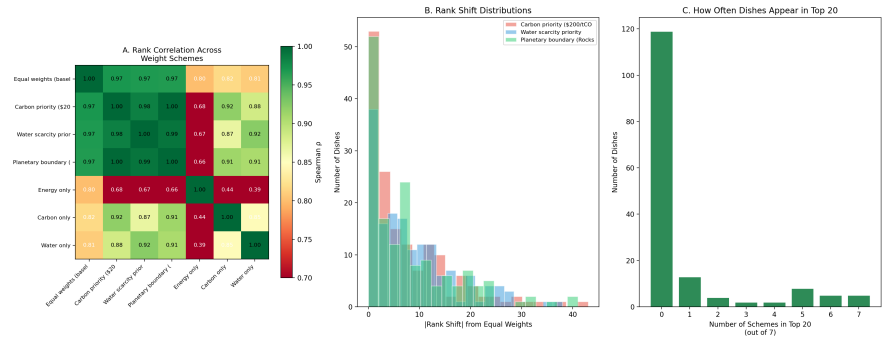


Figure 19: Supplementary Figure S19 | Policy weight sensitivity. DEI rank stability under varying weights w in $\log(\text{DEI})_w = w \cdot \log H - (1 - w) \cdot \log E$. Mean $\rho = 0.828$.

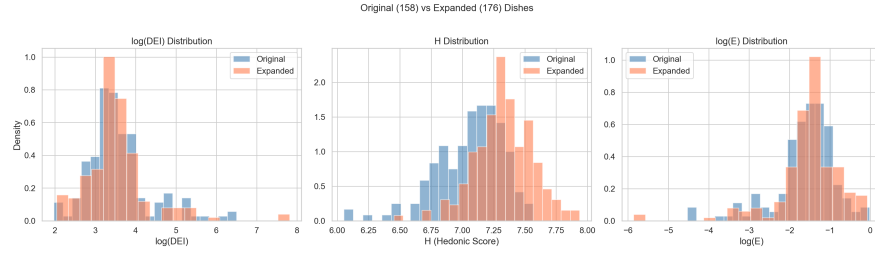


Figure 20: Supplementary Figure S20 | Sample expansion: original vs expanded dishes. $\log(\text{DEI})$ distributions for the original 158 and expanded 176 dishes.

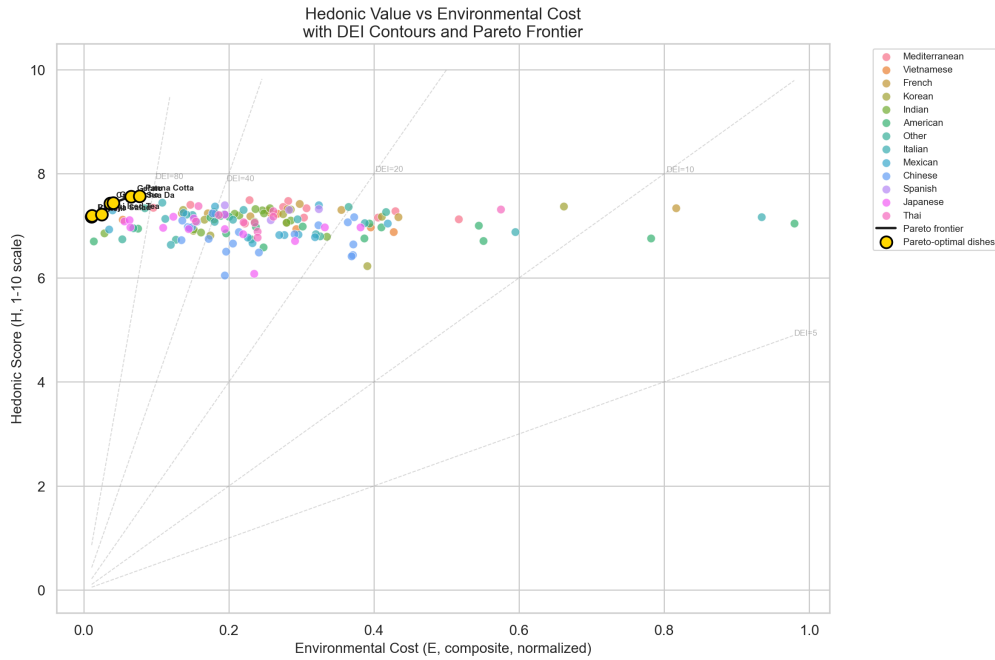


Figure 21: Supplementary Figure S21 | H versus E scatter. The lack of correlation between hedonic score and environmental cost ($r \approx 0$) is the key empirical finding underlying E -dominance.

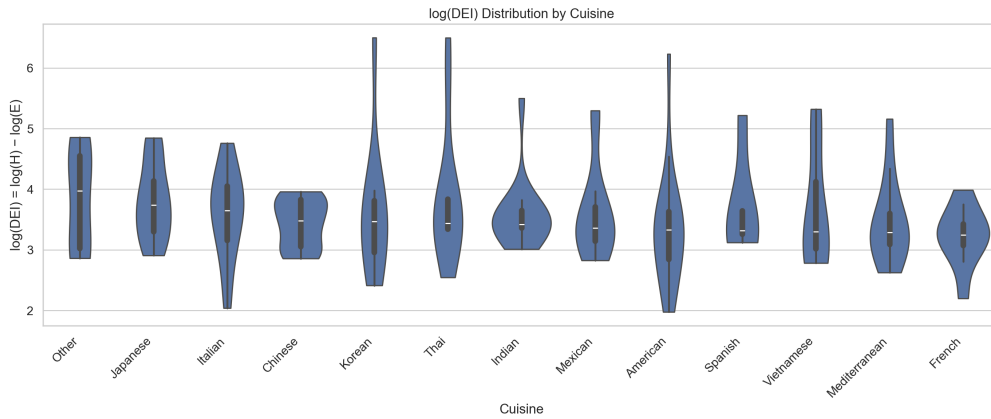


Figure 22: Supplementary Figure S22 | DEI by cuisine. Violin plots showing DEI distribution within each cuisine.

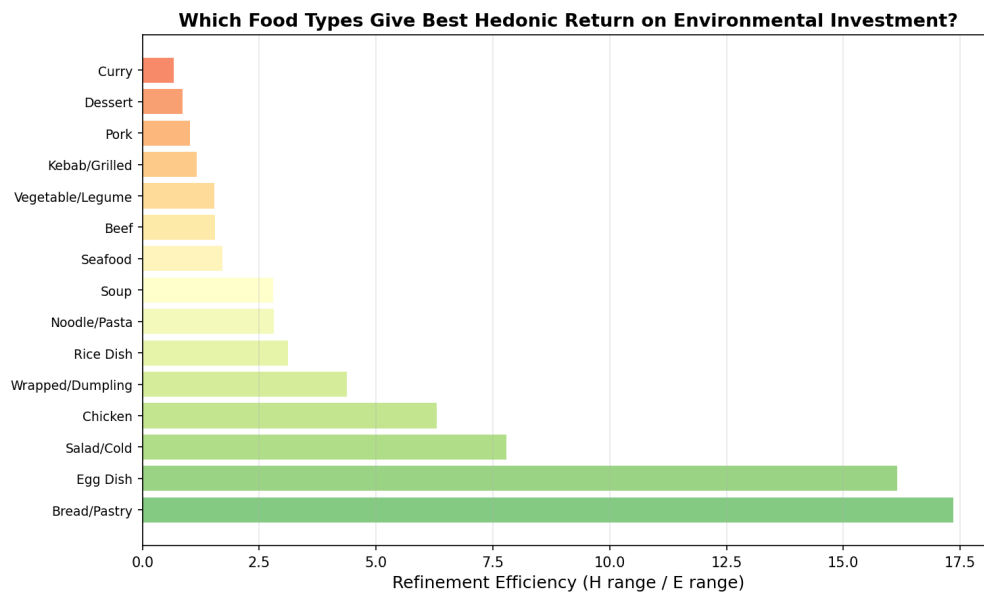


Figure 23: Supplementary Figure S23 | Hedonic waste by dish family. The fraction of environmental cost that yields no hedonic gain, estimated from the refinement cost curve.

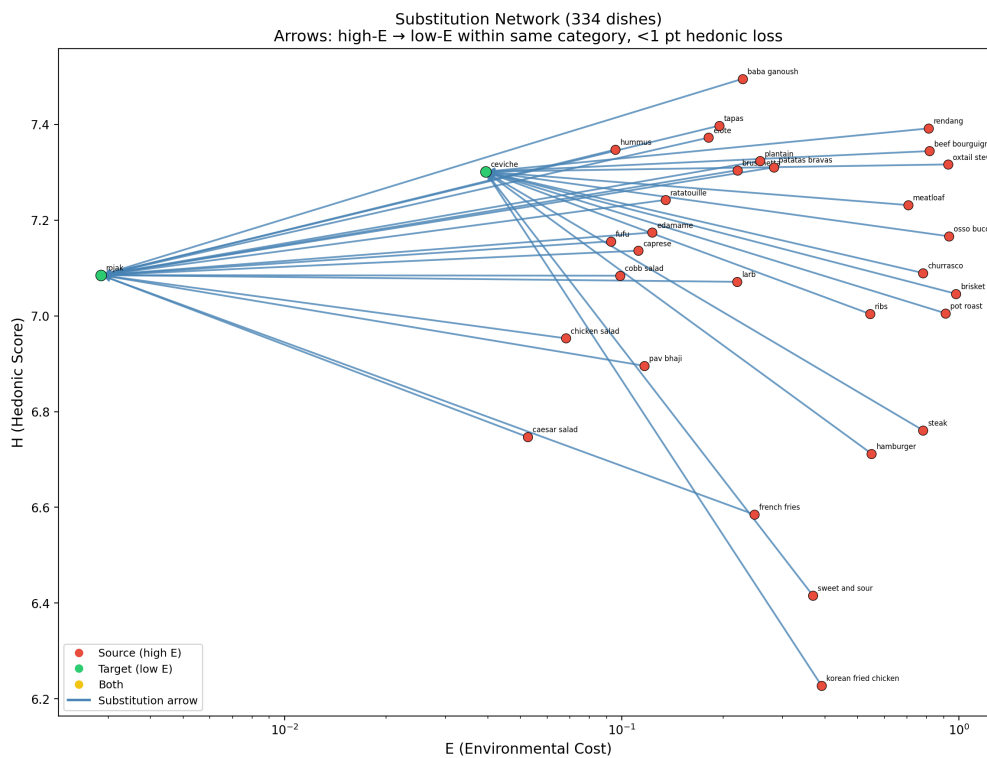


Figure 24: Supplementary Figure S24 | Substitution network. Network graph of viable dish substitutions meeting the four-constraint criterion.