# Text-based inference of moral sentiment change
# Supplementary Material

**Jing Yi Xie**[*], **Renato Ferreira Pinto, Jr.**[*], **Graeme Hirst**, **and Yang Xu**

Department of Computer Science, University of Toronto, Toronto, Canada
`jingyi.xie@mail.utoronto.ca`, `{renato,gh,yangxu}@cs.toronto.edu`

## A  Model evaluation on historical corpora

Table S1 shows average seed word classification accuracy for all models on each moral classification tier for the entire 1800-1999 period. We performed this historical evaluation using Google N-grams embeddings only, since COHA did not contain seed word embeddings for all moral categories in the earliest decades. Similar to the evaluation in modern times, we observe consistent performance above chance in all models.

| Model | Relevance | Polarity | Category |
|-------|-----------|----------|----------|
| Random | 0.50 | 0.50 | 0.10 |
| Centroid | 0.82(1) | 0.89(2) | **0.64(2)** |
| N. Bayes | **0.83(2)** | 0.90(1) | 0.59(2) |
| 1-NN | 0.80(1) | 0.89(2) | 0.57(3) |
| 5-NN | **0.83(1)** | **0.91(1)** | 0.62(3) |
| KDE | **0.83(1)** | **0.91(1)** | **0.64(2)** |

Table S1: Moral seed word classification accuracy for moral relevance, moral polarity, and fine-grained moral categories across models using Google N-grams embeddings. Mean accuracies and standard deviations across all decades in 1800–1999.

## B  Time-shuffled regression analysis

Figure S1 shows the diachronic coefficients for word frequency, length, and concreteness on moral relevance change from multiple regression analaysis, compared to the distribution of coefficients obtained from a control condition of 1,000 shuffled time courses. We observe that all effects become non-significant under the shuffled condition, and a strong diachronic effect of concreteness compared to the control suggests the relevance of this psycholinguistic variable for diachronic change in moral sentiment.
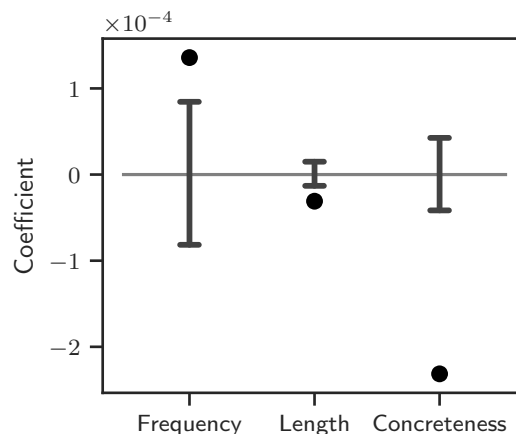
---

[*]Equal contribution.



Figure S1: Multiple regression coefficients of moral relevance change versus control condition. Points show diachronic coefficients, and error bars show 1 standard deviation around the mean control coefficients.

## C  Additional time courses

Figure S2 illustrates the moral trajectories for gay. Although gay originally referred to that which is "skittish/spirited" in the 1800s, it later acquired the meaning of "pertaining to homosexuality" in the 1930s (Kay et al., 2019). This shift in semantics to a subject of more controversy is reflected in the moral relevance plot.
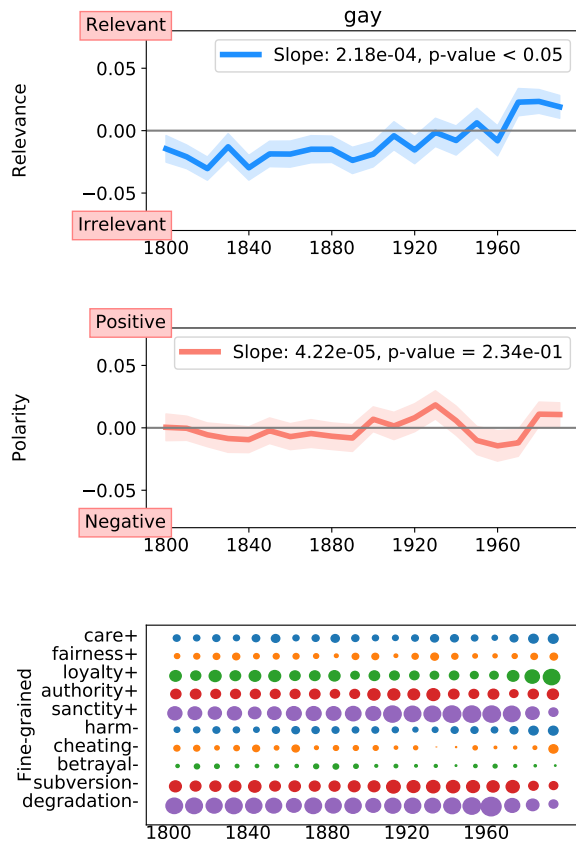
Figure S2: Moral sentiment time courses of *gay* at each of the three levels, inferred by the Centroid model. Time courses at the moral relevance and polarity levels are in log odds ratio, and those for the fine-grained moral categories are represented by circles with sizes proportional to category probabilities.

# References

Christian Kay, Jane Roberts, Michael Samuels, Iren Wotherspoon, and Marc Alexander (eds.). 2019. *The Historical Thesaurus of English*. Glasgow: University of Glasgow. Version 4.21.