

# Systematic Gender Asymmetries in Aesthetic Judgments: An Observational Study of Book Reviewer Preferences

**Lassen, Ida Marie S.**

idamarie@cas.au.dk

Center for Humanities Computing, Aarhus University, Denmark

**Bizzoni, Yuri**

yuri.bizzoni@cc.au.dk

Center for Humanities Computing, Aarhus University, Denmark;  
School of Communication and Culture - Comparative Literature,  
Aarhus University, Denmark

**Thomsen, Mads R.**

madsrt@cc.au.dk

School of Communication and Culture - Comparative Literature,  
Aarhus University, Denmark

**Nielbo, Kristoffer**

kln@cas.au.dk

Center for Humanities Computing, Aarhus University, Denmark

## Introduction

Aesthetic judgments are notoriously complex and subject to considerable variation because aesthetic objects are complex, aesthetic preferences are subjective, and there is a general lack of shared measurement. Literary quality, for instance, can be considered one of the most subjective fields of evaluation, and variation is mostly attributable to noise introduced by individual preferences. Yet the perception of literary quality from many readers over time does show convergent trends: communities tend to establish and update canons (Guillory 1993); specific texts and narratives manage to remain popular (Stephens / McCallum 2013) despite the changing of fashions and political phases, and certain author names become eponymous of literary quality in different countries and throughout the social spectrum (Bloom 2014). Some facets of literary quality can be explained in terms of the literary content (e.g., predictability of content, coherence of the narrative), while others depend on sociocultural priors that introduce systematic variation in aesthetic judgments. It is the latter that are the object of this study, specifically the possible effects of gender on the assessment of literary quality as an example of how aesthetic judgments can be biased by contextual factors.

The problem of literary quality's subjective status becomes even more intriguing when we turn to the challenge of its computational assessment. Most studies assume the possibility of a one-dimensional ground truth by modelling literary quality as a single rating or class associated with a text (Ferrer 2013; Wang et al. 2019; Walsh / Antoniak, 2021). These ground truths are retrieved from various sources: literary critics, book sale numbers, best-seller lists, or crowd-sourced reader opinions. Such approaches have several limitations: Relying only on experts' judgment (e.g., awards, prestigious reviews) biases the model to reflect their preferences, but striving for representativity by crowd-sourcing opinions ends up ignoring important differences in the readers' population. To properly understand the value of these ground truths and develop standardised measures of quality, it is necessary to model possible sources of bias.

## Methods

### Data

The data set covers book reviews published in Danish media in the years 2010-2021. The data were retrieved from the online platform bog.nu's API which collects book reviews published in Danish media. See table 1 for a brief overview of the data set.

Data set overview			
No. of reviews	22 386	No. of different titles	8 406
Male reviewers	14 161	Male authors	5 000
Female reviewers	8 225	Female authors	3 406

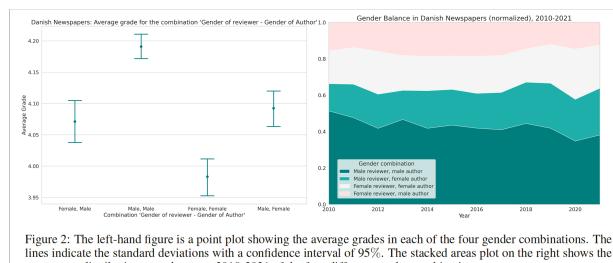
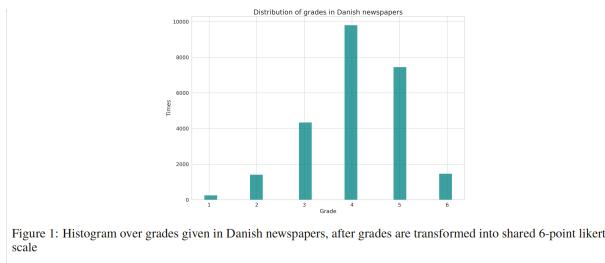
Table 1: An overview of the dataset analysed in this paper.

## Grade Transformation and Estimation

As different media use different grading scales, the grades on bog.nu are transformed to a 100-point grading scale. This approach, however, results in a sparse distribution of grades as the use of the original grading scales maps onto different intervals on the 100-point scale. Instead of this naive approach, we have used the original grade and applied a linear transformation to map all grades to a shared 6-point scale<sup>1</sup>. Mapping from an a-b-point scale to a 1-6-point scale:

$$Y = \lfloor (B - A) \frac{(x - a)}{(b - a)} + A \rfloor = \lfloor (6 - 1) \frac{(x - a)}{(b - a)} + 1 \rfloor = \lfloor \frac{5(x - a)}{b - a} + 1 \rfloor \quad (1)$$

Figure 1 shows the distributions of grades in Danish Newspapers transformed into a shared 6-point scale. See equation (1). Some media do not provide a grade in a given review but only a qualitative review. Bog.nu does, however, provide a quantification of the review, which is estimated by an expert. For reviews written in Danish newspapers, this estimation procedure is used in less than 25% of the cases.



## Gender Retrieval

The original data set from bog.nu does not contain gender for all authors and reviewers. We used an API<sup>2</sup> built on binary gendered name lists to retrieve the missing gender variables. We are aware of the problems with this method and how it rules out other gender identities (Dev et al. 2021). However, a binary representation of gender is necessary for our analysis to understand the existing structures between men and women in contemporary society.

## Model

In order to estimate the effect, the relative effect of author and reviewer gender on the reviewer assigned grade (six-point scale), we solve the following linear model with the null model that #author =  $\beta$ reviewer = 0:

$$y_i^{\text{grade}} = x_i \beta_{\text{author}} * x_i \beta_{\text{reviewer}} + \epsilon_i \quad (2)$$

Where  $y_i$  is the grade of review  $i$ ,  $x_i$  is the predictor value (gender) of review  $i$ ,  $\beta$  represents unknown parameters, and  $\epsilon$  are the error terms.

## Results

Solving for  $\beta$  in the model above with ordinary least square (OLS), we get the following results, see Table 2. The model and all contrasts are statistically significant ( $p < .0001$ ) Conceptually, same-gender (male \* male, female \* female) reviews span the extreme values, while opposite gender (female \* male, male \* female) reviews represent the middle of the distribution, see Fig. 2. Female reviewers reviewing female authors account for the, on average lowest grade. Male reviewers reviewing male authors result in the highest grade, with a 0.2 average grade increase. Opposite-gender reviews are statistically speaking, indistinguishable, but they differ on average 0.1 grade point from same-gender scoring.

Gender combination (reviewer * author)	Average grade	SD	t	p <  t	CI 95%
Intercept [female * female]	3.9829	0.015	268.611	0.0001	[3.954, 4.012]
female * male	0.0882	0.022	3.947	0.0001	[0.044, 0.132]
male * female	0.1095	0.021	5.187	0.0001	[0.068, 0.151]
male * male	0.2081	0.018	11.553	0.0001	[0.173, 0.243]

Table 2: Results for an OLS predicting grade based on the gender combination: reviewer and author. The intercept is the gender combination of female reviewer and female author. The t-test shows that all results are statistically significant.

## Discussion

In line with the results in Thelwall (2019) and Touileb et al. (2020), our results show that the gender of authors as well as of reviewers play a role in literary reviews written in Danish newspapers. In particular, the results above show that men review same-gender more positively than the opposite gender and that women show the reverse pattern, that is, same-gender is reviewed more negatively than the opposite gender. A partial explanation of the behaviour is that if males display a same-gender preference and male reviewers make up the majority of newspaper reviewers, then the gender minority adapt to this preference and develops a same-gender antagonism. At a more general level, the same gender preference of men in aesthetic judgment may reflect a cultural gender antagonism that follows a long historical trajectory. The female opposite gender preference is very likely to follow the same cultural gender antagonism.

We are not arguing that newspapers follow an explicit exclusionary strategy formulated by male reviewers and editors. Whenever we make a judgment, there are essentially two sources of error, bias and noise. While noise is randomly distributed and lacks a systematic explanatory mechanism, biases are systematic and can be explained in terms of a mechanism. Demographic biases often originate in a systemic oppression of minority groups. For the specific review cases, the results are likely to mirror existing oppressive structures in the society such as those found in Dane (2020); MacNell et al. (2015); Johnson and Kirk (2020); Goldin and Rouse (2000); Booth and Leigh (2010); Cole et al. (2004). We expect that majority groups, in general, will define norms and values that result in biased judgments irrespective of societal domain. Denmark is an excellent example of how more or less explicit biases persist in the face of societal countermeasures (ex. equal rights, parity policies) and in general high gender equality.

## Notes

1. It should be noted that most six-point scales have become the standard in many review outlets.
2. <https://genderize.io/>

## Bibliography

- Bloom, Harold** (2014): *The western canon: The books and school of the ages*. Houghton Mifflin Harcourt.  
**Booth, Alison / Leigh, Andrew** (2010): “Do employers discriminate by gender? A field experiment in female-dominated occupations”, in: *Economics Letters* 107.2, pp. 236–238. <https://www.sciencedirect.com/science/article/abs/pii/S0165176510000352>

**Cole, Michael S. / Hubert, S. Feild / Giles, William F.** (2004): “Interaction of recruiter and applicant gender in resume evaluation: a field study”. In: *Sex Roles* 51.9, pp. 597–608. <https://link.springer.com/article/10.1007/s11199-004-5469-1>

**Dane, Alexandra** (2020): *Gender and Prestige in Literature*. Springer.

**Dev, Sunipa / Monajatipoor, Masoud / Ovalle, Anaelia / Subramonian, Arjun / Phillips, Jeff M. / Chang, Kai-Wei** (2021): “Harms of Gender Exclusivity and Challenges in Non-Binary Representation in Language Technologies”, in: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* <https://aclanthology.org/2021.emnlp-main.150/>

**Ferrer, Carolina** (2013): “Canonical values vs. the Law of large numbers: The Canadian Literary Canon in the Age of Big Data”, in: *Rupkatha Journal on Interdisciplinary Studies in Humanities*, 5.3, pp. 81–90. <https://www.semanticscholar.org/paper/Canonical-Values-vs.-the-Law-of-Large-Numbers%3A-The-Ferrer/844a4cfbea9fd16a8fbaedcf3bb77f74d64425b>

**Goldin, Claudia / Rouse, Cecilia** (2000): “Orchestrating impartiality: The impact of “blind” auditions on female musicians”, in: *American economic review*, 90.4, pp. 715–741. <https://www.aeaweb.org/articles?id=10.1257/aer.90.4.715>

**Guillory, John** (1993): *Cultural capital: The problem of literary canon formation*. University of Chicago Press.

**Johnson, Stefanie K / Kirk, Jessica F.** (2020): “Dual-anonymization yields promising results for reducing gender bias: A naturalistic field experiment of applications for Hubble Space Telescope time”, in: *Publications of the Astronomical Society of the Pacific*, 132.1009, p. 034503. <https://iopscience.iop.org/article/10.1088/1538-3873/ab6ce0>

**MacNell, Lillian / Adam Driscoll / Andrea N Hunt** (2015): “What’s in a name: Exposing gender bias in student ratings of teaching”, in: *Innovative Higher Education*, 40.4, pp. 291–303. <https://link.springer.com/article/10.1007/s10755-014-9313-4>

**Stephens, John / Robyn McCallum** (2013): *Retelling stories, framing culture: traditional story and metanarratives in children’s literature*. Routledge.

**Thelwall, Mike** (2019): “Reader and author gender and genre in Goodreads”, in: *Journal of Librarianship and Information Science* 51.2, pp. 403–430. <https://journals.sagepub.com/doi/10.1177/0961000617709061>

**Touileb, Samia / Øvrelid, Lilja / Velldal, Erik** (2020): “Gender and sentiment, critics and authors: a dataset of Norwegian book reviews”, in: *Proceedings of the Second Workshop on Gender Bias in Natural Language Processing*. Barcelona, Spain (Online): Association for Computational Linguistics, pp. 125–138. <https://aclanthology.org/2020.gebnlp-1.11/>

**Walsh, Melanie / Antoniak, Maria** (2021): “The Goodreads ‘Classics’: A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism”, in: *Journal of Cultural Analytics* 4, pp. 243–287. <https://culturalanalytics.org/article/22221-the-goodreads-classics-a-computational-study-of-readers-amazon-and-crowdsourced-amateur-criticism>

**Wang, Xindi / Yucesoy, Burcu / Varol, Onur / Eliassi-Rad, Tina / Barabási, Albert-László** (2019): “Success in books: predicting book sales before publication”, in: *EPJ Data Science*, 8.1, pp. 1–20. <https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-019-0208-6>