

# Milestone 7

*Amanda Su*

*4/15/2020*

## Contents

0.1	Abstract . . . . .	1
0.2	Introduction . . . . .	1
0.3	Literature Review and Paper Review . . . . .	1
0.4	Replication . . . . .	2
0.5	Extension . . . . .	2
0.6	Appendix . . . . .	4
	References . . . . .	4

### 0.1 Abstract

So, Long, and Zhu (2019) built a model that tests if novelists marked as “white” versus “black” produce different narratological effects with respect to the interaction of race and religious authority, in particular, the authority of the Bible. The study ultimately finds that when a white writer cites the Bible, it is less likely that she/he quotes it in a social context compared to when she/he writes about non-Bible related topics. However, when a black writer cites the Bible, it is more likely that she/he quotes it in a social context. I was able to successfully replicate the results of the authors’ paper. For my extension, I decided to reconstruct the paper’s primary model using a Bayesian approach. I found that the results of the model were largely the same the as that of the original. These robust results corroborate the author’s conclusions about how race and writing intersect across more than a century of US fiction

### 0.2 Introduction

TO BE WRITTEN

Can computational methods tell us anything new and interesting about how racial difference is expressed in literature? Do authors of different racial identifications (for example, “white” versus “black”) consistently use different patterns of language, style, and narrative, and if so, what are these patterns? Do they remain stable or change over time?

### 0.3 Literature Review and Paper Review

This scholarly project bridges two scholarly fields historically seen as incompatible: cultural analytics (also known as “computational criticism”) and critical race studies. It does so by discovering generative points of contact between data science and critique, two sets of methods typically viewed as antithetical. Cultural analytics is an emerging field wherein humanist scholars leverage the in- creasing availability of large digital materials and the affordances of new computa- tional tools. This allows them to study, for example, semantic and narratological patterns in the English-language novel at the scale of centuries and across tens-of-thousands of texts. While cutural analytics scholars have taken on an expanding array of topics, including genre and cultural prestige, the topic of race and racial difference has remained relatively understudied.

Since computational methods demand the quantification of one’s objects of study, it’s likely easier to accept measuring a novel’s popularity by sales figures or classifying its genre by diction than labeling it according to discrete racial identifiers. Such labeling is an affront to critical race studies, the mission of which is the deconstruction of racial categories. As such, recent scholarship on the relationship between computation and race has been critique-oriented. Scholars of science and technology, such as Cathy O’Neil and Safiya A. Noble, have documented how computational algorithms used by banks and online search engines intensify racial stratification and oppression by articulating racial minorities as fixed, quantified types that reinforce existing patterns of social inequality. Tara McPherson has shown that the history of modern computation is deeply intertwined with the history of racial formation in the US since the 1960s. The authors of this paper uses a computational model to study race and literature in order to determine both the model’s affordances and its inadequacies. I make use of Jarrett (2007), Stepto (2001), Spillers (2003), and Earhart (2015).

TO BE EDITED AND COMPLETED

## 0.4 Replication

To test their theory about novelists of different races producing different narratological effects in their works with respect to the Bible, the authors constructed a model that explains whether or not a text is “social” as a function of the author’s gender, race, whether or not they cited the Bible as a control variable, the interaction of the race and bible variables, and the random effect for each novel. For the purposes of this paper, moments of sociality are defined as the presence of two or more characters engaged in dialogue or interaction.

I was able to successfully replicate every aspect of the paper.

## 0.5 Extension

**Table 1: Effect of Author Gender, Race, Bible Citation, and Race and Bible’s Interaction on the Sociality of a Text**

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu  
 % Date and time: Thu, Apr 23, 2020 - 00:18:35

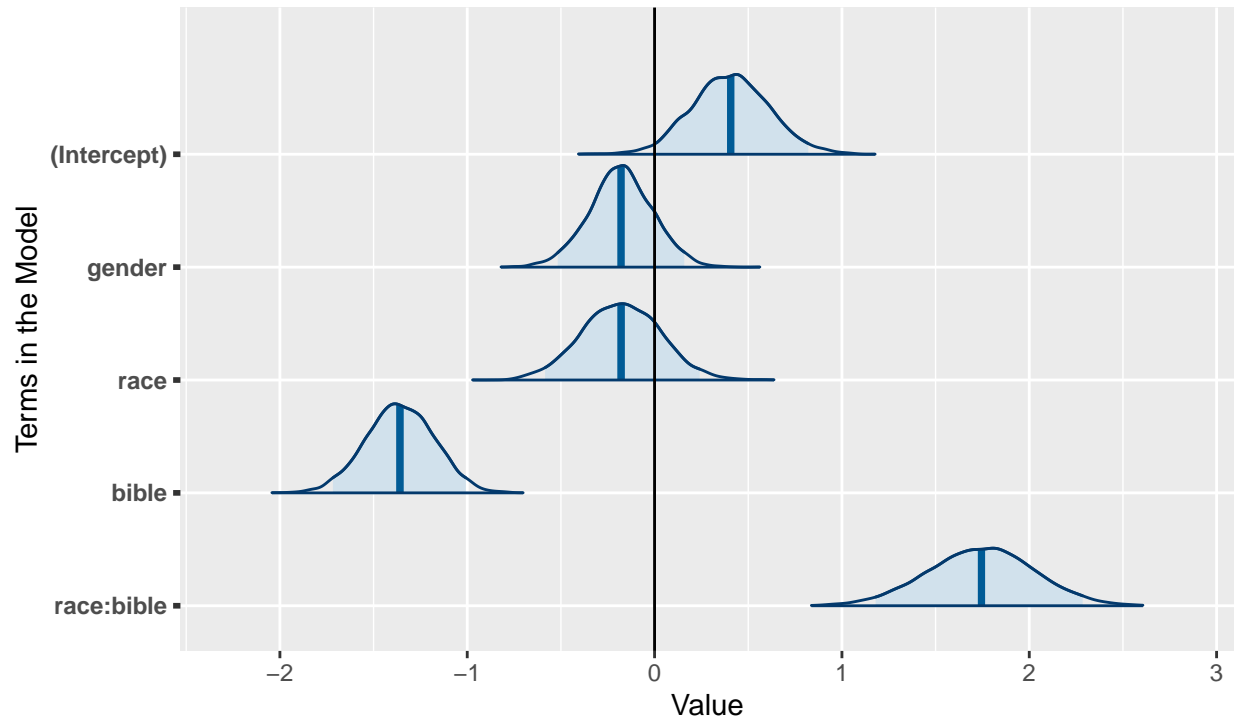
Table 1: Effect of Author Gender, Race, Bible Citation, and Race and Bible’s Interaction on the Sociality of a Text

Statistic	Mean	St. Dev.
Intercept	0.405	0.208
Author’s Gender	−0.179	0.170
Author’s Race	−0.178	0.209
Citation of Bible	−1.359	0.182
Race:Bible	1.742	0.282

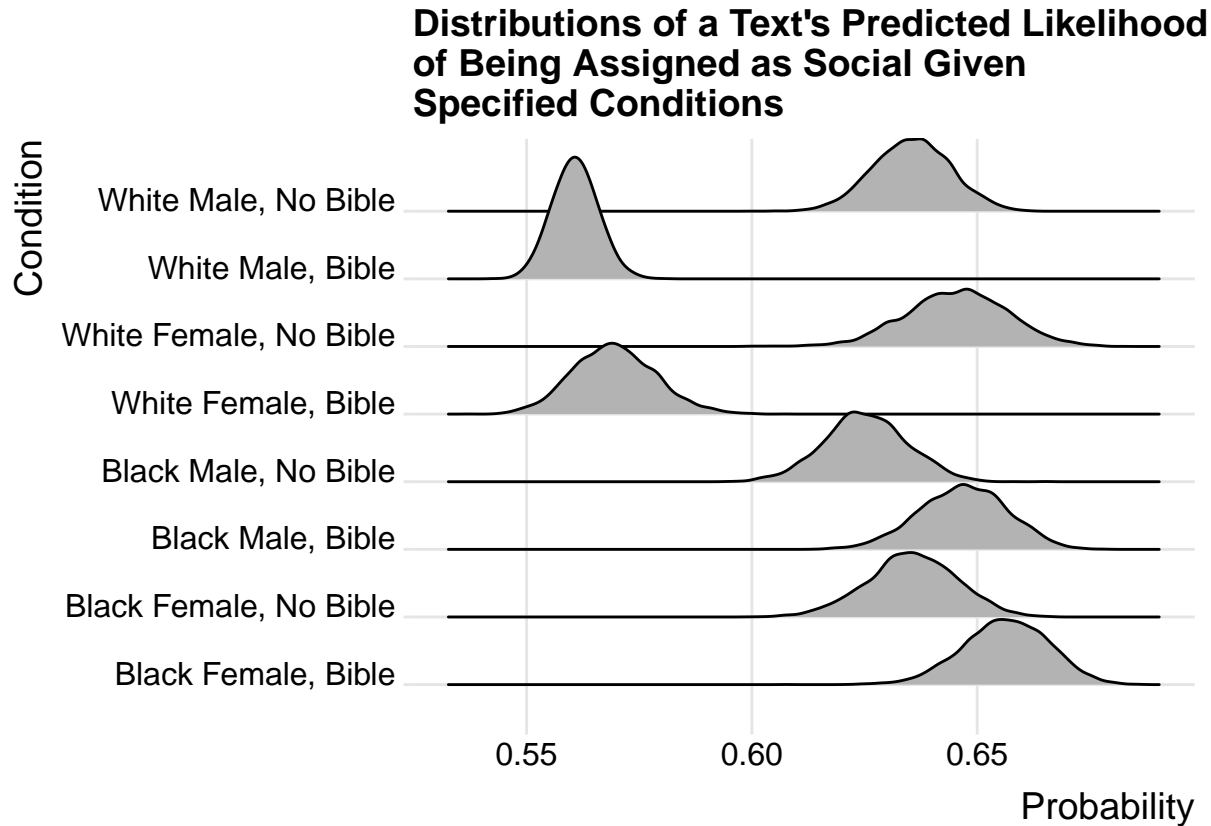
**Graphic 1: Distribution of Coefficients on Author Gender, Race, Bible Citation, and the Interaction of the Latter Two**

### Distributions of Terms in Modified Model

Model explains a text's sociality as a function of author race, gender, citation of Bible, and the interaction of race and bible citation



Graphic 2: Distribution of Predicted Likelihoods of a Text Being Social Given An Author's Race, Gender, and Citation of the Bible in Their Work



Whereas So, Long, and Zhu (2019) decided to perform a maximum likelihood estimation of generalized linear models to determine the predicted values of model coefficients, I perform a full Bayesian estimation to find the average expected values for coefficients. King, Tomz, and Wittenberg (2000) wrote that expected value averages are preferable to predicted values because the latter contains both fundamental and estimation uncertainty, whereas the former only has to account for the estimation uncertainty caused by not having an infinite number of observations. As a result, predicted values have a larger variance than expected values. I ultimately found that the primary results of the authors' paper are largely unchanged even when using a Bayesian approach to create the model.

## 0.6 Appendix

Results from So, Long, and Zhu (2019) were successfully replicated.<sup>1</sup>

## References

- Earhart, Amy. 2015. *Traces of the Old, Uses of the New*. Ann Arbor, MI: University of Michigan Press.
- Jarrett, Gene Andrew. 2007. "Addition by Subtraction: Toward a Literary History of Racial Representation." *Legacy*.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation" 44 (2). *American Journal of Political Science*: 347–61.
- So, Richard Jean, Hoyt Long, and Yuancheng Zhu. 2019. "Race, Writing, and Computation: Racial Difference and the Us Novel, 1880-2000." *Journal of Cultural Analytics*.

<sup>1</sup>All analysis for this paper is available at my Github repository.

Spillers, Hortense. 2003. *Black, White, and in Color: Essays on American Literature and Culture*. Chicago: University of Chicago Press.

Stepso, Robert. 2001. *From Behind the Veil: A Study of Afro-American Narrative*. Urbana, IL: University of Illinois Press.