

SDS 291 Final Project

Laura Edwards, Cameron Darling, Mags McLaughlin, and Natalie Szewczyk

Abstract

Introduction

Survivor is one of cable television's longest-running reality competition shows. Advertised as a 'microcosm for our real world' by executive producer Jeff Probst (Wigler 2019), every season, about 20 people are taken to an exotic location, given little supplies, and divided into tribes. They are then tasked with creating a new society amongst themselves while competing in physical and mental challenges, voting each other out along the way, each player working towards the goal of being the last person remaining to win one million dollars.

Being a competition show focused heavily on the social interactions of its players, the topic of race and its impact on player experience and player outcomes has been discussed at length. Excluding Season 13, Cook's Island, in which players were segregated by race as part of the season's theme, Survivor's season casts have been historically dominated by white players. This likely contributed to many players of color having negative experiences with their castmates going unaddressed, as detailed in an episode of Morning Edition by NPR (Deggan 2020), as well as a Zoom roundtable featuring several past black Survivor players.

As a reaction to the broader cultural reaction to the George Floyd Protests, as well as a petition released on Juneteenth 2020 from some of Survivor's past black players, in November of 2020, CBS announced that beginning in Season 41, 50% of the Survivor castaways would be POC. Diversity measures like CBS' have been implemented in workplaces before, and have been shown to improve firm performance, team creativity, among other similar metrics in a wide variety of contexts (Gomez et al. 2019, Kumar & Gupta 2023). However, do these same benefits translate at a qualitative level in a reality television context? Our study aims to, in part, evaluate the efficacy of the 2020 Survivor diversity initiative on improving the portrayal of POC players themselves, by addressing the following questions: is there a relationship between whether or not a player had a positive, portrayed* initial attitude towards the game and whether or not the player's season occurred before or after the diversity casting rules were implemented? Do other social demographic factors, such as age and gender of a player, moderate this relationship, if it exists?

Methods

Data Source

Our dataset combines existing data gathered by Survivor fans from the ‘castaway_details’ dataset from the survivoR R Package on GitHub with transcribed confessionals transcripts and sentiment analysis done by our team. The ‘castaway_details’ dataset contains details on a player’s gender, POC identity, date of birth, and other detailed demographic information which is not relevant to this study. Per the package author, this demographic information was collected from the Survivor Wiki, Wikipedia, and from individual package contributors.

Study Population

Our team studied the 184 players’ confessionals from Season 36-45, the five seasons preceding (S36-40) and five seasons immediately following (S41-45) the diversity casting rules being put into effect. Looking at the first four episodes of each season, a player’s first confessional which matched the criteria outlined in Figure I was included in the dataset. Each confessional was then analyzed using the VADER sentiment analyzer (version 3.3.2) in Python and coded as positive (scores higher than 0) or negative (scores 0 and below) to measure a player’s portrayed initial attitude towards the game.

Table 1: POC Status of Study Population in Two Main Study Cohorts

	POC	White
Pre-Diversity Rules (S36-40)	32	64
Post-Diversity Rules (S41-45)	52	36

Operational Validity of First Confessionals

We used a player’s first confessional which matched the criteria outlined in Figure I, which attempts to isolate confessionals relating to personal strategy, motivations, and relationships, to capture the first impression that viewers receive of the attitude of a player, which, regardless of the player’s overall season story arc, acts as a foundation with which the viewers contextualize the rest of the player’s game.

```
knitr::include_graphics("Flowchart.jpeg")
```

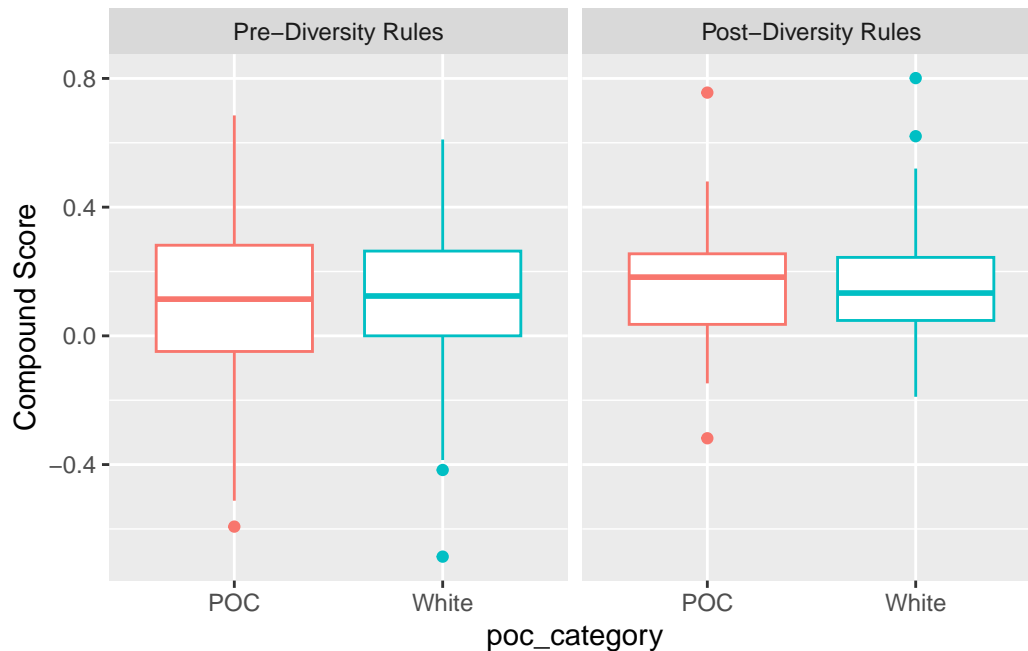


Figure 1: Average Compound Scores of 184 Players from Survivor, Pre- and Post- Implementation of Diversity Casting Rules

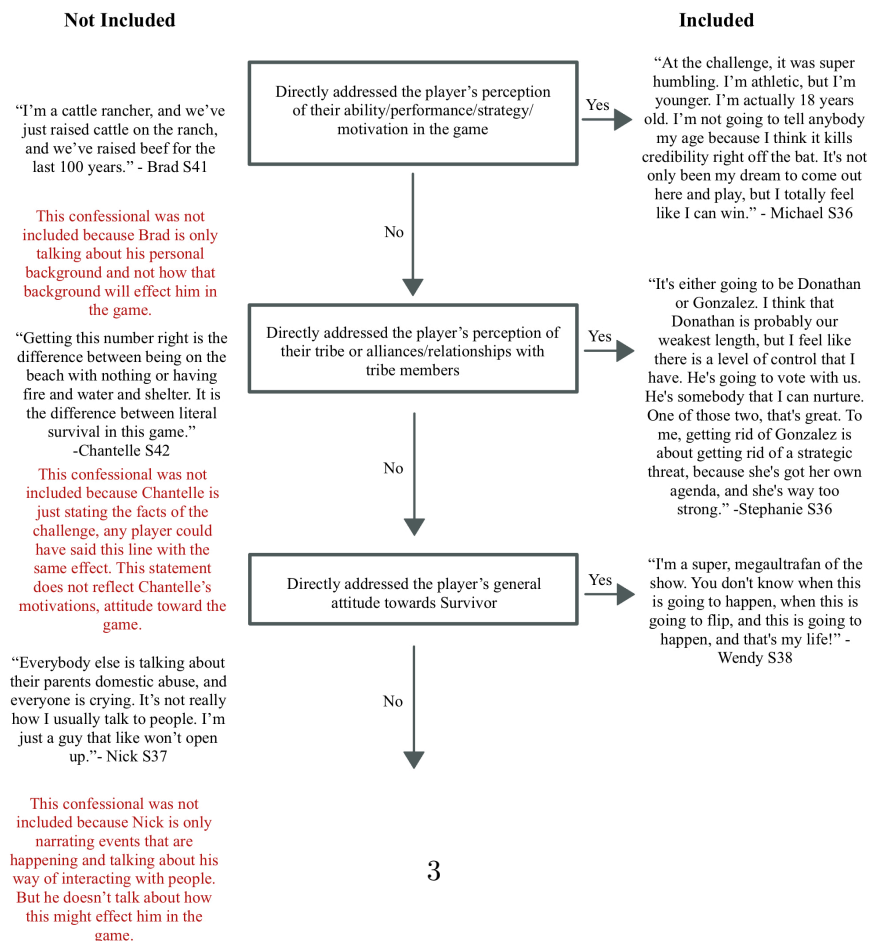


Figure 2: Confessional Selection Diagram

Four players lacked confessionals which met the guidelines outlined in Figure and were coded as negative by default, as our team took a lack of portrayal as having a negative impact on a player’s portrayal.

Dataset

The combined dataset includes information on the gender, age, race, name, and POC status of each player in addition to their confessional, sentiment analysis score, and portrayed initial attitude. Figure 3 is a visual representation of the portrayed attitudes of poc and non-POC players pre- and post- diversity rules. Visually, the stacked bar plots show an increase in the percentage of POC players in the casts post-diversity rules as expected. Of these POC players, the proportion whose initial, portrayed attitudes were analyzed as positive seems to have increased compared to before the diversity casting rules were implemented. However, this increase in positive portrayal also appears with white players post- diversity casting rules as well. This suggests the rough idea that perhaps, the diversity rules may have had no or a very small effect on the nature in which POC players’ initial attitudes were portrayed, although this difference, or lack thereof, will be explored more formally in our statistical analysis.

Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(count)` instead.

Statistical Analysis and Model

Using multiple logistic regression, our study will fit a model to the dataset to investigate the existence of a relationship between the log-likelihood of a portrayed positive initial attitude towards the game with the player’s status as a person of color, before the implementation of diversity casting rules (Seasons 36-40) and after the implementation of diversity casting rules (Seasons 41-45). Our study also will investigate whether this relationship is moderated by the accounting of other key factors of season of appearance, age, and gender. Selection of the demographic factors for our model is based upon the sizable interest and proclaimed importance to success in both the physical and social aspects of gameplay in the Survivor cultural sphere, as well as limitations on data availability and study timeline.

The empirical model used to examine the determinants of a player’s initial, portrayed attitude is specified as:

$$\text{logit}\{P(\text{Attitude}_i = 1 | \text{POC}_i, \text{Div_Rules}_i, \text{Age}_i, \text{Gender}_i, \text{Season}_i)\} = \beta_0 + \beta_1 \text{POC}_i + \beta_2 \text{Div_Rules}_i + \beta_3 \text{Age}_i + \beta_4 \text{Gender}_i + \beta_{5-13} \text{Season}_i + \beta_{14} \text{POC}_i * \text{Div_Rules}_i$$

$$\text{Gender}_i = \begin{cases} 1 & \text{if } i \text{ person identifies as a man} \\ 0 & \text{if } i \text{ person identifies as a woman} \end{cases} \quad (1)$$

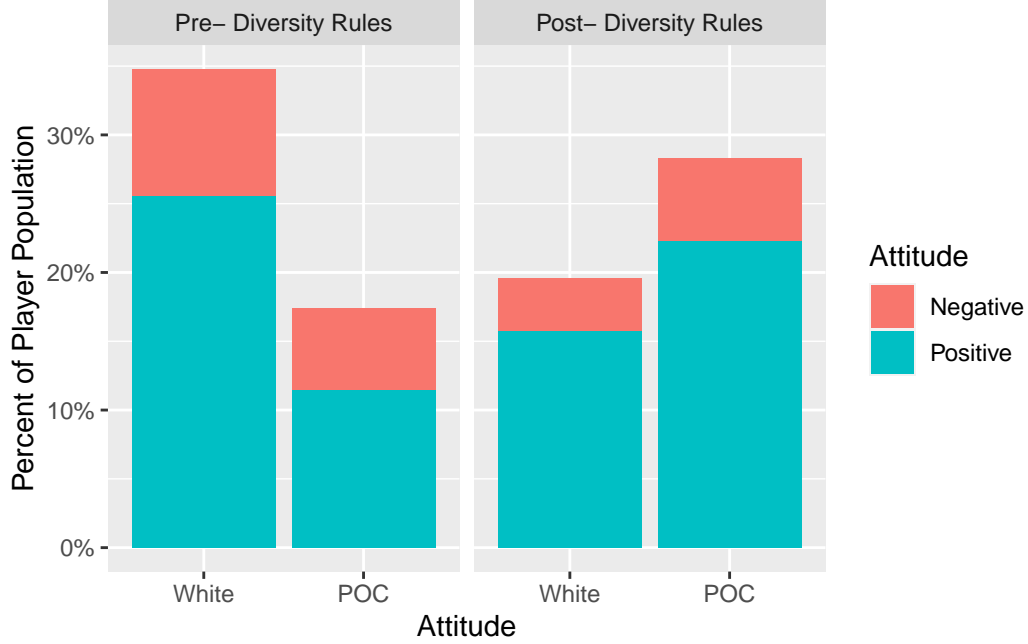


Figure 3: Positive and Negative Portrayed Attitudes of 184 Players from Survivor, Pre- and Post- Implementation of Diversity Casting Rules

where **POC** indicates whether a player is a person of color, **Age** gives a player's age in years with 18 subtracted to account for the minimum age for players, **Gender** indicates whether a player is male or female, and **Div_Rules** indicates the implementation of diversity casting rules on a season, **Season** gives the season number (seasons 37-44, with season 36 and season 45 left out due to singularities), and an interaction term to account for how the presence of the diversity casting rules affected the relationship between a player's POC status and their log odds of a positive, initial portrayed attitude.

We hypothesize that after the implementation of diversity casting rules, the likelihood that a POC player had a positive portrayed initial attitude towards the game meaningfully increased, even after accounting for the age and gender of a player.

Results

We find that for players of the same age and gender playing on the same season pre-diversity rule implementation, the predicted odds of a player's first significant confessional portraying a positive attitude are 34.6% lower for players of color than white players (95% CI: 76.6% lower, 66.1% higher). For players of the same age and gender playing on the same season post-diversity rule implementation, the predicted odds of a player's first significant confessional portraying a positive attitude are 11.3% lower for players of color than white players (95% CI: 70.4% lower,

Table 2: Regression Output of Model Considering Player POC Status, Age, Gender, Season, and Diversity Rules Effect on Log Odds of Initial, Portrayed Attitude of Player

Term	Estimate	Std. Error	z-Statistic	P-Value
Intercept	0.53	0.57	0.93	0.35
POC	-0.47	0.50	-0.95	0.34
Div_Rules	1.40	0.95	1.46	0.14
Age	0.01	0.02	0.26	0.79
Gender	0.32	0.36	0.90	0.37
Season 37	1.21	0.81	1.50	0.13
Season 38	-0.18	0.69	-0.26	0.80
Season 39	0.30	0.71	0.42	0.67
Season 40	0.29	0.74	0.39	0.70
Season 41	-1.50	0.91	-1.65	0.10
Season 42	-0.84	0.94	-0.89	0.37
Season 43	-0.82	0.94	-0.87	0.39
Season 44	-0.05	1.06	-0.04	0.96
POC:Div_Rules	0.35	0.75	0.47	0.64

166% higher). Comparing seasons pre- and post-diversity rule implementation, we find that the difference in the difference in predicted odds ratio for POC versus White players becomes smaller after the diversity rules are implemented, suggesting that the diversity rules have a positive association with the odds that a player of color has a positive, initially portrayed attitude in their first significant confessional.

The results of our logistic regression are tabulated in Table 2.

All of the individual coefficients have p-values greater than the 95% confidence threshold of $p=0.05$. For our predictors of interest in relation to our initial hypotheses—the POC status of the player, the implementation of the diversity rules, as well as a player’s age and gender—, this result indicates that none of these predictors when considered together along with our other predictors controlling for potential season editing clusters were associated with changes in the log odds of an initial, portrayed positive attitude for a player.

Supporting the results of our Wald tests on the individual coefficients, we also find that the overall model is not significant ($p=0.58$), and that an intercept-only model is sufficient to predict the log odds of an initial positive portrayed attitude of a player using a likelihood ratio test. Comparing preliminary models without age and gender predictors to our full model for this analysis, we also found that the addition of these two predictors, both separately ($p = 0.3657$; $p = 0.7924$) and added together ($p = 0.6304$), was insignificant in meaningfully improving model fit to the data. See the Appendix for full results of regression analyses for the reduced models utilizing the age and gender predictors.

add table that shows odds ratios which are much more interpretable than the log odds table that we have

Discussion

Following the cultural reckoning caused by the broad social movement associated with the George Floyd Protests of 2020, many workplaces and businesses responded by implementing diversity, equity, and inclusion (DEI) initiatives. Evaluating the efficacy of these DEI initiatives is an important task and challenge for organizations as they seek to improve workplace equity in the future.

Limitations and Further Research

Conclusion

Appendix

```
survivor_POC <- survivor |> filter(poc == 1)
survivor_White <- survivor |> filter(poc == 0)
num_poc_post <- sum(survivor_POC$div_rules)
num_poc_pre <- nrow(survivor_POC) - num_poc_post
num_white_post <- sum(survivor_White$div_rules)
num_white_pre <- nrow(survivor_White) - num_white_post
```

```
POC_mean_scores <- survivor_POC |> group_by(season) |> summarise(mean = mean(`Compound Score`))
White_mean_scores <- survivor_White |> group_by(season) |> summarise(mean = mean(`Compound Score`))
```

We want the following models:

Model that measures how Attitude is explained by POC status and whether the Div_Rules were there, with the interaction term. This interaction term measures how the presence of div_rules interacts with the relationship between being poc and having a positive attitude.

Model that measures how Attitude is explained by POC status and whether the Div_Rules were there, with the interaction term, as well as age. This measures how the presence of div_rules interacts with the relationship between being poc and having a positive attitude after accounting for age.

Model that measures how Attitude is explained by POC status and whether the Div_Rules were there, with the interaction term, as well as age and gender. This measures how the presence of div_rules interacts with the relationship between being poc and having a positive attitude after accounting for age and gender.

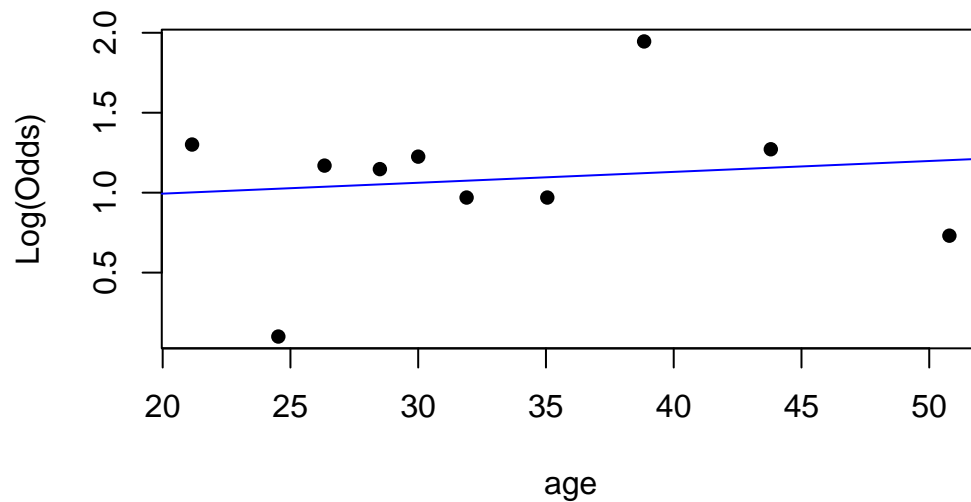
Do we want a model with race? We are not interested in breaking down by race, only POC status, but it seems sensible to have a model that measures how Attitude is explained by race status and whether the Div_Rules were there, with the interaction term, as well as age and gender. This measures how the presence of div_rules interacts with the relationship between being different races.

Summary Stats

Demographics

Assumptions

```
emplogitplot1(Attitude ~ age, data = survivor, ngroups = 10)
```



For age the linear assumption is violated, but this is the best we can do with a linear approximation.

Table 3: Demographic Characteristics of the Players in the Study Population

Characteristic	Value
Gender	
Male - no. (%)	91 (49.46)
Female - no. (%)	93 (50.54)
Age - yr.*	
All	32.80 \pm 8.95
Race - no (%)**	
Asian/Pacific Islander	28 (15.22)
Black	39 (21.11)
Hispanic or Latino	17 (9.24)
White	100 (54.35)
Season - no. (%)	
36	20 (10.87)
37	20 (10.87)
38	18 (9.78)
39	20 (10.87)
40	18 (9.78)
41	17 (9.24)
42	18 (9.78)
43	18 (9.78)
44	17 (9.24)
45	18 (9.78)

Note:

* Mean age of players \pm Std. Deviation, ** Race as provided by the Survivor Wiki