

Extension: Convention Text Analysis – Tokenization Virtue and Vice Comparisons

Jennifer Lin

4/21/2020

Contents

Introduction	1
Comparing Virtue and Vice	2
Grahping the Data	10

Introduction

The Moral Foundations Dictionary captures the foundations in the Moral Foundations Theory under positive (virtue) and negative (vice) categories. The goal is to allow researchers to characterize the type of appeal as it appears in text. As an extension to the Convention Text Analyses, I conduct analyses to compare the virtue and vice appeals for the convention speakers.

The convention speeches that are used in this analysis can be found here: <https://github.com/lin-jennifer/2016NCtranscripts>

The processed data can be found here: <https://lin-jennifer.github.io/MeasuringMorality/data.html>

For the purpose of this walkthrough, I will use the processed data.

I load the data as follows:

```
# Load data
speech <- read.csv("~/Desktop/Working/Moral-Psychology/SpeechAnalysis/quanteda/Composite
  header = TRUE)
```

Comparing Virtue and Vice

Before making comparisons, I filter out the speeches to exclude Invocations and Benedictions.

```
table(speech$Type)

##
## benediction      speech      video
##           7          208          1

speech <- speech[!(speech$Type == "benediction"), ]
```

I also load some packages that I need for this section.

```
library(psych)
library(effsize)

##
## Attaching package: 'effsize'

## The following object is masked from 'package:psych':
##
##      cohen.d
```

Next, I compare a series of t-tests to compare appeals between convention speakers

```
# Harm Virtue
describeBy(speech$care.virtue, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1     1 145 5.97 7.72      4    4.39 2.97   0 49   49 3.25    12.87 0.64
## -----
## group: RNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis se
## X1     1  64 6.42 7.97      4     5 4.45   0 51   51 3.25    14.01 1

t.test(speech$care.virtue ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$care.virtue by speech$Convention
## t = -0.38526, df = 117.15, p-value = 0.7007
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.802246  1.889530
```

```
## sample estimates:
## mean in group DNC mean in group RNC
##          5.965517          6.421875

cohen.d(speech$care.virtue, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.05854796 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.3544670  0.2373711

# Harm Vice
describeBy(speech$care.vice, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean    sd median trimmed  mad min max range skew kurtosis   se
## X1      1 145 3.07 3.93      2    2.26 1.48   0 22    22 2.61      7.96 0.33
## -----
## group: RNC
##   vars   n mean    sd median trimmed  mad min max range skew kurtosis   se
## X1      1  64 4.28 6.68      2    3.02 1.48   0 42    42 3.84      17 0.84

t.test(speech$care.vice ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$care.vice by speech$Convention
## t = -1.352, df = 82.836, p-value = 0.1801
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.9958002  0.5712313
## sample estimates:
## mean in group DNC mean in group RNC
##          3.068966          4.281250

cohen.d(speech$care.vice, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.2458229 (small)
## 95 percent confidence interval:
```

```
##          lower          upper
## -0.54263617  0.05099027

# Fairness Virtue
describeBy(speech$fairness.virtue, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 145 1.74 2.64      1    1.17 1.48   0 16   16 2.72      9.1 0.22
## -----
## group: RNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 64 2.48 4.06      1    1.62 1.48   0 25   25 3.21     13.02 0.51

t.test(speech$fairness.virtue ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$fairness.virtue by speech$Convention
## t = -1.3507, df = 87.35, p-value = 0.1803
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.8448428  0.3519548
## sample estimates:
## mean in group DNC mean in group RNC
##          1.737931          2.484375

cohen.d(speech$fairness.virtue, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.2377741 (small)
## 95 percent confidence interval:
##          lower          upper
## -0.53452635  0.05897814

# Fairness Vice
describeBy(speech$fairness.vice, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1      1 145  0.3 0.88      0    0.11  0   0 8    8 5.4     39.86 0.07
```

```
## -----
## group: RNC
##      vars  n mean  sd median trimmed mad min max range skew kurtosis  se
## X1      1 64 0.36 0.9      0   0.17  0  0  6      6 4.18    21.98 0.11

t.test(speech$fairness.vice ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$fairness.vice by speech$Convention
## t = -0.41822, df = 118.08, p-value = 0.6765
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.3207354  0.2088819
## sample estimates:
## mean in group DNC mean in group RNC
##      0.3034483      0.3593750

cohen.d(speech$fairness.vice, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.06333972 (negligible)
## 95 percent confidence interval:
##      lower      upper
## -0.3592679  0.2325885

# Ingroup Virtue
describeBy(speech$loyalty.virtue, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## X1      1 145 3.33 3.99      2   2.64 1.48  0 31    31  3.2    15.93 0.33
## -----
## group: RNC
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## X1      1 64 5.28 5.83      4   4.15 2.97  0 27    27 2.18    4.63 0.73

t.test(speech$loyalty.virtue ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$loyalty.virtue by speech$Convention
```

```

## t = -2.4348, df = 89.958, p-value = 0.01687
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.5415040 -0.3589271
## sample estimates:
## mean in group DNC mean in group RNC
##      3.331034      5.281250
cohen.d(speech$loyalty.virtue, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.4214214 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.7200643 -0.1227785
# Ingroup Vice
describeBy(speech$loyalty.vice, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
## X1      1 145 0.01 0.12      0      0  0  0  1      1 8.25     66.54 0.01
## -----
## group: RNC
##   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
## X1      1  64 0.06 0.3      0      0  0  0  2      2 5.05     26.33 0.04
t.test(speech$loyalty.vice ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$loyalty.vice by speech$Convention
## t = -1.2491, df = 71.485, p-value = 0.2157
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.12645018  0.02903638
## sample estimates:
## mean in group DNC mean in group RNC
##      0.0137931      0.0625000
cohen.d(speech$loyalty.vice, speech$Convention)

##

```

```
## Cohen's d
##
## d estimate: -0.2521733 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.54903605  0.04468946

# Authority Virtue
describeBy(speech$authority.virtue, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis  se
## X1      1 145 3.17 3.63      2    2.52 1.48   0  23    23 2.48      8.31 0.3
## -----
## group: RNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis  se
## X1      1  64 8.81 8.52      6    7.27 5.93   0  43    43 1.98      4.25 1.07

t.test(speech$authority.virtue ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$authority.virtue by speech$Convention
## t = -5.0954, df = 73.288, p-value = 2.618e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.845964 -3.434208
## sample estimates:
## mean in group DNC mean in group RNC
##      3.172414      8.812500

cohen.d(speech$authority.virtue, speech$Convention)

##
## Cohen's d
##
## d estimate: -1.008798 (large)
## 95 percent confidence interval:
##      lower      upper
## -1.3202442 -0.6973509

# Authority Vice
describeBy(speech$authority.vice, speech$Convention)

##
```

```
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
## X1     1 145 0.03 0.16      0      0  0  0  1     1 5.71    30.81 0.01
## -----
## group: RNC
##   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
## X1     1  64 0.16 0.54      0      0  0  0  3     3 3.66    13.43 0.07

t.test(speech$authority.vice ~ speech$Convention)

##
## Welch Two Sample t-test
##
## data:  speech$authority.vice by speech$Convention
## t = -1.8651, df = 68.188, p-value = 0.06648
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.266316494  0.008988907
## sample estimates:
## mean in group DNC mean in group RNC
##      0.02758621      0.15625000

cohen.d(speech$authority.vice, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.3917643 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.69003146 -0.09349708

# Purity Virtue
describeBy(speech$sanctity.virtue, speech$Convention)

##
## Descriptive statistics by group
## group: DNC
##   vars   n mean   sd median trimmed mad min max range skew kurtosis   se
## X1     1 145 1.53 2.59      0    0.96  0  0 16     16 2.59     8.28 0.22
## -----
## group: RNC
##   vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1     1  64  3.5 3.43    2.5      3 2.22  0 16     16 1.35     1.65 0.43
```



```
t.test(speech$sanctity.virtue ~ speech$Convention)
```

```
##
## Welch Two Sample t-test
##
## data: speech$sanctity.virtue by speech$Convention
## t = -4.1065, df = 96.167, p-value = 8.435e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.920697 -1.017234
## sample estimates:
## mean in group DNC mean in group RNC
## 1.531034 3.500000
```

```
cohen.d(speech$sanctity.virtue, speech$Convention)
```

```
##
## Cohen's d
##
## d estimate: -0.6854335 (medium)
## 95 percent confidence interval:
## lower upper
## -0.9885915 -0.3822754
```

```
# Purity Vice
```

```
describeBy(speech$sanctity.vice, speech$Convention)
```

```
##
## Descriptive statistics by group
## group: DNC
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 145 0.35 0.8 0 0.16 0 0 5 5 3 10.79 0.07
## -----
## group: RNC
## vars n mean sd median trimmed mad min max range skew kurtosis se
## X1 1 64 0.56 0.92 0 0.37 0 0 4 4 1.78 2.74 0.12
```

```
t.test(speech$sanctity.vice ~ speech$Convention)
```

```
##
## Welch Two Sample t-test
##
## data: speech$sanctity.vice by speech$Convention
## t = -1.5846, df = 106, p-value = 0.116
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.47448412 0.05293239
```

```
## sample estimates:
## mean in group DNC mean in group RNC
##          0.3517241          0.5625000
cohen.d(speech$sanctity.vice, speech$Convention)

##
## Cohen's d
##
## d estimate: -0.2520018 (small)
## 95 percent confidence interval:
##      lower      upper
## -0.54886321  0.04485959
```

Grahping the Data

I also create graphs that display the rates that the convention speakers appeal to the virtue and vice aspects of each foundation.

I load some packages needed for this step

```
# Load packages
library(car)

## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:psych':
##
##      logit

library(dplyr)

##
## Attaching package: 'dplyr'
## The following object is masked from 'package:car':
##
##      recode
## The following objects are masked from 'package:stats':
##
##      filter, lag
## The following objects are masked from 'package:base':
##
```

```
##      intersect, setdiff, setequal, union
library(psych)
library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##      %+%, alpha

library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

##
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':
##
##      nasa

library("ggpubr")

## Loading required package: magrittr

library("reshape2")
library(scales)

##
## Attaching package: 'scales'

## The following objects are masked from 'package:psych':
##
##      alpha, rescale

I generate summary statistics for each category for plotting.
# Summary Statistice by convention

HarmVirtue <- speech %>% group_by(Convention) %>% summarize(mean = mean(care.virtue,
  na.rm = TRUE), sd = sd(care.virtue, na.rm = TRUE), n = n(),
  se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(care.virtue),
  min = min(care.virtue), med = median(care.virtue)) %>% mutate(type = "Harm",
  cat = "Virtue")

HarmVice <- speech %>% group_by(Convention) %>% summarize(mean = mean(care.vice,
  na.rm = TRUE), sd = sd(care.vice, na.rm = TRUE), n = n(),
```

```

se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(care.vice),
min = min(care.vice), med = median(care.vice)) %>% mutate(type = "Harm",
cat = "Vice")

FairnessVirtue <- speech %>% group_by(Convention) %>% summarize(mean = mean(fairness.virtue,
na.rm = TRUE), sd = sd(fairness.virtue, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(fairness.virtue),
min = min(fairness.virtue), med = median(fairness.virtue)) %>%
mutate(type = "Fairness", cat = "Virtue")

FairnessVice <- speech %>% group_by(Convention) %>% summarize(mean = mean(fairness.vice,
na.rm = TRUE), sd = sd(fairness.vice, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(fairness.vice),
min = min(fairness.vice), med = median(fairness.vice)) %>%
mutate(type = "Fairness", cat = "Vice")

IngroupVirtue <- speech %>% group_by(Convention) %>% summarize(mean = mean(loyalty.virtue,
na.rm = TRUE), sd = sd(loyalty.virtue, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(loyalty.virtue),
min = min(loyalty.virtue), med = median(loyalty.virtue)) %>%
mutate(type = "Ingroup", cat = "Virtue")

IngroupVice <- speech %>% group_by(Convention) %>% summarize(mean = mean(loyalty.vice,
na.rm = TRUE), sd = sd(loyalty.vice, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(loyalty.vice),
min = min(loyalty.vice), med = median(loyalty.vice)) %>%
mutate(type = "Ingroup", cat = "Vice")

AuthorityVirtue <- speech %>% group_by(Convention) %>% summarize(mean = mean(authority.virtue,
na.rm = TRUE), sd = sd(authority.virtue, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(authority.virtue),
min = min(authority.virtue), med = median(authority.virtue)) %>%
mutate(type = "Authority", cat = "Virtue")

AuthorityVice <- speech %>% group_by(Convention) %>% summarize(mean = mean(authority.vice,
na.rm = TRUE), sd = sd(authority.vice, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(authority.vice),
min = min(authority.vice), med = median(authority.vice)) %>%
mutate(type = "Authority", cat = "Vice")

PurityVirtue <- speech %>% group_by(Convention) %>% summarize(mean = mean(sanctity.virtue,
na.rm = TRUE), sd = sd(sanctity.virtue, na.rm = TRUE), n = n(),
se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(sanctity.virtue),
min = min(sanctity.virtue), med = median(sanctity.virtue)) %>%

```

```

mutate(type = "Purity", cat = "Virtue")

PurityVice <- speech %>% group_by(Convention) %>% summarize(mean = mean(sanctity.vice,
  na.rm = TRUE), sd = sd(sanctity.vice, na.rm = TRUE), n = n(),
  se = sd/sqrt(n), ci = qt(0.975, df = n - 1) * se, max = max(sanctity.vice),
  min = min(sanctity.vice), med = median(sanctity.vice)) %>%
  mutate(type = "Purity", cat = "Vice")

# Combine each of the outputs
token <- rbind(HarmVirtue, HarmVice, FairnessVirtue, FairnessVice,
  IngroupVirtue, IngroupVice, AuthorityVirtue, AuthorityVice,
  PurityVirtue, PurityVice)

# Organize label order for the foundations
token$type <- factor(token$type, levels = c("Harm", "Fairness",
  "Ingroup", "Authority", "Purity"))
token$cat <- factor(token$cat, levels = c("Virtue", "Vice"))

```

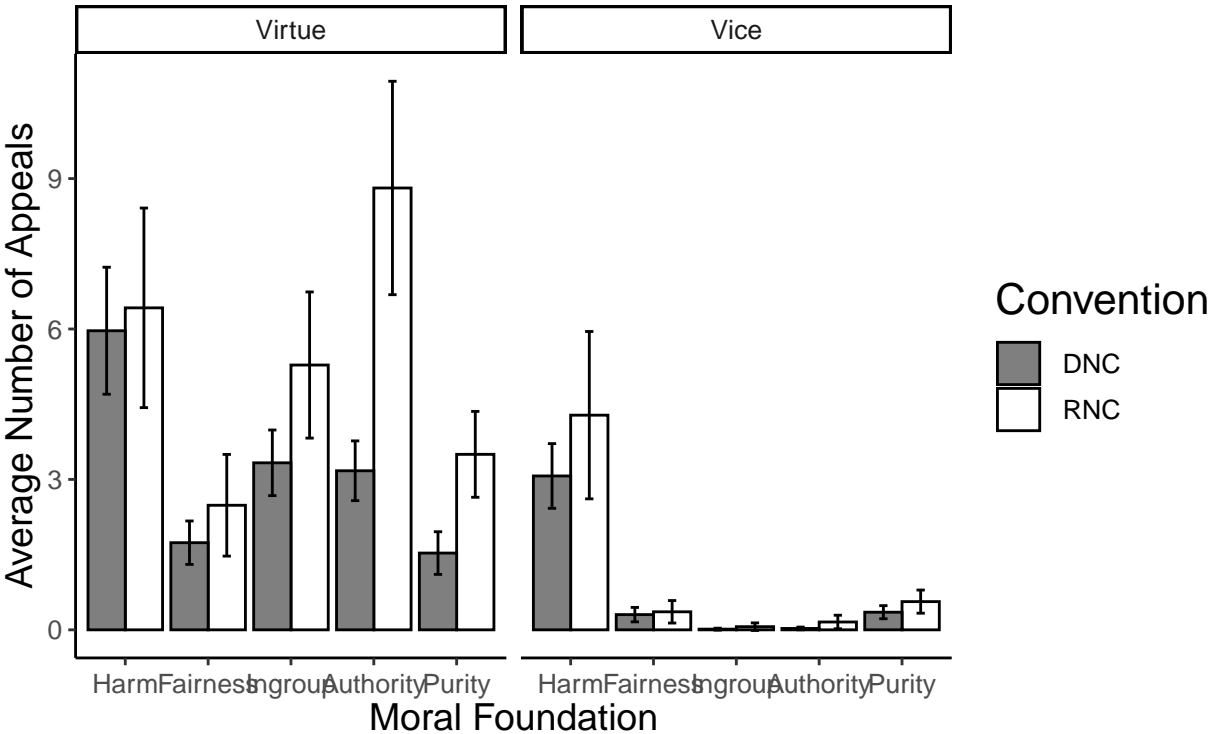
Then I graph

```

# Create Graph with foundation on x axis and fill by
# convention type
ggplot(token, aes(x = type, y = mean, fill = Convention)) + geom_bar(stat = "identity",
  position = position_dodge(), color = "black") + geom_errorbar(aes(ymin = mean -
  ci, ymax = mean + ci), width = 0.2, position = position_dodge(0.9)) +
  ggtitle("Moral Appeals in Political Speeches") + theme_classic() +
  xlab("Moral Foundation") + ylab("Average Number of Appeals") +
  labs(caption = "Source: 2016 RNC and DNC") + facet_wrap(~cat) +
  theme(text = element_text(size = 12, colour = "black"), axis.title = element_text(s
  colour = "black"), title = element_text(size = 16, colour = "black"),
  plot.caption = element_text(size = 10, color = "black"),
  axis.text.x = element_text(angle = 0, hjust = 0.5, vjust = 0.5),
  plot.title = element_text(hjust = 0.5)) + scale_x_discrete(labels = wrap_format
  scale_fill_manual("Convention", values = c(DNC = "grey50",
  RNC = "white"))

```

Moral Appeals in Political Speeches



Source: 2016 RNC and DNC