Gender coding reliability

Introduction

To establish the reliability of gender coding, a sample of characters was coded by a secondary coder. For each game, 10 characters were randomly chosen with the probability of being chosen being in proportion to the amount of dialogue they spoke. No characters were repeated within series. See the file `getDataForCodingReliability.R` for specific details. This resulted in 444 characters that were second-coded (some games had fewer than 10 characters, or used game data to define the gender of characters rather than human coding). This represents about 3% of all characters in the corpus.

A secondary coder, who was blind to the original coding, coded each of the characters in this sample. To measure agreement, we use Cohen's kappa, which is designed for measuring inter-rater reliability while taking in the agreement that would come about by chance (e.g. because the proportions of cases in different groups are unbalanced).

Load libraries:

```
library(rjson)
library(psych)
```

Load data:

```
d = read.csv("reliabilityCoding_S.csv",stringsAsFactors = F)
d = d[!is.na(d$gender),]
d$gender.Orig = NA

for(folder in unique(d$folder)){
  js = fromJSON(file = paste0("../",folder,"meta.json"))
  g = js$characterGroups
  g2 = rep(names(js$characterGroups),sapply(js$characterGroups,length))
  names(g2) = unlist(js$characterGroups)

  d[d$folder==folder,]$gender.Orig = g2[d[d$folder==folder,]$charName]
}
# Some data changed after reliability review

d[d$game=="Final Fantasy VII" & d$charName=="Waitress",]$gender.Orig = "female"
d[d$game=="Final Fantasy VII" & d$charName=="Barkeeper",]$gender.Orig = "male"
d[d$game=="King's Quest III: To Heir Is Human" & d$charName=="Squirrel",]$gender.Orig = "neutral"
d[d$game=="King's Quest III: To Heir Is Human" & d$charName=="Fish",]$gender.Orig = "neutral"
d[d$game=="King's Quest III: To Heir Is Human" & d$charName=="Bird",]$gender.Orig = "neutral"
d[d$game=="King's Quest III: To Heir Is Human" & d$charName=="Mouse",]$gender.Orig = "neutral"
d[d$game=="Persona 4" & d$charName=="Kimono-Clad Woman",]$gender.Orig = "female"
d[d$game=="Final Fantasy VII Remake" & d$charName=="Elmyra",]$gender.Orig = "male"
d[d$game=="King's Quest VIII" & d$charName=="Gryphs",]$gender.Orig = "neutral"

d$gender[d$gender %in% c("neutral", "not coded","not gendered")] = "neutral"
```

Total coding

The agreement matrix for 1st and 2nd coding, including all categories coded:

```
table(paste0("2nd coding:",d$gender),paste0("1st coding:",d$gender.Orig))
```

```
##
```

```
##                       1st coding:female 1st coding:male 1st coding:NA
##   2nd coding:female                 114               2             2
##   2nd coding:male                     1             291             0
##   2nd coding:neutral                  0               7             1
##
##                       1st coding:neutral
##   2nd coding:female                    0
##   2nd coding:male                      6
##   2nd coding:neutral                  20
```

Coders agreed for 96.37% of cases.

Formal measure of agreement:

```
cohen.kappa(x=cbind(d$gender,d$gender.Orig))
```

```
## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
##                  lower estimate upper
## unweighted kappa  0.89     0.92  0.96
## weighted kappa    0.94     0.94  0.94
##
##  Number of subjects = 441
```

According to Landis & Koch (1977), this represents "almost perfect" agreement. There was disagreement for NA characters which is explored below.

Coding gendered versus neutral

One hurdle in coding is finding information on the characters, particularly generic characters. Below is a comparison of whether the two coders agreed that there was enough information to make a gendered coding or not:

```
d$codable = d$gender!="neutral"
d$codable.Orig = !d$gender.Orig %in% c("neutral","unknown","not coded", "cut")

table(paste0("2nd coding:", d$codable),paste0("1st coding:",d$codable.Orig))
```

```
##
##                     1st coding:FALSE 1st coding:TRUE
##   2nd coding:FALSE                20               8
##   2nd coding:TRUE                  6             410
```

```
cohen.kappa(x=cbind(d$codable, d$codable.Orig))
```

```
## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
##                  lower estimate upper
## unweighted kappa  0.59     0.72  0.86
## weighted kappa    0.59     0.72  0.86
##
##  Number of subjects = 444
```

There were 7 cases where the 1st coding suggested the characters were neutral while the 2nd coding gave them genders:

```
knitr::kable(d[!d$codable.Orig & d$codable,c("game","charName","Comments")],row.names = F)
```

game	charName	Comments
Final Fantasy V	Moogle	Lenna calls the moogle “him” and “he”, though Butz uses “it”
Final Fantasy XII	Cloudborne Resident (tan bangaa in green vest)	https://youtu.be/h8yPPqyqQdw?t=404
Final Fantasy XII	Rabanastran (blindfolded tan bangaa SG)	https://jegged.com/Games/Final-Fantasy-XII/Side-Quests/Viera-Matchmaking.html
King’s Quest III: To Heir Is Human	Cat	Called “he”
King’s Quest VIII	Gryphs	Male voice acting
The Secret of Monkey Island	Head	Not many gender cues. From lore from other games, it seems this is Sepp the Navigator. In the special edition it’s voiced by a male actor, https://youtu.be/vQReJFaFHHA?t=872

The Head from Monkey Island can be inferred as male, but only from content outside the original game. Therefore, it’s kept as neutral. There are two disagreements about two species in Final Fantasy XII, so these are kept neutral to be consistent with the rest of the coding for that game.

The other cases have positive evidence for gender signifiers which were uncovered in the 2nd coding, so were changed.

There were 7 cases where the 2nd coding suggested the characters were neutral while the 1st coding gave them genders:

```
knitr::kable(d[d$codable.Orig & !d$codable,c("game","charName","Comments")],row.names = F)
```

game	charName	Comments
Final Fantasy II	Man in Armor	Difficult to code https://youtu.be/dvRSK2cGHyc?t=1038, https://youtu.be/ocA3cHTibWs?t=3000
Final Fantasy II	Wandering Soldier	
Final Fantasy V	Soldier	Not much to go on, https://youtu.be/HAVPyGVQY0w?t=153
Kingdom Hearts III	Everyone	
Mass Effect 3	Geth Prime	Male voice, not many gender cues. https://youtu.be/IG2yry7-1zU?t=34
The Secret of Monkey Island	Guy 1	We see their legs (boots, trousers), easy to assume male but not much direct evidence. https://youtu.be/wsyV_lwYDVI?t=135
Star Wars: Knights of the Old Republic	Phirk	Aqualish, voice is in other language
Star Wars: Knights of the Old Republic	Nubassa	According to wiki lore, females look different, though there are few gender sigifiers

The coding for these were kept as they were.

Coding male versus female

Below is a table comparing only characters where both coders chose either “male” or “female”.

```
maleOrFemale = d$gender %in% c("male","female") & d$gender.Orig %in% c("male","female")

table(paste0("2nd coding:",d[maleOrFemale,]$gender),
      paste0("1st coding:",d[maleOrFemale,]$gender.Orig))
```

```
##
##                     1st coding:female 1st coding:male
##   2nd coding:female               114               2
##   2nd coding:male                   1             291
```

```
cohen.kappa(x=cbind(d[maleOrFemale,]$gender,d[maleOrFemale,]$gender.Orig))
```

```
## Warning in cohen.kappa1(x, w = w, n.obs = n.obs, alpha = alpha, levels =
## levels): upper or lower confidence interval exceed abs(1) and set to +/- 1.
```

```
## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
##                  lower estimate upper
## unweighted kappa  0.96     0.98     1
## weighted kappa    0.96     0.98     1
##
##  Number of subjects = 408
```

There were two cases where the initial coding was male and the 2nd coding was female. The first is the "Presidential Aide" in Final Fantasy VIII. They wear generic clothing, and so the initial coding is likely correct. The second is for "Elmyra" from Final Fantasy VII Remake, who is Aerith's mother and clearly female, so this is a case of an error in the initial coding.

There was one case where the initial coding was female and the 2nd coding was male. This is for the "Steward" in Final Fantasy XII. The second coder had identified a similar but different character in a play-through video.

Conclusion

The agreement was very high. The majority of disagreements involved whether a character was gendered or neutral, rather than whether they were male or female. Only one case of clear incorrect coding was identified.

References

Landis, J.R.; Koch, G.G. (1977). "The measurement of observer agreement for categorical data". Biometrics. 33 (1): 159–174.