Gender bias in Video Game dialogue: Supplementary materials

Stephanie Rennick, Seán G. Roberts, Melanie Clinton, E. I., Liana Oh, Charlotte Clooney, Edward Healy

July 15, 2022

Contents

1 Introduction

This set of supplementary materials describes each step of the creation and analysis of the Video Games Dialogue Corpus. Parts of it are written in Rmarkdown, which includes R code that implements an analysis, the output of that code, and plain commentary and explanations. This supporting materials document itself is compiled from the output of several documents. Each individual document and the original data and R code are available in the accompanying online repository.

The rest of this introduction gives a broad overview of the methods.

Sample

A sample of 50 video games was selected in the Role-Playing Game (RPG) genre where dialogue was a major game mechanic. The sample was chosen to be representative of several characteristics such as: publication date (balanced from 1986 to 2020), style ('Western' vs. 'Japanese' RPGs), and target audience (rated for 'Everybody', 'Teen' and 'Adult' by the Entertainment Software Rating Board). The games included ones developed by large companies (e.g. franchises such as *Final Fantasy, Persona, Mass Effect, Dragon Age, The Elder Scrolls, Kingdom Hearts*) and from smaller developers (e.g. *Monkey Island, Stardew Valley*). All games either individually sold, or belong to series that sold, at least 1 million copies worldwide. Dialogue for each game was located from a range of sources including data directly from the game code and public websites such as wikis and fan-made transcripts. See SI 1.2 for details.

Script parsing

For each game script, a custom python program was written which scraped and parsed the script into a common format. This parser used systematic pattern recognition, but also applied specific manual edits listed in the metadata files. There were approximately 20,000 manual edits applied to the corpus, mostly fixing mappings between character names and lines of dialogue. The scraping and parsing programs are available in an online repository alongside programs for calculating all the measures and statistics presented in this paper (`https://osf.io/b2qcg/?view_only=c194016e73544b60b57bccff453dd93a`).

The game script format represented lines of dialogue paired with the name of the character who spoke them, as well as actions and changes in location. The format used a recursive JSON structure in order to represent dialogue trees common in games.

The game scripts were validated with a systematic error-checking procedure (see SI 1.2). Transcription errors in the source were identified by finding a video of the game being played, choosing random dialogue in the video, and checking that this dialogue appears accurately in the corpus. Parsing errors from the automatic parsers were identified by manually checking random lines of dialogue. For each game, 15 lines were checked for transcription errors and 5 lines were checked for parsing errors. Any errors were raised as issues on the GitHub repository, and fixed. After this, a second round of error checking and fixing was conducted following the same steps as above.

Gender Coding

Conferred gender of characters was coded manually according to a set of defeasible indicators, as discussed above and in more detail in SI 1.3. The coding scheme did not assume binary gender. Evidence for edge cases is documented in the corpus repository. Where there was insufficient evidence for a character's gender, they were labelled as "neutral" (around 7.6% of characters).

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. Agreement between coders was 'almost perfect' (Landis & Koch, 1977, raw agreement = 96%, Cohen's kappa = 0.92 [0.89, 0.96], see SI 1.15).

Measures

The measures of dialogue length and readability were obtained using the python module *textatistic* (https://pypi.org/project/textatistic/) that was designed for looking at gender differences in large text corpora (Hengel, 2020). Length of dialogue was measured in number of words, number of lines, number of sentences and number of syllables. All of these measures were correlated with each other with $r > 0.98$ (measured at the group level), indicating that the measures are robust.

Several of the games in the sample were originally written in Japanese, so there may be differences in estimates of female dialogue in the original script versus the English translation. To test this, we analysed three versions of *Chrono Trigger*: the original Japanese script and two English translations. The measures of dialogue length were highly similar between all texts (correlation between number of English words and Japanese characters per line r = 0.93) and all estimates of the proportion of female dialogue are within 0.7

percentage points of each other (see SI 1.13). This suggests that the general gender biases are not caused by translation.

Statistical methods

The aim of the statistical measures is to assess whether there is a statistically significant bias in the distribution of dialogue by gender within a game. Standard parametric tests are inappropriate because the data is highly non-independent (words belong to lines, and lines belong to coherent characters) and highly skewed (a small number of characters say a lot while a large number of characters say little). Instead, a permutation framework was used that compares an observed measure with the range of measures that would be expected if there really was no bias. This was done as follows (for more details, see SI 1.4).

The proportion of dialogue by gender is assessed in comparison to two baselines which reflect two possible sources of bias. The first source is a 'character bias' where more male characters are included in the game than female characters, which has a knock-on effect on the proportion of dialogue for each gender. A hypothetical script was generated where the mapping between gender categories and individuals is randomly determined, with each character having an independent and equal probability of being male or female. That is, the link between gender and characters is randomised to remove any potential bias. Generating 100,000 scripts created a distribution of probable values for the proportion of female dialogue if there was no bias. This was compared to the true proportion of female dialogue. This produced a z-score that represents the strength of the bias (in number of standard deviations away from the expected mean), and a p-value that represents how likely the baseline process results in a measure that is more extreme than the observed distribution. Lower p-values indicate that the baseline model assumptions are unlikely to hold, and suggests that the imbalance in dialogue is due to the imbalance in the proportion of characters.

Similarly, the second possible source of bias is the 'dialogue bias' where the average male character is given more dialogue than the average female character. To model a baseline to test this, a hypothetical script is generated where the mapping between gender and characters remains unchanged, but the mapping between characters and lines is randomly permuted (all of one character's lines can be swapped for all of another character's lines). This maintains the same proportion of female characters as the real data, but the proportion of dialogue for female characters will vary if they are given systematically less to say than male characters. 100,000 scripts were generated to create a distribution of values, to which the true proportion of female dialogue can be compared.

References

Hengel, E. (2020) Publishing while female, CEPR Press.

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

2 Corpus Design

Corpus design

Introduction

This document describes the design of the corpus, including the sample of games used, the software pipeline, the format of the data, and the error checking procedures.

Corpus Design

50 games were selected for the corpus. All games were Role Playing Games (RPGs) with central mechanics for dialogue. They all have high sales figures: All games either individually sold, or belong to series that sold, at least 1 million copies worldwide. Every game or series frequently features in lists of the top RPGs of all time. For example, IGN's top 100 RPGs of all time and Game Informer's top 100 RPGs of all time contain all the games in the corpus (or at least one game from each series). The exception is the King's Quest and Monkey Island games, which sold relatively poorly on release (compared to modern games) and are sometimes thought of as being "adventure games" or "point-and-click" games, though they fit the definition of RPGs. However, both of these series are frequently discussed as being highly influential on the medium (e.g. https://latinamericanpost.com/35312-monkey-island-one-of-the-most-important-video-games-in-history, https://gamerant.com/monkey-island-games-development-history-evolution-30-years/, https://vocal.media/geeks/legacy-of-king-s-quest-in-modern-gaming, and https://videogamesuncovered.com/features/a-history-of-kings-quest/). The final requirement was that the games needed to have an accessible source of dialogue available.

We aimed to collect a balance of games according to three factors. The first was RPG style, including Western RPGs and Japanese-style RPGs, since those are the dominant types. Secondly, the target audience age. For this latter, we used three categories based on the the official ESRB ratings: "Child" (ESRB ratings "Everyone", "Everyone 10+"), "Teen" (ESRB rating "Teen"), and "Adult" (ESRB ratings "Mature 17+" and "Adults Only 18+"). The table below shows the distribution of game series with the number of games in brackets.

Rating	Western	JRPG	Totals
Child	Stardew Valley, KQ, Monkey Island (13)	Super Mario RPG, KH (5)	18
Teen	Horizon Zero Dawn, Start Wars: KOTOR (2)	Chrono-Trigger, FF (19)	20
Adult	Mass Effect, Elder Scrolls, Dragon Age (9)	Persona (3)	12
Totals	24	26	50

KQ = King's Quest; KH = Kingdom Hearts, FF = Final Fantasy.

Finally, we also aimed for a balance of games across time between 1985 and 2020:

```
yBoundaries = c(1984,1989,1994,1999,2004,2009,2014,2021)
allGames$yearCat = cut(allGames$year,c(1984,1989,1994,1999,2004,2009,2014,2021),
                       include.lowest = T,
                       labels = paste(yBoundaries[1:7]+1,yBoundaries[2:8],sep=" - "))

x = tapply(allGames$shortName,allGames$yearCat,paste,collapse="; ")
n = tapply(allGames$shortName,allGames$yearCat,length)
xx = data.frame(Years = names(x),games = x, Num = n)
xx$Years = gsub(","," - ",xx$Years)
xx$Years = gsub("\\(","",xx$Years)
xx$Years = gsub("\\]","",xx$Years)
xx$Years = gsub("\\[","",xx$Years)
```

```r
knitr::kable(xx,row.names = F)
```

Years	games	Num
1985 - 1989	FFI; FFII; KQ3; KQ1; KQ2; KQ4	6
1990 - 1994	FFIV; FFV; FFVI; KQ5; KQ6; KQ7; MI1; MI2	8
1995 - 1999	ChronoTrigger; FFVII; FFVIII; KQ8; MI3; SMario; Daggerfall	7
2000 - 2004	FFIX; FFX; FFX2; KH; KOTOR; Morrowind	6
2005 - 2009	DAO; FFXII; FFXIII; KH2; ME1; Persona3; Persona4; Oblivion	8
2010 - 2014	DragonAge2; FFXIV; FFXIII-2; FFXIII-LR; KH3D; ME2; ME3; Skyrim	8
2015 - 2021	FFXV; FFVII-R; HorizonZeroDawn; KQChapters; KH3; Persona5; StardewValley	7

Given the difficulties of balancing these three factors, and the constraint on being able to find dialogue sources, the corpus is relatively well balanced, at least at the game level. However, the corpus is not balanced in terms of number of words per game. There is a difference of two orders of magnitude between the game with most dialogue and the game with least dialogue. This is mainly an effect of the change in technology, storage capacity, and development budget for video games over the last three decades.

```r
gs = read.csv("../results/generalStats.csv",stringsAsFactors = F)
gs = gs[gs$alternativeMeasure=="False",]
gs = gs[gs$group=="TOTAL",]
knitr::kable(gs[order(gs$words),c("game","words")],row.names = F)
```

game	words
King's Quest II: Romancing the Throne	402
King's Quest I: Quest for the Crown	1396
King's Quest IV: The Perils of Rosella	1772
Final Fantasy	2763
King's Quest III: To Heir Is Human	3151
Final Fantasy II	8689
Monkey Island 2: LeChuck's Revenge	9285
King's Quest V	10833
Final Fantasy V	12408
Final Fantasy XIII	12674
Super Mario RPG: Legend of the Seven Stars	13755
The Secret of Monkey Island	14619
Final Fantasy VI	14750
King's Quest VIII	15880
Final Fantasy IV	17822
Kingdom Hearts 3D: Dream Drop Distance	18484
Kingdom Hearts	19847
King's Quest VII: The Princeless Bride	21358
The Curse of Monkey Island	30577
Final Fantasy X-2	31635
Chrono Trigger	37982
Kingdom Hearts III	40808
The Elder Scrolls II: Daggerfall	43855
Persona 3	44447
Kingdom Hearts II	47843
Lightning Returns: Final Fantasy XIII	47916
Horizon Zero Dawn	50348
Final Fantasy VIII	51368
Stardew Valley	53870
Final Fantasy XIII-2	68922
King's Quest VI	72074
Final Fantasy XV	74792
Final Fantasy X	84290

game	words
Final Fantasy VII Remake	86487
Final Fantasy IX	99245
Final Fantasy VII	100584
King's Quest Chapters	113053
Final Fantasy XII	128742
The Elder Scrolls V: Skyrim	150041
Persona 4	156753
The Elder Scrolls III: Morrowind	173666
Mass Effect 2	270209
Dragon Age 2	281116
Mass Effect 3	361358
Mass Effect	385744
Persona 5	385817
Star Wars: Knights of the Old Republic	439667
Final Fantasy XIV	689360
Dragon Age: Origins	701258
The Elder Scrolls IV: Oblivion	777177

The distribution of words by RPG type is skewed towards Western RPGs:

```
Wseries = c("Dragon Age", "Horizon", "King's Quest",
            "Mass Effect","Monkey Island","The Elder Scrolls",
            "Star Wars: Knights of the Old Republic",
            "Stardew Valley")
rpgTypeWords = tapply(gs$words, gs$series %in% Wseries,sum)
rpgTypeWords = rbind(rpgTypeWords,paste0(round(100*prop.table(rpgTypeWords),2),"%"))
rpgTypeWords[1,] = format(as.numeric(rpgTypeWords[1,]), big.mark = ",")
colnames(rpgTypeWords) = c("JRPG", "WRPG")
knitr::kable(rpgTypeWords, row.names = F)
```

JRPG	WRPG
2,298,183	3,982,709
36.59%	63.41%

The distribution of ratings is skewed towards adult titles:

```
ratings = c("Dragon Age" = "Adult", "Horizon"="Teen",
            "King's Quest" = "Child", "Mass Effect" = "Adult",
            "Monkey Island" = "Child","The Elder Scrolls" = "Adult",
            "Star Wars: Knights of the Old Republic" = "Teen",
            "Stardew Valley" = "Child", "Chrono Trigger" = "Teen",
            "Final Fantasy" = "Teen", "Kingdom Hearts" = "Child",
            "Persona" = "Adult", "Super Mario RPG"= "Child")
gs$rating = ratings[gs$series]
ratingWords = tapply(gs$words, factor(gs$rating,levels = c("Child","Teen","Adult")),sum)
ratingWords = rbind(ratingWords,paste0(round(100*prop.table(ratingWords),2),"%"))
ratingWords[1,] = format(as.numeric(ratingWords[1,]), big.mark = ",")
knitr::kable(ratingWords, row.names = F)
```

Child	Teen	Adult
489,007	2,060,444	3,731,441
7.79%	32.8%	59.41%

The distribution is also biased towards the early 2000s:

```
gs$YearCat = allGames[match(gs$game,allGames$game),]$yearCat
yWords = t(t(tapply(gs$words,gs$YearCat,sum)))
yWords = cbind(yWords, paste0(round(100*prop.table(yWords),2),"%"))
yWords[,1] = format(as.numeric(yWords[,1]), big.mark = ",")
colnames(yWords) = c("Words","%")
knitr::kable(yWords, row.names = T)
```

	Words	%
1985 - 1989	18,173	0.29%
1990 - 1994	173,149	2.76%
1995 - 1999	294,001	4.68%
2000 - 2004	848,350	13.51%
2005 - 2009	2,254,638	35.9%
2010 - 2014	1,887,406	30.05%
2015 - 2021	805,175	12.82%

Despite these imbalances, we suggest that the corpus is still representative of an average gaming experience. In any case, deciding how to sub-sample each game to create a balanced corpus would depend on the particular research question and type of balance aimed for. For example, if the aim was to balance characters, character groups, certain stages of the game, or depth of dialogue tree. We incorporate as much of each game as possible, so that future studies can make their own decisions.

Where balance is a key issue for statistical methods, we run additional analyses on a random sub-sample where each game has the same number of lines of dialogue (see further below), or calculate measures of dialogue in words per-thousand words within a game.

Sources

The dialogue scripts were sourced from a variety of source types, including public fan transcripts, public wikis and directly from game data. They also varied in completeness from “complete” (virtually all dialogue that a player could experience), to “high” (most dialogue a player would experience on a typical play-through of a game), and “sample” (e.g. a single play-through of the game without alternative dialogue choices, or only transcriptions of portions of the game). See other sections of the SI for analyses of the reliability and representativeness of these sources.

We are very grateful for the work that fans put into organising these sources.

Table of games and sources:

Game	Year	Source	Type	Completeness
Chrono Trigger	1995	https://www.chronocompendium.com/Term/Retranslation.html	game data	complete
Dragon Age: Origins	2009	Game Data	game data	complete
Dragon Age 2	2011	Game Data	game data	complete
Final Fantasy	1987	https://archive.rpgamer.com/games/ff/ff1/info/ff1_script.txt	fan transcript	complete
Final Fantasy II	1988	https://gamefaqs.gamespot.com/ps/916670-final-fantasy-ii/faqs/61436	fan transcript	complete
Final Fantasy IV	1991	https://gamefaqs.gamespot.com/ds/939425-final-fantasy-iv/faqs/53978	fan transcript	high
Final Fantasy V	1992	http://www.finalfantasyquotes.com/ff5/script	fan transcript	high
Final Fantasy VI	1994	https://gamefaqs.gamespot.com/snes/554041-final-fantasy-iii/faqs/70118	fan transcript	high
Final Fantasy VII	1997	http://www.yinza.com/Fandom/Script/	fan transcript	high
Final Fantasy VIII	1999	https://www.neoseeker.com/finalfantasy8/faqs/136092-final-fantasy-viii-script-a.html	fan transcript	high
Final Fantasy IX	2000	https://gamefaqs.gamespot.com/ps/197338-final-fantasy-ix/faqs/42207	fan transcript	high
Final Fantasy X	2001	http://auronlu.istad.org/ffx-script/	fan transcript	high
Final Fantasy X-2	2003	https://www.ffcompendium.com/h/faqs/ffx2scriptaschthehated.txt	fan transcript	high
Final Fantasy XII	2006	http://ffnerdery.blogspot.com/p/final-fantasy-xii.html	fan transcript	high
Final Fantasy XIII	2009	https://en.wikiquote.org/wiki/Final_Fantasy_XIII#Dialogue	wiki	sample
Final Fantasy XIV	2010	https://ffxiv.gamerescape.com	wiki	sample
Final Fantasy XIII-2	2011	https://gamefaqs.gamespot.com/pc/846193-final-fantasy-xiii-2/faqs/64861	fan transcript	high
Lightning Returns: Final Fantasy XIII	2014	https://www.youtube.com/watch?v=Kl_EgMs2V4A	fan transcript	sample
Final Fantasy XV	2016	https://thelifestream.net/final-fantasy-xv-lore/final-fantasy-xv-chapter-by-chapter-lore-exposition-and-development/	fan transcript	high
Final Fantasy VII Remake	2020	https://finalfantasy.fandom.com/wiki/Final_Fantasy_VII_Remake_script	wiki	high
Horizon Zero Dawn	2017	https://game-scripts.fandom.com/wiki/Horizon_Zero_Dawn	fan transcript	sample
King's Quest III: To Heir Is Human	1986	https://kingsquest.fandom.com/wiki/KQ3_transcript	game data	complete
King's Quest I: Quest for the Crown	1987	https://kingsquest.fandom.com/wiki/KQ1SCI_transcript	game data	complete
King's Quest II: Romancing the Throne	1987	https://kingsquest.fandom.com/wiki/KQ2_transcript	game data	complete
King's Quest IV: The Perils of Rosella	1988	https://kingsquest.fandom.com/wiki/KQ4SCI_transcript	game data	complete
King's Quest V	1992	https://kingsquest.fandom.com/wiki/KQ5NES_transcript	game data	high
King's Quest VI	1992	https://kingsquest.fandom.com/wiki/KQ6_transcript	game data	complete
King's Quest VII: The Princeless Bride	1994	https://kingsquest.fandom.com/wiki/KQ7_transcript	game data	complete
King's Quest VIII	1998	https://kingsquest.fandom.com/wiki/KQ8_transcript#Connor	game data	complete
King's Quest Chapters	2015	https://kingsquest.fandom.com/wiki/KQC1_transcript	game data	high
Kingdom Hearts	2002	https://transcripts.fandom.com/wiki/Kingdom_Hearts	fan transcript	high
Kingdom Hearts II	2005	https://transcripts.fandom.com/wiki/Kingdom_Hearts_II	fan transcript	high
Kingdom Hearts 3D: Dream Drop Distance	2012	https://gamefaqs.gamespot.com/3ds/997779-kingdom-hearts-3d-dream-drop-distance/faqs/65008	fan transcript	high
Kingdom Hearts III	2019	https://gamefaqs.gamespot.com/ps4/718920-kingdom-hearts-iii/faqs/78466	fan transcript	high
Mass Effect	2007	See Readme	game data	complete
Mass Effect 2	2010	Dump from custom branch of Legendary Explorer: https://github.com/ME3Tweaks/LegendaryExplorer	game data	complete
Mass Effect 3	2012	Dump from custom branch of Legendary Explorer: https://github.com/ME3Tweaks/LegendaryExplorer	game data	complete
The Secret of Monkey Island	1990	https://gamefaqs.gamespot.com/pc/562681-the-secret-of-monkey-island/faqs/23891	fan transcript	high
Monkey Island 2: LeChuck's Revenge	1991	https://gamefaqs.gamespot.com/pc/562680-monkey-island-2-lechucks-revenge/faqs/79490	fan transcript	sample
The Curse of Monkey Island	1997	https://gamefaqs.gamespot.com/pc/29083-the-curse-of-monkey-island/faqs/60819	fan transcript	high
Persona 3	2006	https://lparchive.org/Persona-3/	fan transcript	sample
Persona 4	2008	https://lparchive.org/Persona-4/	fan transcript	sample
Persona 5	2016	https://lparchive.org/Persona-5/	fan transcript	sample
Star Wars: Knights of the Old Republic	2003	https://github.com/hmi-utwente/video-game-text-corpora/blob/master/Star%20Wars:%20Knights%20of%20the%20Old%20Republic/data/dataset_20200716.csv	game data	complete
Stardew Valley	2016	https://drive.google.com/drive/folders/0BwyXuxAqGS7ueVdFX2dQSUVzcUk	game data	high
Super Mario RPG: Legend of the Seven Stars	1996	https://gamefaqs.gamespot.com/snes/588739-super-mario-rpg-legend-of-the-seven-stars/faqs/30431	fan transcript	high
The Elder Scrolls II: Daggerfall	1996	https://github.com/Interkarma/daggerfall-unity	game data	high
The Elder Scrolls III: Morrowind	2002	https://elderscrolls.fandom.com/wiki/	wiki	sample
The Elder Scrolls IV: Oblivion	2006	https://www.mediafire.com/file/bhkqiqjhfib0waa/dialogueExport.txt/file	game data	complete
The Elder Scrolls V: Skyrim	2011	https://gamefaqs.gamespot.com/pc/615805-the-elder-scrolls-v-skyrim/faqs/69918	fan transcript	high

Corpus pipeline

The pipeline for converting the original sources into the standard format involves several steps:

1. Scrape the online source data into temporary local files.
2. Parse the local files into a standard format using a custom parser, and clean and unify the character names using metadata (see below).
3. Count words and other statistics in the parsed data to produce statistics.
4. Analyse the statistics into formal tests and visualisations.

The corpus pipeline is visualised below:

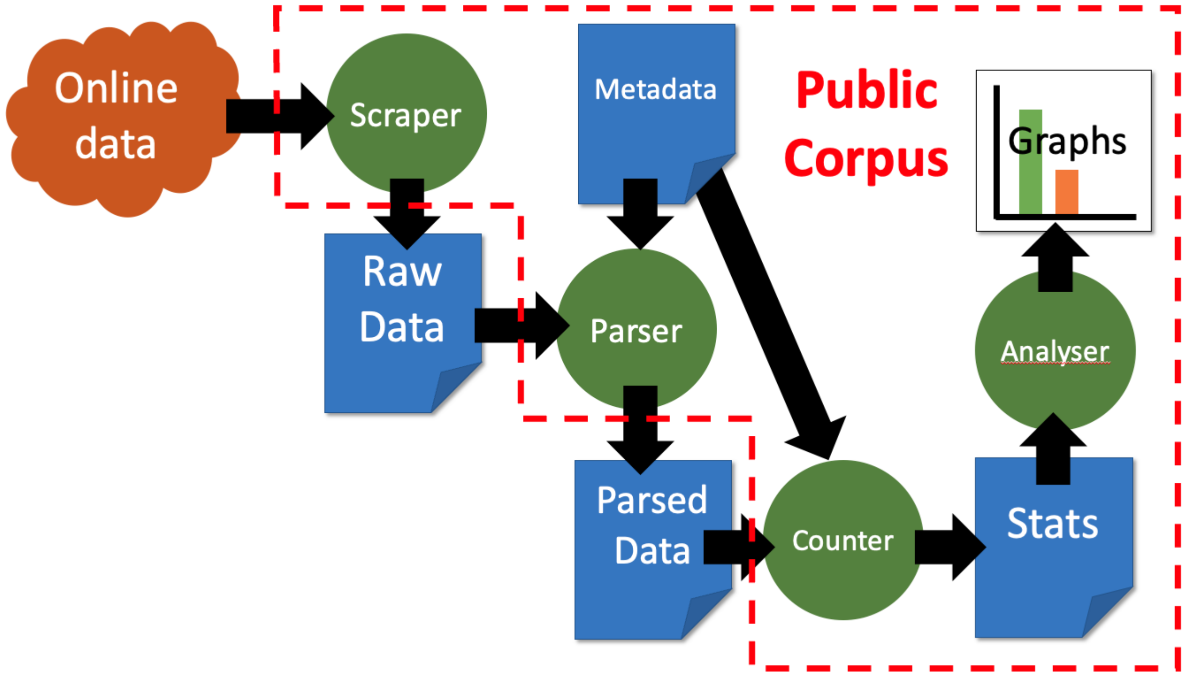


Figure 1: The corpus pipeline

We approach these tasks using replicatable methods in order to create a "self-inflating corpus". The entire pipeline for each game is implemented in python and R code. This means that we can release the code and metadata publicly as a github repository. Now, other researchers can re-run the code in order to re-create the corpus, without needing to share copyrighted materials directly. In addition, the corpus can continue to be expanded and edited in a centralised and consistent way.

The pipeline is set up so that a single parser can be applied to several games. The hope was that code could be re-used. In reality, nearly every game needed its own special parser because of the specifics of the source format or game mechanics. As a result, the parsers account for 10,000 lines of code in the repository. The metadata has around 28,000 lines of gender coding and character name unification details.

Standard format for video game dialogue

The canonical form of the data in the corpus is a JSON format. This is a plain text format that can be read with an ordinary text editor or any open-source JSON tool. This format was chosen because: it can capture pairings of data and recursive structures; is portable and open-source; it is machine-readable (e.g. it is compatible with python dictionaries); and it looks like a screenplay script to human readers.

The script for a game is a list of dictionaries. Each dictionary has a main key which represents the name of the character who is speaking. The value associated with this key is the dialogue they speak. Below is an example of three lines of dialogue from Final Fantasy VII between the characters Barret and Cloud:

```
[
{"Barret": "The planet's full of Mako energy. People here use it every day."},
{"ACTION": "Cloud shrugs."},
{"Barret": "It's the life blood of this planet."},
{"Cloud": "I'm not here for a lecture. Let's just hurry."}
]
```

According to the JSON format, square brackets enclose a list of items separated by commas, and curly brackets enclose a set of key:value pairs.

There are a number of reserved main keys that indicate particular kinds of information. These include:

- ACTION: the value is a description of the action (see the example above).
- LOCATION: the value is a description of the location.
- SYSTEM: the value is a transcription of non-diegetic text that appears to the player, but is not spoken by in-game characters. This includes menu text and tutorial text.
- CHOICE: a branching choice (see below)
- GOTO: The script continues at another location (see below)
- STATUS: A description of some contextual status, used to interpret branching choices (see below)

The dictionary can optionally include other keys, as long as they begin with an underscore, which might convey contextual information. In the example below from Oblivion, the extra data includes the race of the character, the emotion information assigned to the face animator, and the quest ID that this dialogue appears within.

```
{"Velwyn Benirus": "I got the door open. The rest is up to you.",
    "_Race": "Imperial",
    "_Emotion": "Fear 90",
    "_Quest": "0000BCD7"}
```

Branching dialogue is handled using a recursive structure. The main key is labelled "CHOICE" and its value is a list of possible outcomes. Each outcome is a list of dialogue dictionaries, like the normal format. Any of the dialogue dictionaries can itself be a choice structure, allowing recursive branching.

Figures 2 and 3 show a recursive branching dialogue from Final Fantasy VII between Cloud (the player character) and Aerith. There are several possible binary choices, indicated by different shades of the same colour:

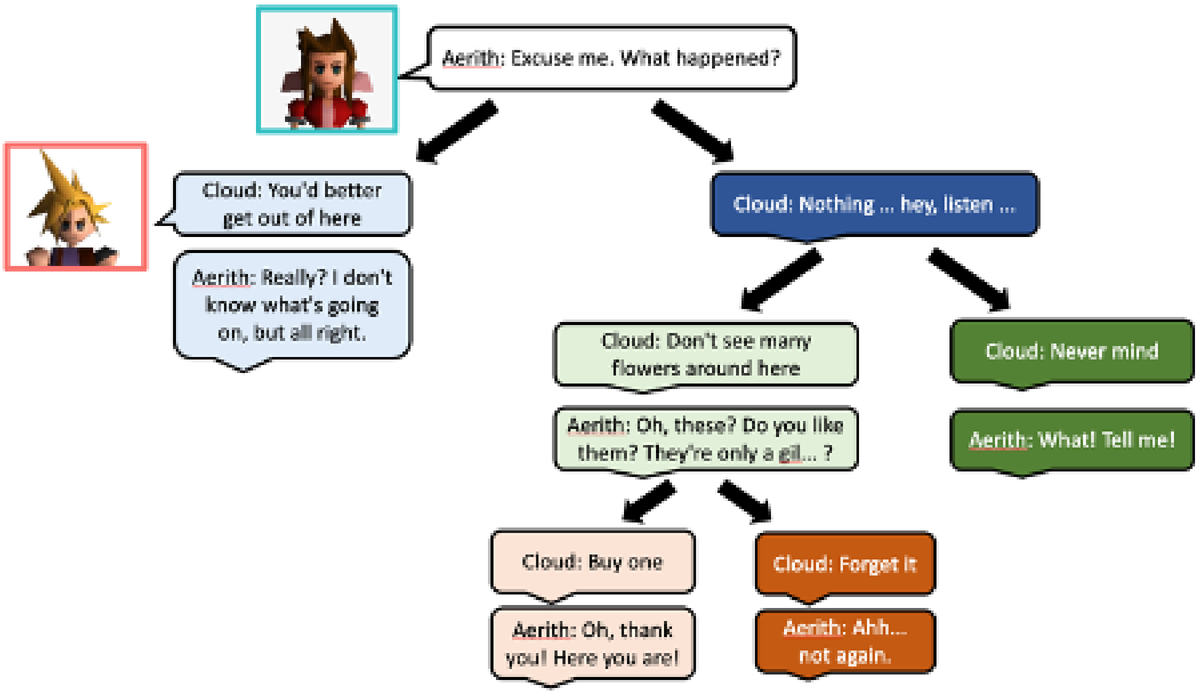


Figure 2: A visualisation of a branching tree structure in Final Fantasy VII

```
{"Aerith": "Excuse me. What happened?"},
{"CHOICE": [
      [
        {"Cloud": "You'd better get out of here."},
        {"Aerith": "Really? I don't know what's going on, but all right."}],
      [
        {"Cloud": "Nothing... hey, listen..."},
        {"CHOICE": [
              [
               {"Cloud": "Don't see many flowers around here"},
               {"Aerith": "Oh, these? Do you like them? They're only a gil... ?"},
               {"CHOICE": [
                          [
                                 {"Cloud": "Buy one"},
                                 {"Aerith": "Oh, thank you! Here you are!"}],
                          [
                                 {"Cloud": "Forget it"},
                                 {"Aerith": "Ahh... not again."}]]}],
              [
               {"Cloud": "Never mind"},
               {"Aerith": "What! Tell me!"},
              ]
        ]}
      ]
  ]}
```

Typically, each outcome will represent the consequences of a player choice. The first entry in each outcome will indicate the trigger for that outcome (e.g. choosing to buy a flower or not). However, the format is designed to capture any type of outcome, including those triggered by: game conditions (e.g. a certain quest is complete); player character statuses (e.g. player character is female); different possible character responses; or random choices. One outcome can be empty, indicating that there's a possibility of hearing no additional dialogue.

In the example below from Skyrim, Ralof responds differently based on the player character's class:

```
{"CHOICE": [
    [
        {"STATUS": "Player race/questline"},
        {"Ralof": "Mage, eh? Well, to each his own."}],
    [
        {"STATUS": "Player race/questline"},
        {"Ralof": "Warrior, good! Those stars will guide you to honor and glory."}],
    [
        {"STATUS": "Player race/questline"},
        {"Ralof": "Thief, eh? It's never too late to take charge of your own fate."}
    ]
]}
```

Although a branching structure can technically represent all types of outcome, some games are more succinctly represented using a mix of branching structures and links between parts of this structure. To handle this, each dialogue dictionary can be given a unique ID (assigned to the key "_ID"). Then there is a reserved main key "GOTO" paired with an ID. This is an instruction that the script resumes at the dictionary with the given ID.

In the example below from Mass Effect 3, Din Korlack's line changes depending on the gender of the player character Shepard. If Shepard is female, then the first outcome is experienced. If not, then the second outcome is experienced, but then there's a GOTO line that rejoins the script at Zaeed's first line. That is, the text dialogue experienced by the player is identical except for the pronouns.

```
{"CHOICE": [
    [
        {"STATUS": "Shepard is female (12/0)"},
        {"Din Korlack": "Shepard's investigating. She's... a recent acquaintance.", "_ID": "689365"}
        {"Zaeed Massani": "How recent?", "_ID": "689367"},
        {"Din Korlack": "Very.", "_ID": "689368"},
        {"Zaeed Massani": "Shit. All right, I'm listening.", "_ID": "689369"}],
    [
        {"STATUS": " (-1/0)"},
        {"Din Korlack": "Shepard's investigating. He's... a recent acquaintance.", "_ID": "689366"},
        {"GOTO": "689367"}]]}
```

Metadata format

The metadata file is a JSON format file with the following fields:

- "game": Full name of the game.
- "series": Name of the series(e.g. "Final Fantasy").
- "year": Year of publication.
- "source": Web source for the raw script.
- "sourceFeatures": What the source contains, see below.
- "characterInfoSource" (optional): Source for a wiki-style listing for automatic extraction of character features.
- "sampleOnly" (optional): True if the source is only a small sample of the full script.
- "notes" (optional): Any coder notes about the data.
- "parserParameters": parameters for the parser. Must include "parser" (name of the parser that's used) and "fileType" (extension of files in the 'raw' folder to parse, 'html' by default). See the parsers for further arguments that can be passed.
- "mainPlayerCharacters": list of main playable characters (e.g. ["Cloud"])
- "characterGroups": Dictionary of groups and the characters that are members of each group (see below)
- "aliases": A mapping from alternative names to canonical names. This helps the parser fix spelling mistakes and unify character dialogue written under alternative names, (e.g. before their name is known, `"Flower girl": "Aerith"`).

'sourceFeatures' is a dictionary with the following properties:

- "type": One of 'fan transcript', 'game data', 'wiki'
- "completeness": One of 'sample', 'high', 'complete'
- "dialogueOrder": true (appropriate for studying transisions between speakers) or false (some other order, e.g. ordered by)
- "choices": What is the coverage of dialogue choices?
  - "NA" (game has no choices)
  - "not included"
  - "partial"
  - "complete"

Character groups

The characterGroups field in the metadata is a mapping from group names to a list of character names who are members of that group. This is used to code each character's gender.

The group labels can be any string, and there can be as many groups as is necessary to capture the diversity in the character groupings. Character names should be the final canonical names, after the aliases are applied.

```
    "characterGroups": {
        "male": [
            "Cloud",
            "Barret",
            ...
            ],
        "female": [
            "Tifa",
            "Aerith",
            ...
            ],
        "neutral": [
            "Chocobo",
            "Jenova",
            ...
            ],
        ...
        }
```

Aliases

Sometimes, a character has multiple names in the script. This can happen if:

- The character is disguised as another character (e.g. Prince Edgar is transformed into King Otar in King's Quest VII).
- The character speaks before revealing their name (e.g. Aerith in Final Fantasy VII)
- The name is shortened (e.g. "Red" instead of "Red XIII" in Final Fantasy VII)
- There are stage directions in the name (e.g. "Cara [to Mid]" in Final Fantasy V)
- There is variation in upper case/lower case letters (e.g. "Shinra manager" and "Shinra Manager").
- There is a typo in the script.

These issues can be fixed by adding alias information to the metadata. This is placed after the "characterGroups". It includes a list of 'wrong' names and what they should be corrected to. E.g. below all instances of "Flower girl" are converted to "Aerith".

```
"aliases": {
        "Flower girl": "Aerith",
        "Aries": "Aerith",
        "Muuki": "Mukki",
        "Red": "Red XIII",
        "Shinra manager": "Shinra Manager",
        "Village headman": "Village Headman",
        ...
    }
```

Some scripts assign one line of dialogue to multiple characters if they're saying the same thing at the same time. This can lead to some 'character names' like "Cloud & Aerith". These can be split into individual lines for each character by using a list in the aliases (instead of just a character name string):

```
"aliases": {
    "Cloud & Aerith": ["Cloud", "Aerith"],
    "Biggs, Jessie, & Wedge": ["Biggs","Jessie", "Wedge"]
    }
```

Sometimes, multiple characters are given the same label if they are not known to the transcriber or the player at the time of speaking. Characters can be identified by line of dialogue. In the example below, the label "???" is converted based on the line of dialogue. For example, if the dialogue matches "Ha ha ha ha. I'm so ... lonely", then it will be converted to "Birdo":

```
"aliases":{
            "???": {
              "Birdo": ["Ha ha ha ha. I'm so ... lonely.",
                              "Oh ... If you had played with me,",
                              "Thanks!"],
                    "Jinx": ["You did well for your inexperience, Jagger."],
                      }
            }
```

Note that:

- Matching is done by checking if the dialogue in the script *starts with* the line of dialogue in the metadata, so there's no need to include the whole line, just a recognisable portion.
- If a line of dialogue in the script does not match any of the lines in the metadata, the name remains as it is.
- Alias changes apply during parsing, so the incorrect names won't appear in the script. Therefore, when including aliases, the names in the "characterGroups" list should reflect the corrected name or individual name, not the original name.

Error checking

This section describes the procedure for error checking in order to ensure that the data is accurate and representative.

If the data source is directly from the game files, then only the check for false positives and parsing errors is required. Otherwise, both tests below are required.

After the checks have been carried out, the results should be added into the metadata after the "source" entry. for example:

```
"errorChecks": {
  "truePositive_numTestsDone": "5",
  "truePositive_numParsingErrors": "0",
  "truePositive_numSourceErrors": "1",
  "truePositive_notes": "One line inaccurate transcript: [EXAMPLE]",
  "falsePositive_numTestsDone": "5",
  "falsePositive_numErrors": "1",
  "falsePositive_notes": "Parsing error: no space after full stop."
}
```

A Github issues should be raised for any problem that could potentially be fixed.

Check for true positives and transcription errors

Follow the procedure below to check for true positives (lines that are in the source that are in the game) compared with transcription errors (lines that may have been mis-transcribed in the source).

1. Find a video on YouTube of someone playing the game. Try to find one that documents an entire play through the game (rather than clips), without mods, and that is not a speed run or specialist run (e.g. pacifist). Typically, "let's play" videos will be suitable.
2. If a run is split over several videos, choose one at random.
3. Choose a random place in the video. This website will help you do that: https://correlation-machine.com/VideoGameCorpus/randomVideoLocation.html
4. Find the next piece of dialogue. If you reach the end of the video, loop around to the beginning. Look at up to three lines of dialogue that are spoken together.
5. Search the data.json file to answer:

- Does the dialogue in the video exist in the corpus? (ignoring small errors in punctuation, capitalisation, and also ignoring typos - the question is whether the line is represented somehow)
- Is the text of the transcript of the video accurate? Note that consecutive lines spoken by the same character are collapsed into one line in the corpus.
- Is the structure of the conversation correct? (Are options defined in "CHOICE" structures? Are all options available? Does the sequence match?). Note that the dialogue in the game may be randomised, optional or status-dependent, so all lines in the corpus may not appear in the video. The question is whether the dialogue in the video is covered by the corpus.
- If there are any errors, can we identify the source?
  - Error in parsing program.
  - Error in original transcript source.

Repeat steps 2-5 for 5 parts of the video.

Check for false positives and parsing errors

Repeat the following procedure 5 times to check for false positives (lines in the source that are not really in the game) and parsing errors (lines that are in the data, but parsed incorrectly in terms of character assignment or dialogue structure).

Pick a random line in the corpus data.json file. Confirm that:

- The character name is plausible (not some possible parsing error like "and so").

- There are no strange typographic characters.
- There are no obvious parsing errors (e.g. another character's dialogue line enclosed in the dialogue string, words not separated properly). If it is possible, find this line in the source transcript. This might involve finding the location of the source in the meta.json file, then using a google search like: "site:http://www.yinza.com/Fandom/Script/ "This is a church in the""
- Confirm that the line in the source has been correctly parsed into the corpus.

Number of errors

As part of the error checking processes, and for other processes, 212 bug reports were filed to the Github repository and all were fixed.

The number of errors discovered in the corpus using the method above are summarised below. However, the process above used as a process of due diligence during development which was intended to find and fix errors. Therefore, they are not ideal as measures of the final quality of the data. Still, they demonstrate relatively low levels of error. We were initially concerned with the quality of transcription in fan transcripts, but this type of error was rare.

Load the data from meta.json files:

```
library(rjson)
stats = read.csv("../results/generalStats.csv",stringsAsFactors = F)
# Remove alternative measures
stats = stats[stats$alternativeMeasure!="True",]
stats = stats[!is.na(stats$words),]
folders = unique(stats$folder)
d = NULL
for(folder in folders){
  js = fromJSON(file = paste0(folder,"meta.json"))
  if(!is.null(js$`error checks`)){
    errorChecks = js$`error checks`
    dx = data.frame(folder = folder,
       TruePosTestsRun = errorChecks$truePositive_numTestsDone,
       TruePosTranscriptErrors = errorChecks$truePositive_numParsingErrors,
       TruePosSourceErrors = errorChecks$truePositive_numSourceErrors,
       FalsePosTestsRun = errorChecks$falsePositive_numTestsDone,
       FalsePosErrors = errorChecks$falsePositive_numErrors)
    d = rbind(d,dx)
  }
}
d$TruePosTestsRun = as.numeric(d$TruePosTestsRun)
d$TruePosTranscriptErrors = as.numeric(d$TruePosTranscriptErrors)
d$TruePosSourceErrors = as.numeric(d$TruePosSourceErrors)
d$FalsePosTestsRun = as.numeric(d$FalsePosTestsRun)
```

```
## Warning: NAs introduced by coercion
```

```
d$FalsePosErrors = as.numeric(d$FalsePosErrors)
```

```
## Warning: NAs introduced by coercion
```

Mean proportion of tests with parsing errors:

```
d$propTranscriptionErrors = d$TruePosTranscriptErrors/d$TruePosTestsRun
mean(d$propTranscriptionErrors,na.rm=T)
```

```
## [1] 0
```

Mean proportion of tests with source errors:

```
d$propSourceErrors = d$TruePosSourceErrors/d$TruePosTestsRun
mean(d$propSourceErrors, na.rm=T)
```

```
## [1] 0.08541667
```

```r
range(d$propSourceErrors, na.rm=T)
```

```
## [1] 0.0 0.5
```

Mean proportion of tests with false positive errors:

```r
d$propFalsePosErrors = d$FalsePosErrors/d$FalsePosTestsRun
mean(d$propFalsePosErrors, na.rm=T)
```

```
## [1] 0.005
```

There were no transcription errors and only one false positive error logged during error checking. There were many issues during development, but these were identified by different means and fixed.

The main source of errors were source errors, which were observed in 17 games. For example, in the play-through video of Skyrim that was chosen, conversations with some NPCs were observed that were not in the source transcript (hence the "completeness" of this game being 'high' instead of 'complete'). Or for Monkey Island, the source only records one path through a dialogue tree, so there was dialogue that was not captured. Both of these cases were known limitations of the source.

Overall, the data exhibited high accuracy and representativeness.

3 Coding gender for characters in video games

Coding gender for characters in video games

Introduction

The aim is to code the (Western) player community's interpretation of the gender of each character. Any label can be used as a gender category. Letters should be lowercase. Some conventions help the unification of the data later on. If you assign a character to the male category, use the label "male" (rather than "man", "boy" etc.). If you assign a character to the female category use the label "female" (rather than "woman", "lady", "girl" etc.). Characters can be coded as having no gender (e.g. "genderless"), but only with positive evidence.

Characters can belong to multiple groups. For example, a character might belong to both the groups "female" and "trans".

You might assign characters to the "neutral" category if:

- They are supernatural or cosmic beings, or very different species, or inanimate object for which interpreting gender using the indicators below is very difficult.
- The name refers to a group (e.g. "people") or general species (e.g. "Chocobo") which might include more than one gender.
- They have a generic name (e.g. "Guard", "Villager") AND their gender isn't readily identifiable from the indicators below. This includes where it isn't possible to identify the specific character. They are an invisible third person narrator, especially in text-based adventures. (Although sometimes the gender of the narrator can be determined and should be coded accordingly, for example where they are a named character – such as Varric in Dragon Age 2 – or where there is sufficient indication of their gender from their voice/pronouns used etc.)
- The dialogue comes from the game system not tied to a character, e.g. in tutorials.

Indicators

There are various indicators used to code the gender of the game characters. Some are more authoritative than others, but any individual indicator is defeasible. Usually several indicators are required to make a decision.

1. Category defined by community wiki

Some games have community-written wikis or discussion boards with pages for each character. Some of these define the gender of the character (e.g. https://finalfantasy.fandom.com/wiki/Cloud_Strife under "Gender" in the info bar on the right). The community wiki categorisation should not necessarily be taken as the gold standard. Firstly, the category can be assigned erroneously (as appears to be the case for some minor characters), debatable, or change according to the language/release of the game. For example, in Mass Effect, the Asari species is, according to the game lore, a mono-gender species. However, all Asari characters have visual cues, pronouns, and voice actors which suggest players will confer female gender to them.

Secondly, coding on wikis are likely based on a set of gender indicators such as the ones listed below. So the information in wikis can be seen as 'shorthand' for a gender coding that takes into account many factors. However, in most cases, these will agree with the coding scheme presented here.

2. Character name

The gender of some characters is revealed in their name, e.g. "King Graham". However, the names in the corpus data are sometimes given by the transcribers rather than the game, so other indicators may be needed when coding characters like "Adolescent Male #1".

3. Appearance

Visual appearance may provide clear cues to gender, including visual signifiers (e.g. pink bows or makeup are often restricted to female characters) or secondary sex characteristics (e.g. breasts). This is not to suggest that all and only females wear makeup, or have specific secondary sex characteristics. Their use in a game, however, is frequently a choice by developers in order to indicate that such characters are not male. Some wiki pages have higher-resolution fan art for characters (though these may not be reliable).

4. Claims that the character themselves make in dialogue Characters rarely announce their gender directly, but still might bring it up in conversation. E.g. in King's Quest Chapter 1:

```
Amaya: "I'm a woman of action, if you couldn't tell."
```

Or in Stardew valley:

```
"Abigail": "Oh, it's because I'm a girl... isn't it? Ugh..."
```

5. How other game characters refer to them

Including pronouns, but this can be unreliable and switch between translations. E.g. https://finalfantasy.fandom.com/wiki/Quina_Quen

6. Pronouns used in community wikis

In a wiki page about a character, they may be referred to by a specific pronoun that might give you a clue.

7. Gender of the voice actor

If the voice actor is known, this might provide a clue to the gender, but may be quite unreliable.

Edge cases

For the majority of characters, coding is straightforward, as there are multiple indicators of gender present and in agreement. Any coding decisions in edge cases should be documented in the "README.md" files for each game. Difficult cases tend to fall into one of five categories:

1. Internal ambiguity

in a small range of cases, the gender of some characters is treated inconsistently within a game. Quina Quen from Final Fantasy IX falls into this category: Quina is referred to by characters in the English translation of the game as 's/he' but using the possessive pronoun 'his'. Narratively, Quina is a member of a genderless race called the Qu, but the game mechanics are inconsistent in their treatment of Quina, for instance, certain spells that only affect male characters affect Quina, and certain abilities that only affect female characters do not. This inconsistency affects players' experience of a character's gender (see e.g. discussions of Quina, here or here).

2. External disagreement

We are aiming to code a typical player's experience of characters' gender, but there are cases in which players disagree. For example, dialogue in Final Fantasy VII suggests that the character 'Big Bro' (also known as 'Beautiful Bro') is a cross-dressing man, but some players read Big Bro as a trans woman. In the Final Fantasy VII remake, Big Bro is replaced with a character called 'Jules' who the wiki describes as an effeminate man, but some players have interpreted as a trans man (e.g. Big Bro in Final Fantasy VII: e.g. here).

For cases of type (1) and (2) we looked for agreement between as many indicators as possible, and documented decisions in the readme.md files for each game.

3. Counting

Determining whether two or more character appearances are instances of the same character, or different characters, can be challenging. These cases tend to fall into one of two varieties. Firstly, there are generics: NPCs that share a character model and cannot be distinguished by name, voice acting, appearance or contextual clues. Soldiers and guards often fall into this category, but townsfolk and other background characters can as well. Where it is possible to consistently differentiate such NPCs we have done so using aliases (e.g. if two identical soldiers are guarding a gate, but one is on the left and the other on the right). Where it is not possible to do so, we have noted that the character is a generic rather than an unique individual in a game's readme.md file. The second variety is where characters appear in disguise or have undergone a transformation changing their visual appearance. Depending on contextual clues, and in particular, what the player can be expected to know, we either consider the different versions of a character a single person, or multiple people. For instance, in King's Quest VI, a genie appears to the player in different forms that vary by age and gender. These interactions occur before the player has met the genie in their true form, and the player is unlikely to realise as they experience these disguises that they belong to a male genie. In this case we coded each disguise of the genie as a separate character. By contrast, a character who cross-dresses in a scene but is easily recognisable (such as Cloud in Final Fantasy VII) is not recorded as a separate character. Decisions and their rationale are documented in the readme.md file for each game.

4. Gendered objects

Some video game characters do not have a physical form usually associated with having a gender, such as talking books or vending machines. However, a game may nonetheless treat such a character as gendered, and players confer gender on them, based on their voice acting, the pronouns other characters use to describe them, and so forth. Such cases are documented in the readme.md file for each game.

5. Insufficient evidence

Sometimes there is insufficient evidence available to categorize the gender of a character. This most often occurs for very minor characters, that appear off-screen or in the distance, and are unvoiced. However there are exceptions, particularly with characters from fantastic races or with unusual physical features. For instance, the 'bangaa' species from Final Fantasy XII are not sexually dimorphic, and nothing about their physical appearance is indicative of their gender, including their clothes and accessories. Some bangaa are indicated to be male or female in gender through the use of pronouns and voice acting, but most of the bangaa NPCs are unvoiced. Where insufficient evidence was available, characters were coded as 'neutral'.

4 Analysing Gender Balance in Video Games

Analysing gender balance in video game dialogue

Introduction

This section looks at the proportion of dialogue in video games for female and male characters. The tests below test:

- Descriptive statistics of gender balance
- Change over time
- Statistical difference from hypothetical baselines.

The amount of dialogue can be measured in several different ways, including the number of syllables, words, sentences and script lines. We prefer a measure in words because this measure is:

- Reliable: Words are not totally straightforward to measure, but there are standard solutions. In contrast, estimates of syllables and sentences rely on additional assumptions.
- Valid: Words are a more valid measure of text length than the number of lines, because a line can vary from one word to several sentences.

Basic statistics

Load libraries

```
library(ggplot2)
library(ggrepel)
library(GGally)
```

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

```
library(grid)
library(rjson)
library(tidyr)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(mgcv)
```

```
## Loading required package: nlme
```

```
## This is mgcv 1.8-38. For overview type 'help("mgcv-package")'.
```

```
library(tidymv)
```

```
## tidymv will be deprecated. Users are recommended
##     to check out the in-progress replacement tidygam
##     (https://github.com/stefanocoretta/tidygam).
```

```
library(ggpubr)
```

Load statistics for all groups for all games and work out proportion of female dialogue compared to all female and male dialogue for different measures of length.

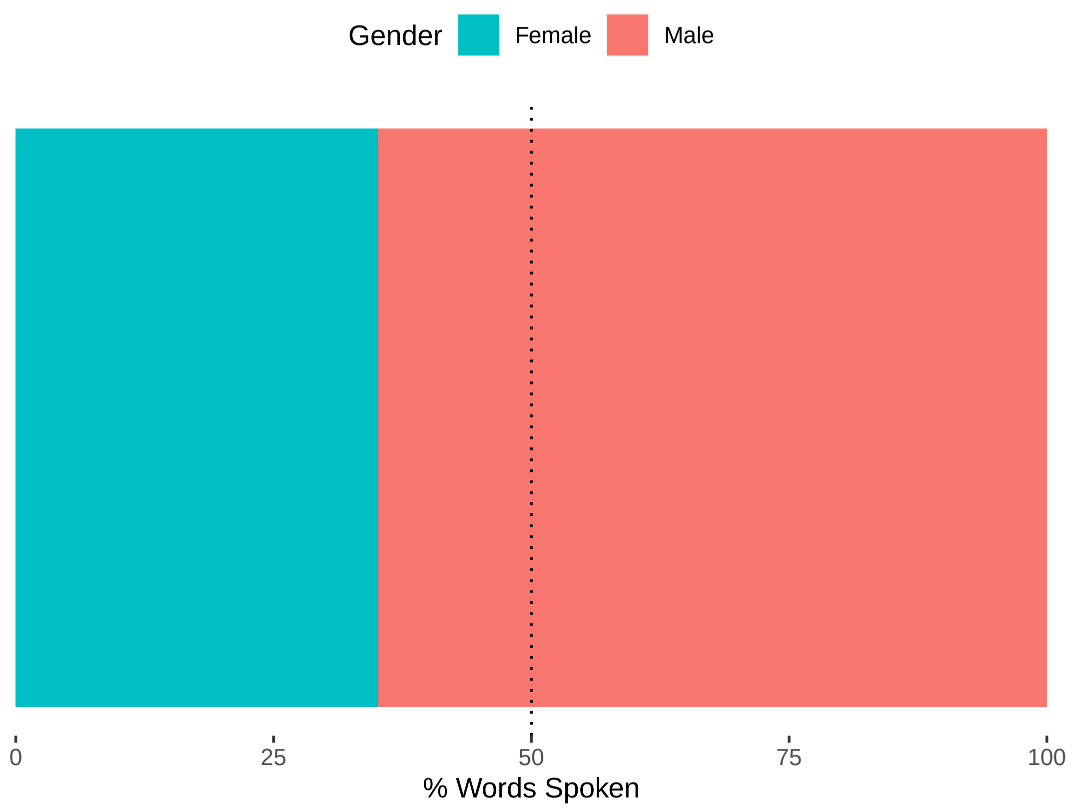
```
stats = read.csv("../results/generalStats.csv",stringsAsFactors = F)
# Remove alternative measures
stats = stats[stats$alternativeMeasure!="True",]
stats = stats[!is.na(stats$words),]
d = NULL
folders = unique(stats$folder)
for(folder in folders){
  sxM = stats[stats$folder==folder & stats$group == "male",]
  sxF = stats[stats$folder==folder & stats$group == "female",]
  js = fromJSON(file = paste0(folder,"meta.json"))
  if(nrow(sxM)>0 & nrow(sxF)>0){
    d = rbind(d,
      data.frame(
        folder = folder,
        series = sxF$series,
        game = sxF$game,
        lines = sxF$lines / (sxF$lines + sxM$lines),
        words = sxF$words / (sxF$words + sxM$words),
        sentences = sxF$sentences / (sxF$sentences + sxM$sentences),
        syllables = sxF$syllables / (sxF$syllables + sxM$syllables),
        fw = sxF$words,
        mw = sxM$words,
        year = js$year,
        shortName = tail(strsplit(folder,"/")[[1]],1),
        femCharProp = sxF$numCharacters / (sxF$numCharacters+sxM$numCharacters),
        maleWordsPerChar = 1000* ((sxM$words/(sxM$words+sxF$words)) / sxM$numCharacters),
        femaleWordsPerChar = 1000* ((sxF$words/(sxM$words+sxF$words)) / sxF$numCharacters)
      ))
  }
}
shortNameChanges = list(
  c("KingdomHearts", "KH"),
  c("KingsQuest", "KQ"),
  c("_Remake", "-R"),
  c("TheSecretOfMonkeyIsland", "MI1"),
  c("MonkeyIsland2", "MI2"),
  c("TheCurseOfMonkeyIsland", "MI3"),
  c("SuperMarioRPG", "SMario"),
  c("FFX_B", "FFX"),
  c("MassEffect1", "ME1"),
  c("MassEffect2", "ME2"),
  c("MassEffect3C", "ME3"),
  c("B$", ""),
  c("DS$", ""),
  c("StarWarsKOTOR","KOTOR"),
  c("DragonAgeOrigins","DAO"),
  c("_","")
)
for(snc in shortNameChanges){
  d$shortName = gsub(snc[1],snc[2],d$shortName)
}
```

Overall proportion of female dialogue:

```r
allGamesPropFemale = (sum(stats[stats$group=="female",]$words,na.rm=T)/
                        (sum(stats[stats$group=="female",]$words,na.rm=T)+
                           sum(stats[stats$group=="male",]$words,na.rm=T)))*100
allGamesPropFemale
```

```
## [1] 35.16371
```

```r
allGamesPropMale = 100-allGamesPropFemale
dx = data.frame(
  Gender=factor(c("Male","Female"),levels=c("Male","Female")),
  percentageWords=c(allGamesPropMale,allGamesPropFemale))
mainStatGraph = ggplot(dx,aes(x=1,y=percentageWords,fill=Gender))+ geom_bar(stat='identity')+
  geom_hline(yintercept=50,linetype="dotted") +
  coord_flip(ylim = c(0,100)) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.text.y = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = "top") +
  scale_fill_discrete(breaks=c("Female","Male"),name="Gender")+
  ylab("% Words Spoken") +
  xlab("")
mainStatGraph
```




```r
pdf('../results/graphs/OverallWordsByGender.pdf',width=6,height=3)
mainStatGraph
dev.off()
```

```
## pdf
##   2
```

Plots for specific series:

```r
seriesToPlot = names(table(d$series)[table(d$series)>1])
for(series in seriesToPlot){
  dx = d[d$series==series,]
  dx$game = factor(dx$game, levels = unique(dx$game[order(dx$year,decreasing = T)]))
  dx$words.m = 1 - dx$words
  sx = pivot_longer(dx,c("words","words.m"))
  sx$group = factor(sx$name,
                    levels=c("words.m","words"),
                    labels=c("Male","Female"))
  sx$measurement = sx$value*100

  gx = ggplot(sx, aes(x=game,y=measurement,fill=group)) +
    geom_bar(stat='identity')+
    geom_hline(yintercept=50,linetype="dotted") +
    coord_flip(ylim = c(0,100)) +
    theme(panel.grid.major = element_blank(),
          panel.grid.minor = element_blank(),
          panel.background = element_blank(),
          legend.position = "top") +
    scale_fill_discrete(breaks=c("Female","Male"),name="Gender")+
    ylab("% Words Spoken") +
    xlab("")
  gx

  # write
  fileName = paste0("../results/graphs/series/",series,".pdf")
  fileName = gsub(" ","_",fileName)
  pdf(fileName,width=5,height=4)
  gx
  dev.off()
}
```

Summary of proportion of female dialogue according to different measures:

```r
knitr::kable(d[,c("game","syllables","words","sentences","lines")],digits = 3)
```

game	syllables	words	sentences	lines
The Elder Scrolls III: Morrowind	0.328	0.332	0.343	0.353
The Elder Scrolls IV: Oblivion	0.408	0.408	0.413	0.422
The Elder Scrolls II: Daggerfall	0.215	0.212	0.200	0.174
The Elder Scrolls V: Skyrim	0.306	0.305	0.305	0.305
Final Fantasy V	0.267	0.270	0.312	0.329
Final Fantasy X	0.323	0.322	0.318	0.303
Final Fantasy X-2	0.478	0.484	0.548	0.606
Final Fantasy XV	0.200	0.195	0.144	0.113
Final Fantasy XIV	0.333	0.334	0.341	0.358
Final Fantasy VIII	0.318	0.322	0.299	0.320
Final Fantasy VII Remake	0.369	0.372	0.383	0.368
Final Fantasy	0.230	0.229	0.242	0.259
Final Fantasy VI	0.180	0.182	0.234	0.219
Final Fantasy IV	0.106	0.106	0.127	0.136
Lightning Returns: Final Fantasy XIII	0.542	0.546	0.581	0.606
Final Fantasy XII	0.273	0.272	0.271	0.262
Final Fantasy IX	0.262	0.262	0.273	0.270
Final Fantasy XIII-2	0.413	0.412	0.411	0.408
Final Fantasy XIII	0.419	0.421	0.444	0.454
Final Fantasy II	0.297	0.298	0.296	0.274
Final Fantasy VII	0.258	0.259	0.272	0.282

game	syllables	words	sentences	lines
Super Mario RPG: Legend of the Seven Stars	0.166	0.166	0.181	0.161
Star Wars: Knights of the Old Republic	0.280	0.286	0.293	0.283
Chrono Trigger	0.375	0.380	0.397	0.416
Horizon Zero Dawn	0.441	0.444	0.482	0.535
Monkey Island 2: LeChuck's Revenge	0.172	0.168	0.159	0.131
The Secret of Monkey Island	0.079	0.079	0.089	0.067
The Curse of Monkey Island	0.069	0.067	0.060	0.052
King's Quest I: Quest for the Crown	0.123	0.112	0.088	0.118
King's Quest VI	0.062	0.064	0.071	0.059
King's Quest VIII	0.142	0.142	0.143	0.113
King's Quest VII: The Princeless Bride	0.490	0.492	0.529	0.562
King's Quest Chapters	0.344	0.349	0.355	0.363
King's Quest V	0.295	0.293	0.277	0.249
King's Quest II: Romancing the Throne	0.797	0.798	0.689	0.692
King's Quest III: To Heir Is Human	0.239	0.239	0.223	0.233
King's Quest IV: The Perils of Rosella	0.807	0.800	0.779	0.754
Kingdom Hearts	0.228	0.221	0.199	0.182
Kingdom Hearts II	0.155	0.155	0.150	0.143
Kingdom Hearts III	0.139	0.140	0.135	0.126
Kingdom Hearts 3D: Dream Drop Distance	0.067	0.067	0.064	0.070
Dragon Age 2	0.400	0.401	0.415	0.409
Dragon Age: Origins	0.322	0.322	0.318	0.314
Persona 3	0.431	0.430	0.459	0.481
Persona 4	0.478	0.476	0.498	0.487
Persona 5	0.396	0.396	0.421	0.401
Mass Effect 3	0.407	0.403	0.407	0.401
Mass Effect	0.417	0.418	0.415	0.412
Mass Effect 2	0.441	0.435	0.405	0.415
Stardew Valley	0.450	0.452	0.427	0.461

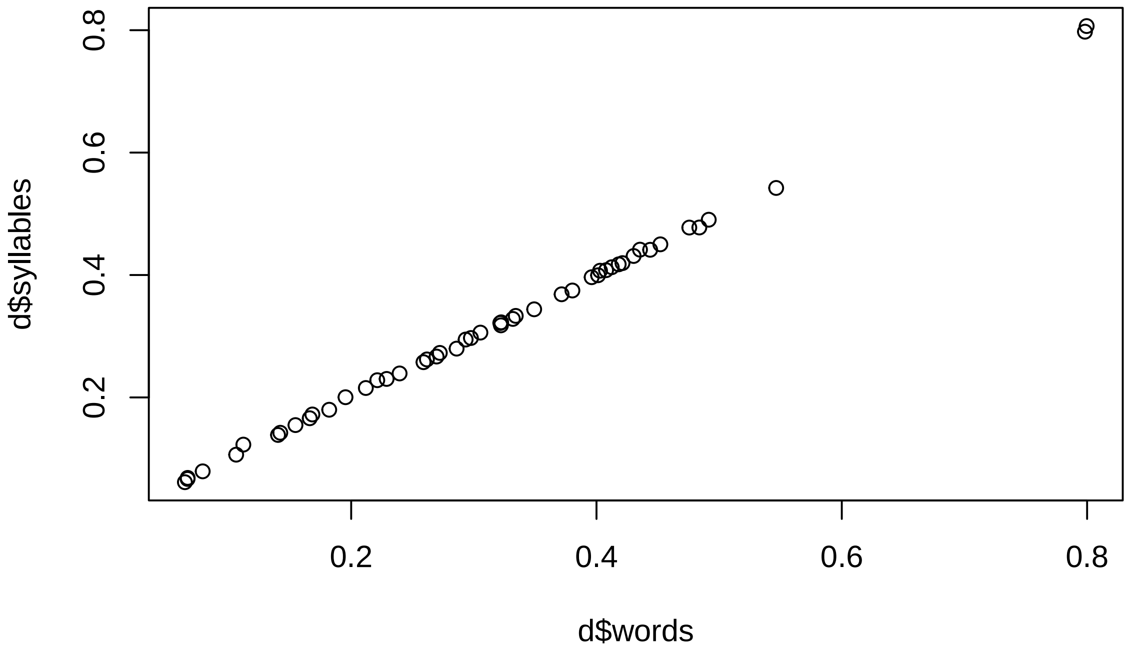
Write some stats for the paper:

```
# Write general stats
minFemaleProp = d[d$words==min(d$words),]
maxFemaleProp = d[d$words==max(d$words),]

minFemalePropStr = paste0(round(100*minFemaleProp$words),"\\% (\\emph{",minFemaleProp$game,"})")
maxFemalePropStr = paste0(round(100*maxFemaleProp$words),"\\% (\\emph{",maxFemaleProp$game,"})")

cat(minFemalePropStr,file="../results/latexStats/gameWithMinimumFemaleWords.tex")
cat(maxFemalePropStr,file="../results/latexStats/gameWithMaximumFemaleWords.tex")

cat(round(100*d[d$game=="Final Fantasy XV",]$words),
 file = "../results/latexStats/FFXV_PropFemaleDialogue.tex")
```

The proportion of female dialogue ranges from 6% (*King's Quest VI*) to 80% (*King's Quest IV: The Perils of Rosella*).

The estimates from different measures of length are highly correlated:

```
plot(d$words,d$syllables)
```




```
cor(d$words,d$syllables,method = 'k')
```

```
## [1] 0.9967347
```

```
cor(d$words,d$sentences)
```

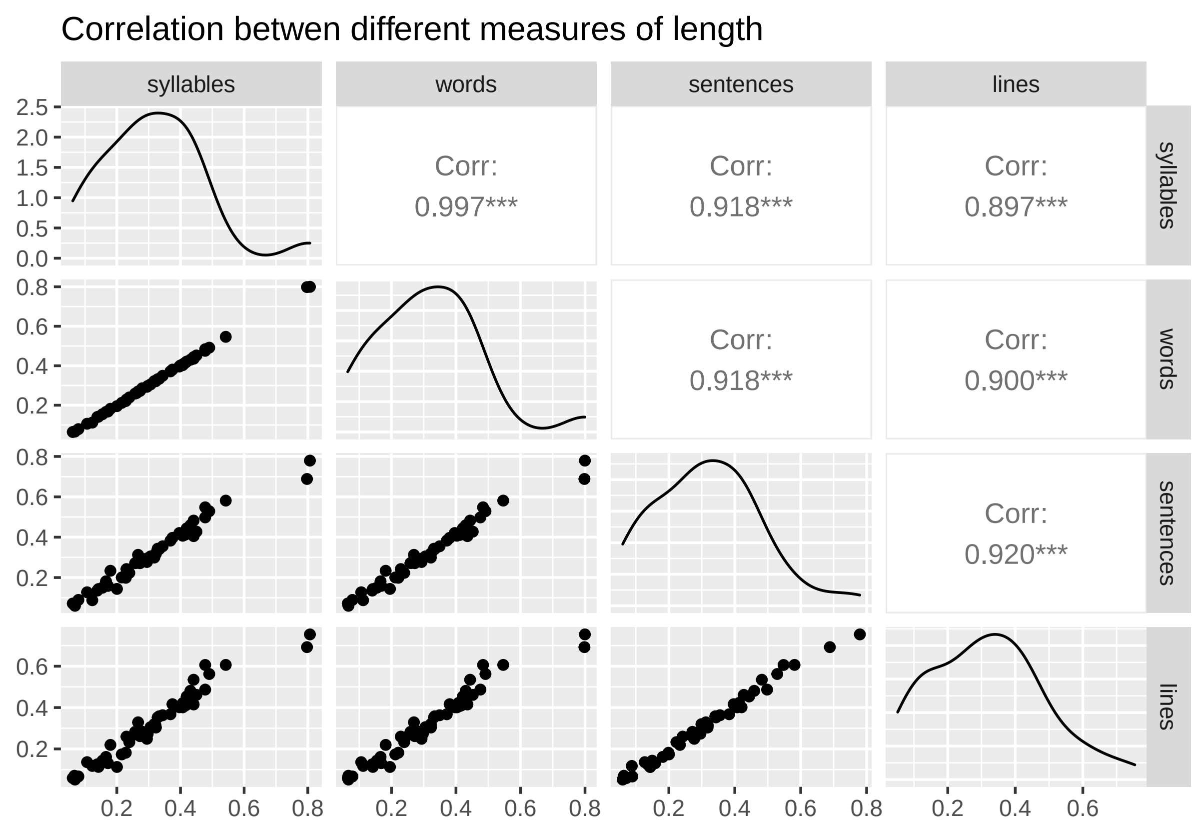
```
## [1] 0.9857857
```

```
cor(d$syllables,d$sentences)
```

```
## [1] 0.983933
```

Correlation between all measures, using Kendall's correlation for significance:

```
measures = c("syllables","words","sentences","lines")
corK = function(data,mapping,...){
  ggally_cor(data,mapping,method="kendall")
  }
ggpairs(d[,measures],
        upper = list(continuous = corK),
        title="Correlation betwen different measures of length")
```



What are the maximum differences between measures of syllables, words and sentences?

```
mx = c("syllables","words","sentences")
d$maxDiff = apply(d[,mx],1,function(X){max(abs(outer(X,X,"-")))})
```

For 95% of games, the measures of syllables, words, and sentences were within 5.5 percentage points of each other. Two games had higher differences. The first is King's Quest 2, which has a very small amount of dialogue, so small differences in counts can lead to large differences in proportions.

The second is Final Fantasy X-2 (7 percentage points difference), which has a higher estimate for sentences than for words or syllables.

```
x2 = read.csv("../data/FinalFantasy/FFX2/stats_by_character.csv",stringsAsFactors = F)
x2 = x2[x2$group %in% c('male','female'),]
summary(lm(words~lines*group, data=x2[x2$group %in% c("male","female"),]))
```

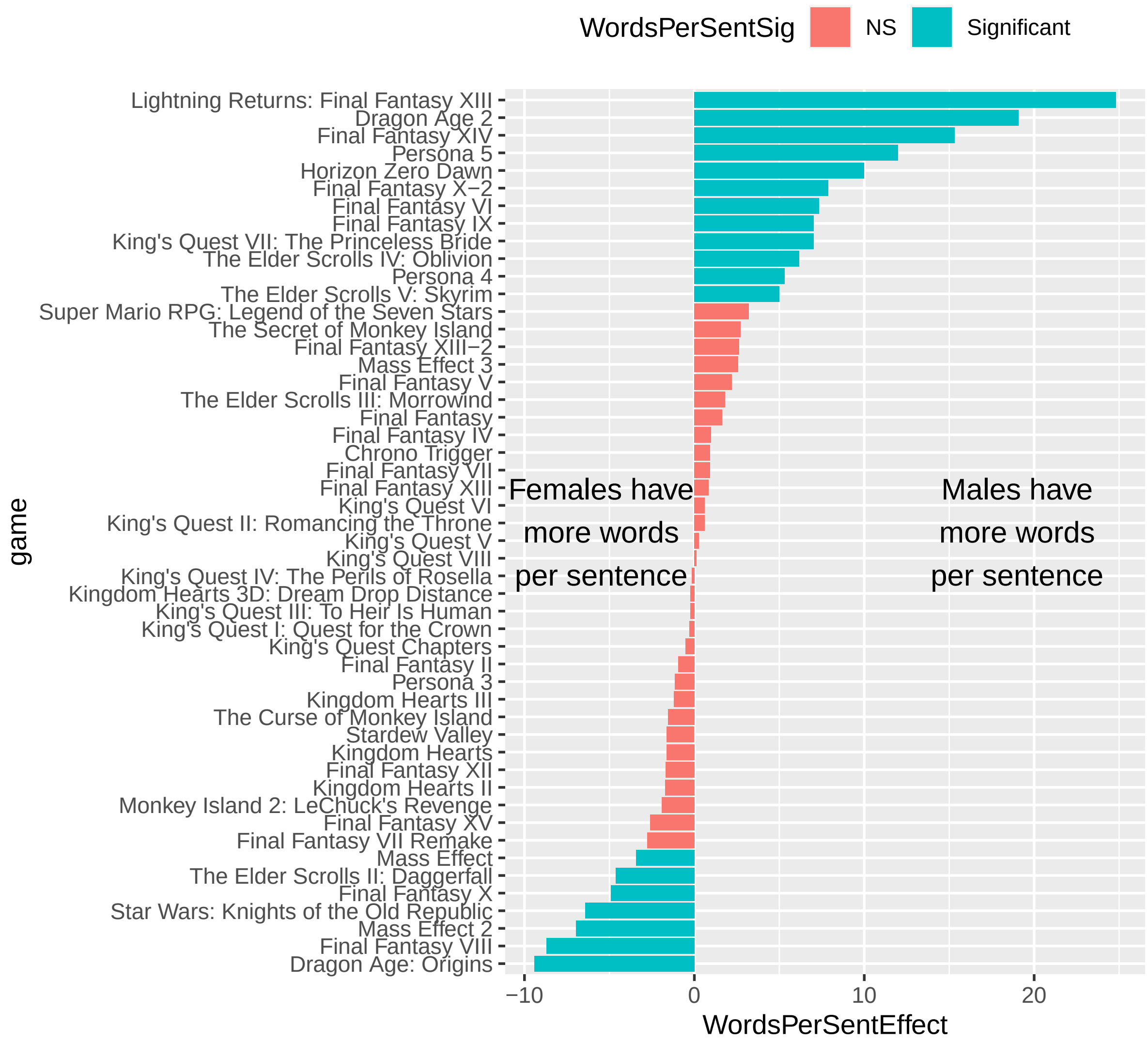
```
##
## Call:
## lm(formula = words ~ lines * group, data = x2[x2$group %in% c("male",
##     "female"), ])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -698.25  -60.89  -15.50   -1.02 1278.82
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      84.2100    39.9749   2.107   0.0374 *
## lines             6.7689     0.2051  32.996  < 2e-16 ***
## groupmale       -80.0665    46.6103  -1.718   0.0886 .
## lines:groupmale   5.9070     0.9604   6.151 1.25e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 188.4 on 111 degrees of freedom
## Multiple R-squared:  0.9248, Adjusted R-squared:  0.9227
## F-statistic: 454.7 on 3 and 111 DF,  p-value: < 2.2e-16
```

While the overall proportion of female dialogue is high, FFX-2 has large gender differences in the number of words per sentence. Female speak on average 3.4 words per sentence, and males speak 4.4. We note that the Dale-Chall Readability is also higher for males than females, which agrees with this.

We note that, across games, there is considerable variation in the gender differences in words per sentence:

```
d$WordsPerSentEffect = NA
d$WordsPerSentSig = NA
for(folder in folders){
  x = read.csv(paste0(folder,"stats_by_character.csv"),stringsAsFactors = F)
  if(nrow(x)>0){
    x = x[x$group %in% c("male","female"),]
    sx = summary(lm(words~sentences*group, data=x))
    t.val = sx$coefficients[4,3]
    sig = sx$coefficients[4,4]< (0.05/nrow(d))
    sig = c("NS","Significant")[sig+1]
    d[match(folder,d$folder),]$WordsPerSentEffect = t.val
    d[match(folder,d$folder),]$WordsPerSentSig = sig
  }
}
d$game = factor(d$game,levels=d$game[order(d$WordsPerSentEffect)])

a1 = grobTree(textGrob("Males have\nmore words\nper sentence", x=0.8,  y=0.5))
a2 = grobTree(textGrob("Females have\nmore words\nper sentence", x=0.15,  y=0.5))
ggplot(d,aes(x=WordsPerSentEffect,y=game,fill=WordsPerSentSig))+
  geom_bar(stat="identity") +
  annotation_custom(a1)+
  annotation_custom(a2)+
  theme(legend.position = "top")
```

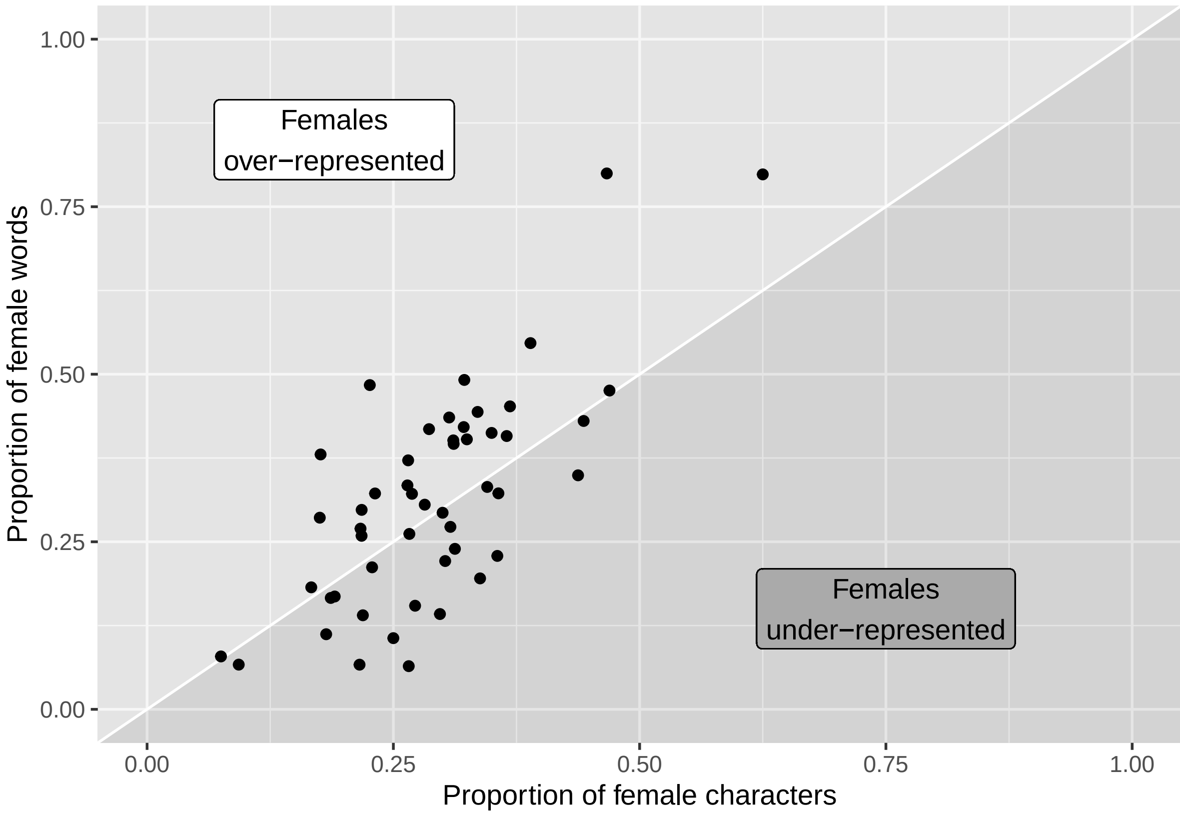


Proportion of characters

Relationship between dialogue proportions and character proportions:

```
poly = data.frame(
  x = c(-1,2,2),
  y = c(-1,-1,2),
  id = c(1,1,1))

bgColours = c("#458fc2","#ec0085")
bgColours = c("#000000","#AAAAAA")

wordsVCharacters = ggplot(d,aes(y=words,x=femCharProp)) +
  geom_polygon(data = data.frame(
    x = c(-1,2,2,-1,-1,2),
    y = c(-1,-1,2,-1,2,2),
    grp = c("a","a","a","b",'b','b')),
               aes(x = x, y = y,fill=grp),alpha=0.1)+
  scale_fill_manual(breaks=c('a','b'),values=bgColours)+
  coord_cartesian(ylim=c(0,1),xlim=c(0,1))+
  xlab("Proportion of female characters") +
  ylab("Proportion of female words") +
  geom_abline(intercept=0,slope=1,colour="white") +
```

```
  geom_label(label="Females\nover-represented",x=0.19,y=0.85) +
  geom_label(label="Females\nunder-represented",x=0.75,y=0.15, fill="#AAAAAA")+
  theme(legend.position = 'none')

wordsVCharacters + geom_point()
```




```
pdf("../results/graphs/WordsVsCharacters_noPoints.pdf",width=4,height = 3.5)
  wordsVCharacters + geom_point(alpha=0)
dev.off()
```

```
## pdf
##   2
```

```
pdf("../results/graphs/WordsVsCharacters.pdf", width=4,height = 3.5)
  wordsVCharacters + geom_point()
dev.off()
```

```
## pdf
##   2
```

There is a significant correlation between the proportion of female characters and the proportion of female dialogue:

```
cor.test(d$words,d$femCharProp)
```

```
##
##  Pearson's product-moment correlation
##
## data:  d$words and d$femCharProp
## t = 7.5226, df = 48, p-value = 1.168e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5749401 0.8416053
## sample estimates:
```

```
##       cor
## 0.7355718
```

Words per character

Is there a difference in the average amount of dialogue given to the average male character compared to the average female character? One confounding aspect of the data is that different games have different amounts of dialogue. So instead of raw words per character, for each game we calculate the average number of words given to a character *per 1000 words of dialogue*. That is, for every 1000 words of dialogue observed, how many go to a single male character on average? And how many to a single female character on average?

On average, a female character says 17.2 words per 1000 words of dialogue.

On average, a male character says 15.3 words per 1000 words of dialogue.

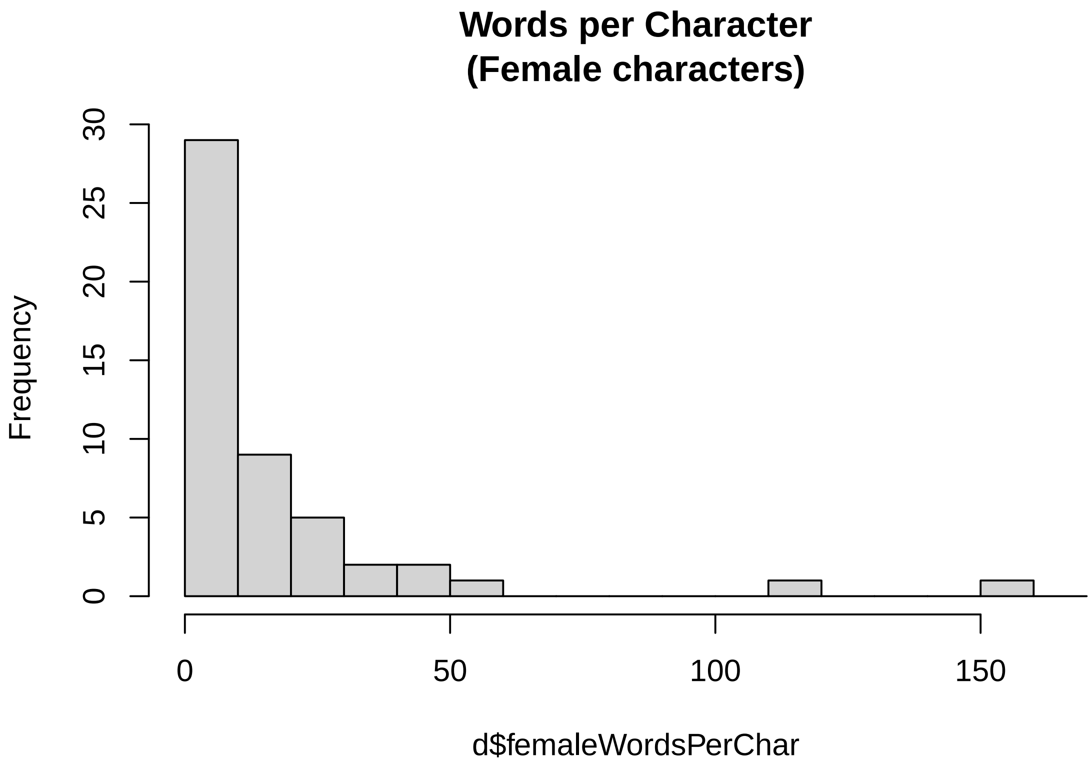
That is, the average female character says slightly more than the average male character. This difference is not significant:

```
t.test(d$femaleWordsPerChar,d$maleWordsPerChar,paired = T)
```

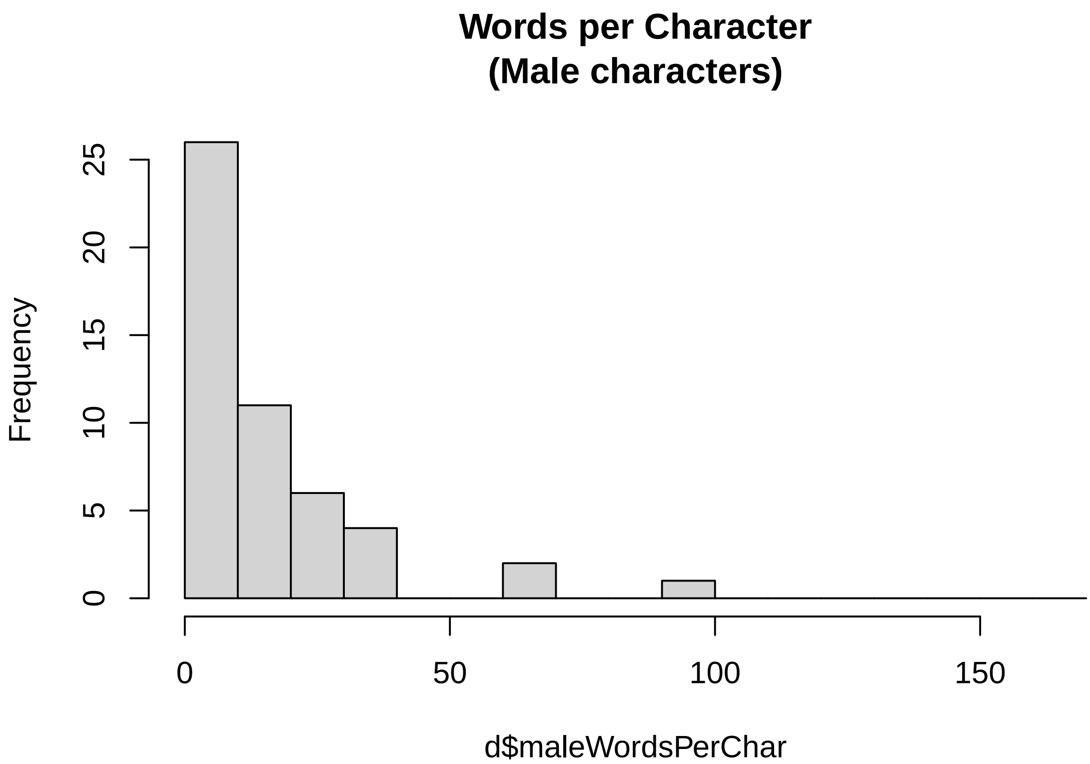
```
##
##  Paired t-test
##
## data:  d$femaleWordsPerChar and d$maleWordsPerChar
## t = 0.64566, df = 49, p-value = 0.5215
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.984454  7.756792
## sample estimates:
## mean of the differences
##                1.886169
```

Here is the distribution over games:

```
hist(d$femaleWordsPerChar, main="Words per Character\n(Female characters)",
     breaks = seq(0,175,by=10))
```




```
hist(d$maleWordsPerChar, main="Words per Character\n(Male characters)",
     breaks = seq(0,175,by=10))
```

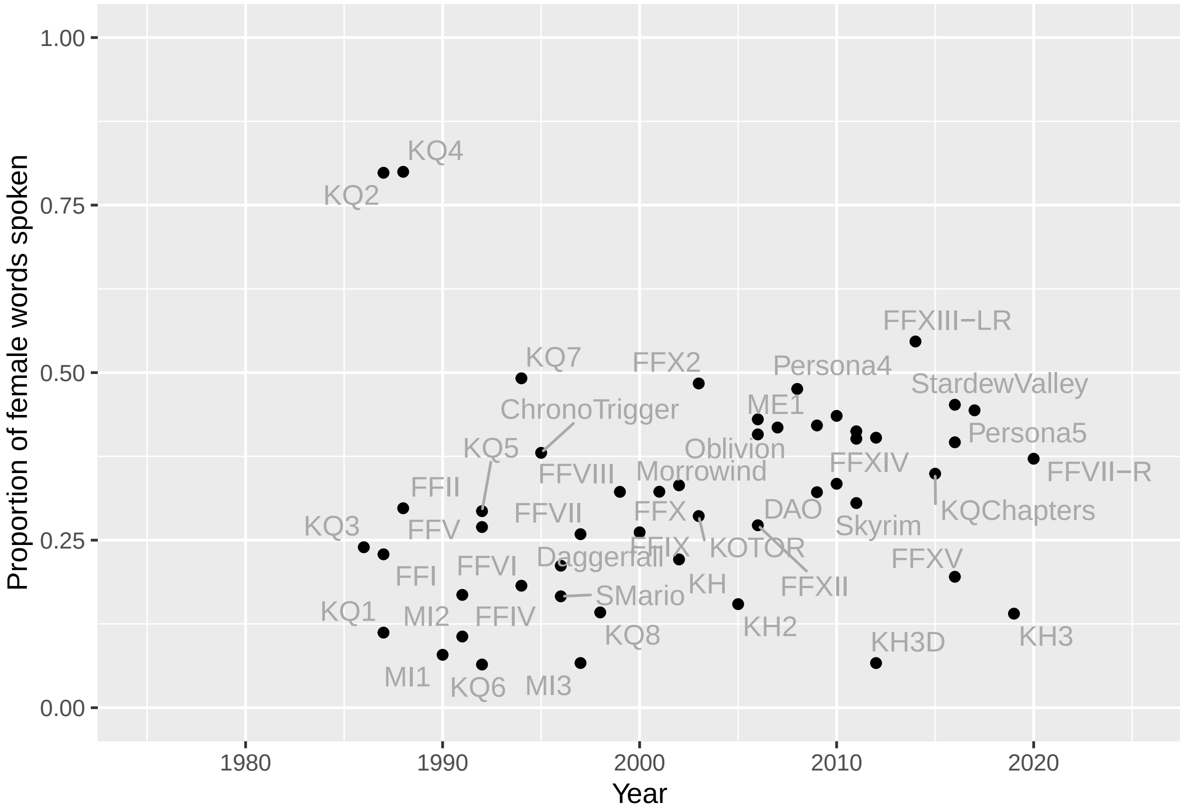


Change over time

This section tests whether the proportion of words spoken by female characters has changed over time. First we can visualise the data:

```
ggplot(d,
       aes(x=year,y=words)) +
  geom_point() +
  coord_cartesian(ylim=c(0,1),xlim=c(1975,2025))+
  ylab("Proportion of female words spoken") +
  xlab("Year") +
  geom_text_repel(aes(label=shortName),color="dark gray")
```

```
## Warning: ggrepel: 7 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



There are clearly two outliers: King's Quest 1 and King's Quest 2. These both have very little dialogue.

Test relationship between proportion of female characters and time:

```
cor.test(d$words,d$year)
```

```
##
##  Pearson's product-moment correlation
##
## data:  d$words and d$year
## t = 1.1528, df = 48, p-value = 0.2547
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1196790  0.4231517
## sample estimates:
##       cor
## 0.1641365
```

There is no significant relationship overall. However, removing the two outliers shows a significant positive correlation:

```
dNoOutliers = d[!d$shortName %in% c("KQ2","KQ4"),]
propFemaleOverTime = cor.test(dNoOutliers$words,dNoOutliers$year)
propFemaleOverTime
```

```
##
##  Pearson's product-moment correlation
##
## data:  dNoOutliers$words and dNoOutliers$year
## t = 3.6394, df = 46, p-value = 0.0006897
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.2179794 0.6673123
## sample estimates:
##       cor
## 0.4728292
```

Let's contextualise the linear relationship:

```
# Increase in prop female over time
pm = lm(words~year, data=dNoOutliers)
incPropFemalePerDecade = round((pm$coefficients['year']*100) * 10,1)
incPropFemalePerDecade
```

```
## year
##  6.3
```

```
# Year of parity
parity = ceiling((0.5 - pm$coefficients[1])/pm$coefficients[2])
```

The proportion of female dialogue is increasing by 6.3 percentage points per decade. If this rate continues, parity will be reached by 2036.

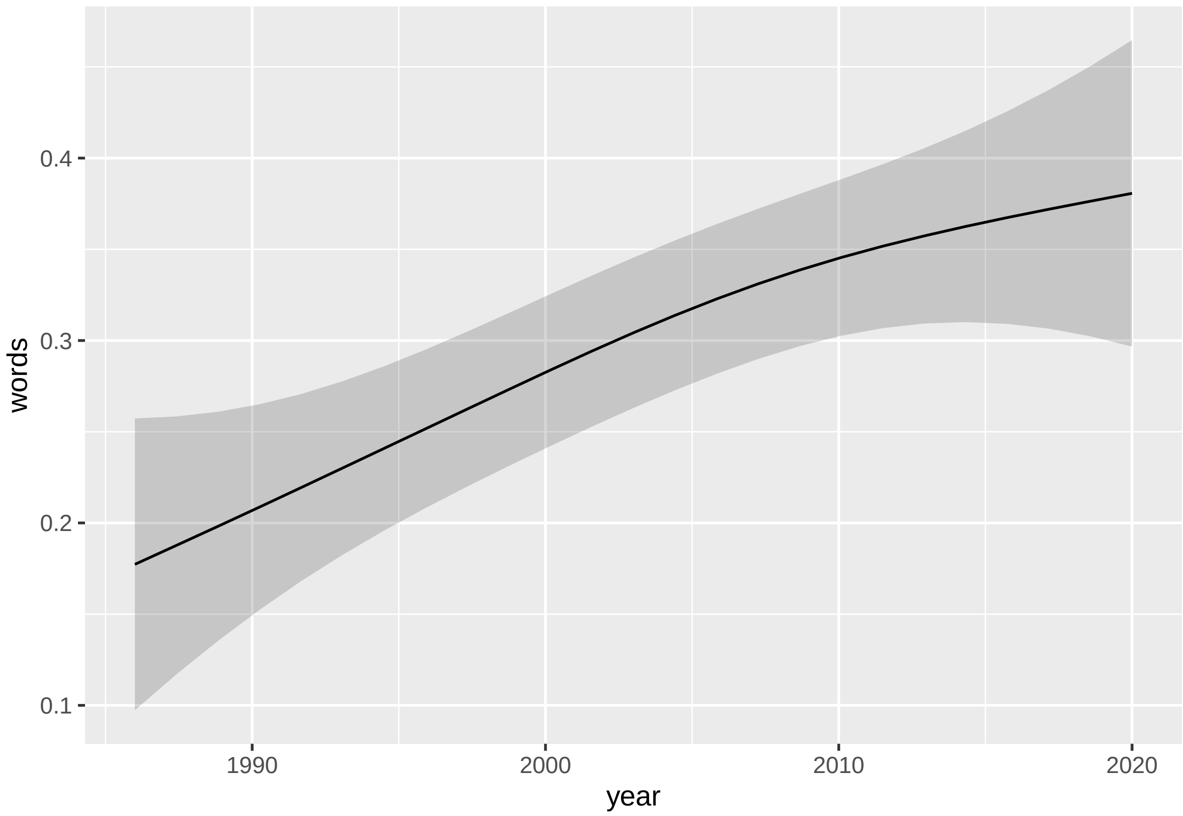
We can also test whether the change over time is non-linear, using a General Additive Model (GAM):

```
gam0 = bam(words ~ s(year), data=dNoOutliers)
# The EDF is a measure of non-linearity
summary(gam0)$edf
```

```
## [1] 1.520483
```

The EDF is slightly above 1, which is an indication of non-linearity. The GAM suggests that the increase in female proportions is slowing down, with the average seeming to plateau around 40%.

```
plot_smooths(gam0,year)
```



However, the GAM does not significantly improve the fit of the model compared to the linear model:

```
lrtest(pm,gam0)
```

```
## Likelihood ratio test
##
## Model 1: words ~ year
## Model 2: words ~ s(year)
##      #Df LogLik      Df  Chisq Pr(>Chisq)
## 1 3.0000 36.652
## 2 3.8797 37.312 0.87972 1.3207     0.2505
```

Therefore we prefer the linear model.

Write some stats for the paper:

```
propFemaleOverTimeStr = paste0(
  "$r$=",round(propFemaleOverTime$estimate,2),
  " [",round(propFemaleOverTime$conf.int[1],2),",",
    round(propFemaleOverTime$conf.int[2],2),"]",
  ", $n$=",nrow(d),
  ", $p$=",round(propFemaleOverTime$p.value,3))
cat(propFemaleOverTimeStr, file="../results/latexStats/propFemaleOverTime.tex")
cat(incPropFemalePerDecade, file="../results/latexStats/incPropFemalePerDecade.tex")
cat(min(d$year), file="../results/latexStats/earliestYear.tex")
cat(max(d$year), file="../results/latexStats/latestYear.tex")
cat(parity, file="../results/latexStats/yearOfParity.tex")
```

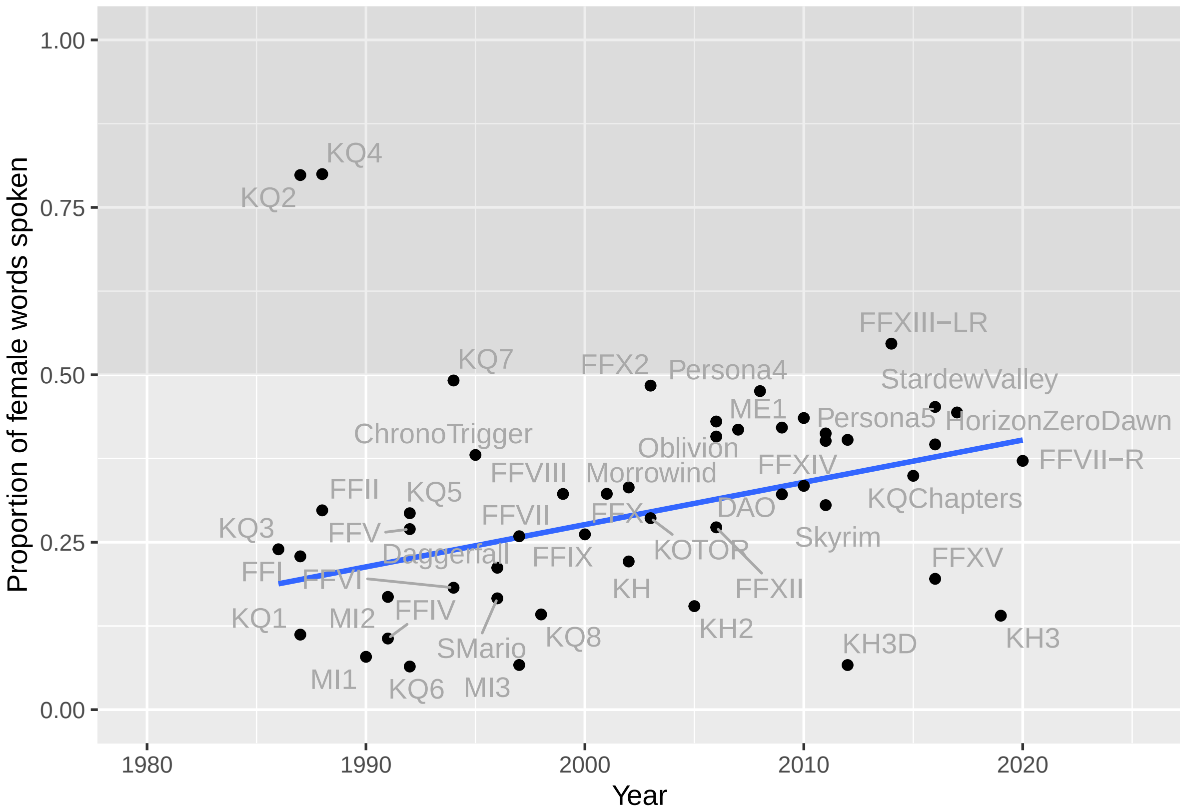
Plot for paper:

```
changeOverTime = ggplot(d,
       aes(x=year,y=words)) +
  annotate("rect", xmin = 0, xmax = 3000,
           ymin = 0.5, ymax = 1.5, alpha = .1) +
```

```
#  annotate("rect", xmin = 0, xmax = 3000,
#            ymin = 0.5, ymax = 1.5, alpha = .1,fill="#ec0085") +
#  annotate("rect", xmin = 0, xmax = 3000,
#            ymin = -0.5, ymax = 0.5, alpha = .1,fill="#458fc2") +
  stat_smooth(method=lm, se = F,data = dNoOutliers) +
  geom_point() +
  coord_cartesian(ylim=c(0,1),xlim=c(1980,2025))+
  ylab("Proportion of female words spoken") +
  xlab("Year") +
  geom_text_repel(aes(label=shortName),color="dark gray")
changeOverTime
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: ggrepel: 6 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```




```
pdf('../results/graphs/ChangeOverTime.pdf', width=8,height=6)
changeOverTime
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
dev.off()
```

```
## pdf
##   2
```

Change in proportions of female characters

There is no significant correlation in the total data:

```r
cor.test(d$femCharProp, d$year)
```

```
##
##  Pearson's product-moment correlation
##
## data:  d$femCharProp and d$year
## t = 1.3362, df = 48, p-value = 0.1878
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.09392719  0.44429971
## sample estimates:
##       cor
## 0.1893715
```

But removing one early outlier (King's Quest II) shows that there is a positive trend:

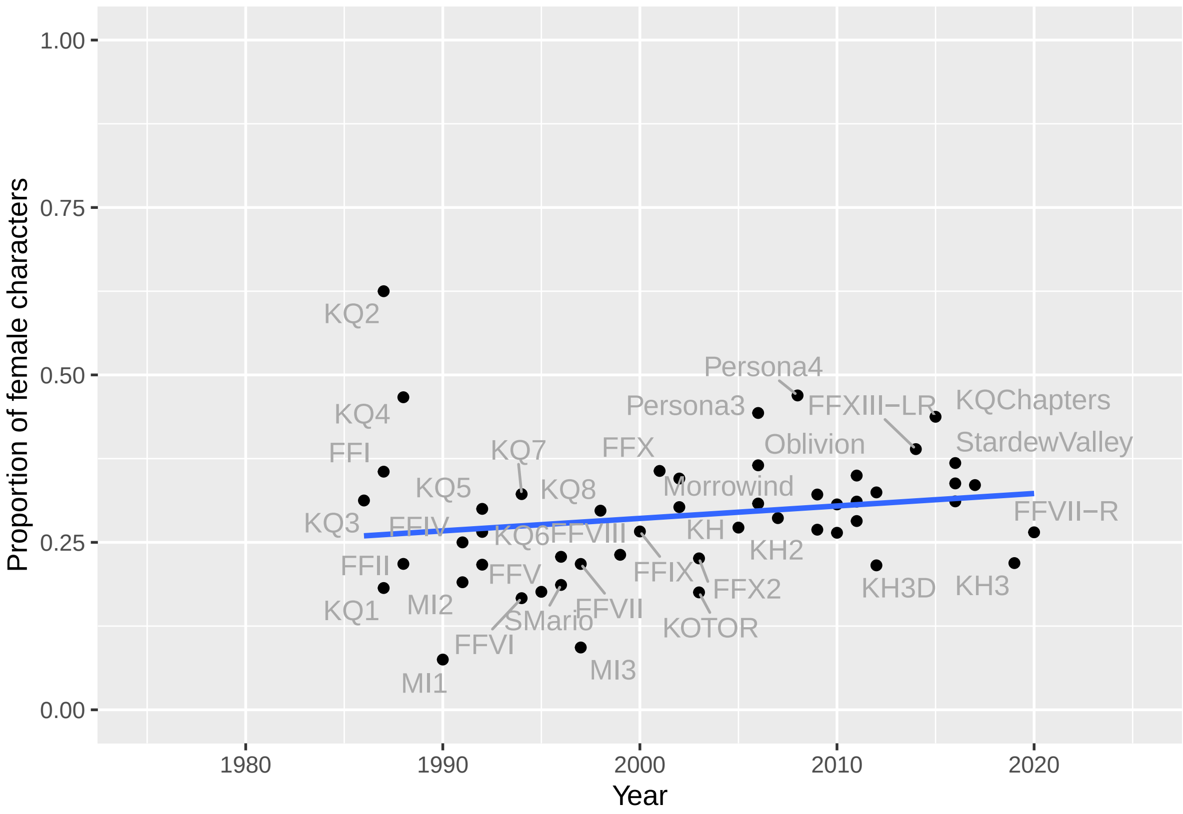
```r
dx = d[d$game != "King's Quest II: Romancing the Throne",]
cor.test(dx$femCharProp, dx$year)
```

```
##
##  Pearson's product-moment correlation
##
## data:  dx$femCharProp and dx$year
## t = 2.5814, df = 47, p-value = 0.01302
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.07901041 0.57645514
## sample estimates:
##       cor
## 0.3523781
```

```r
ggplot(d,
       aes(x=year,y=femCharProp)) +
  geom_point() +
  stat_smooth(method=lm, se = F) +
  coord_cartesian(ylim=c(0,1),xlim=c(1975,2025))+
  ylab("Proportion of female characters") +
  xlab("Year") +
  geom_text_repel(aes(label=shortName),color="dark gray")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: ggrepel: 15 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```


Proportion of female characters
1.00
0.75
0.50
0.25
0.00
1980
1990
2000
2010
2020
Year
KQ2
Persona4
Persona3
FFXIII−LR
KQChapters
KQ4
Oblivion
StardewValley
FFI
KQ7
FFX
KQ5
KQ8
Morrowind
KQ3
FFIV
KQ6
FFVIII
KH
FFVII−R
KH2
FFII
FFIX
FFV
FFX2
KH3D
KH3
KQ1
MI2
FFVII
SMario
KOTOR
FFVI
MI3
MI1

Assessing gender bias with random baselines

This section analyses the gender balance in the dialogue of the games in the corpus. It begins with an illustration of the statistical methods applied to *Final Fantasy VII*.

The proportion of dialogue by gender is assessed in comparison to two "baselines". These reflect two possible sources of bias in the amount of dialogue spoken by each gender:

- How many characters of each gender there are (i.e. are there more male than female characters).
- How much dialogue each character is given (i.e. are males given more dialogue than females).

Load libraries

```
library(vcd)
library(MASS)
library(ggplot2)
library(ggrepel)
```

Demonstration: Gender differences in Final Fantasy VII

Let's load the data for Final Fantasy VII:

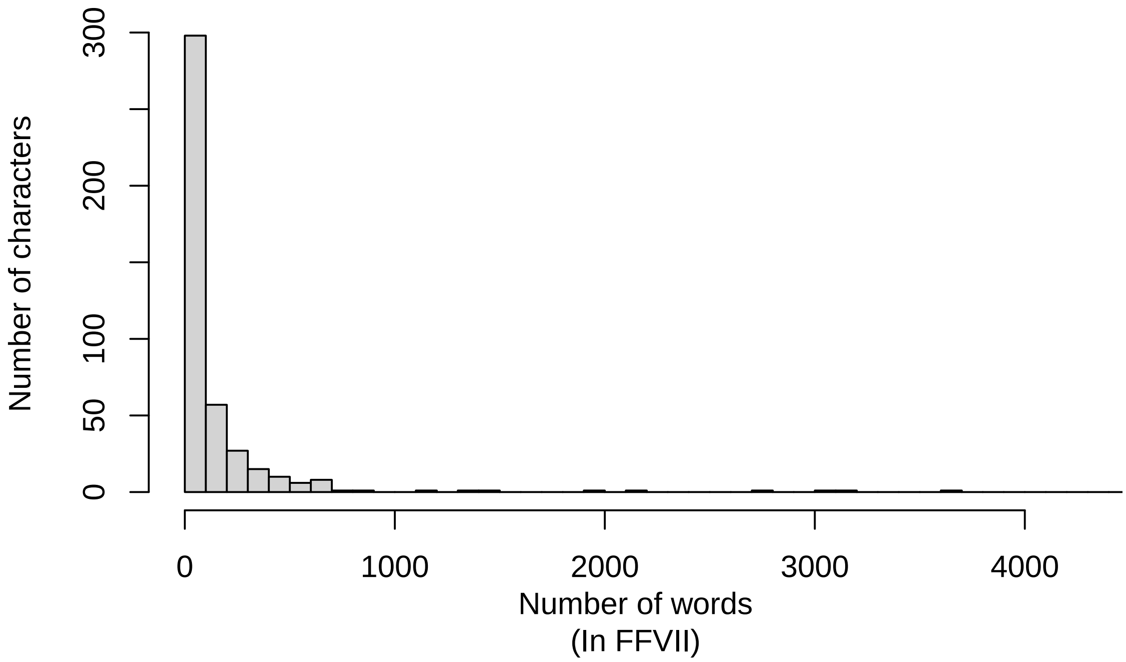
```
d = read.csv("../data/FinalFantasy/FFVII/stats_by_character.csv",stringsAsFactors = F)
cor(d[d$group=="female",]$words,d[d$group=="female",]$lines)
```

```
## [1] 0.9940972
```

Distribution of words by character

What does the distribution of lines per character look like? Here is a plot, showing the number of words of dialogue for each character. The shape of how these values are distributed across the scale is called the "distribution".

```
lengthVar = "words"
# estimate the parameters
distr = d[,lengthVar]
#plot(1:length(distr),sort(distr),xlab="Character",ylab="Words of dialogue")
hist(distr, freq = TRUE, breaks = 100, xlim = c(0, quantile(distr, 0.99)),
     ylab="Number of characters",
     xlab="Number of words\n(In FFVII)", main="")
```



The distribution is very skewed: a small number of characters have a lot of dialogue, and a large number of characters have a small amount of dialogue. For example, 297 characters have less than 100 words of dialogue each, while only 8 characters have over 3,000 words of dialogue each.

We can formally test whether the distribution fits an ideal "normal" (bell-curve), an exponential distribution or a t-distribution:

```
# goodness of fit test
normFit = fitdistr(distr,"normal")
ks.test(distr, "pnorm", normFit$estimate)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  distr
## D = 0.96568, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
expFit = fitdistr(distr, "exponential")
ks.test(distr, "pexp", expFit$estimate)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  distr
## D = 0.34796, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
tFit = fitdistr(distr, "t")
ks.test(distr, "pt", tFit$estimate)
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
```

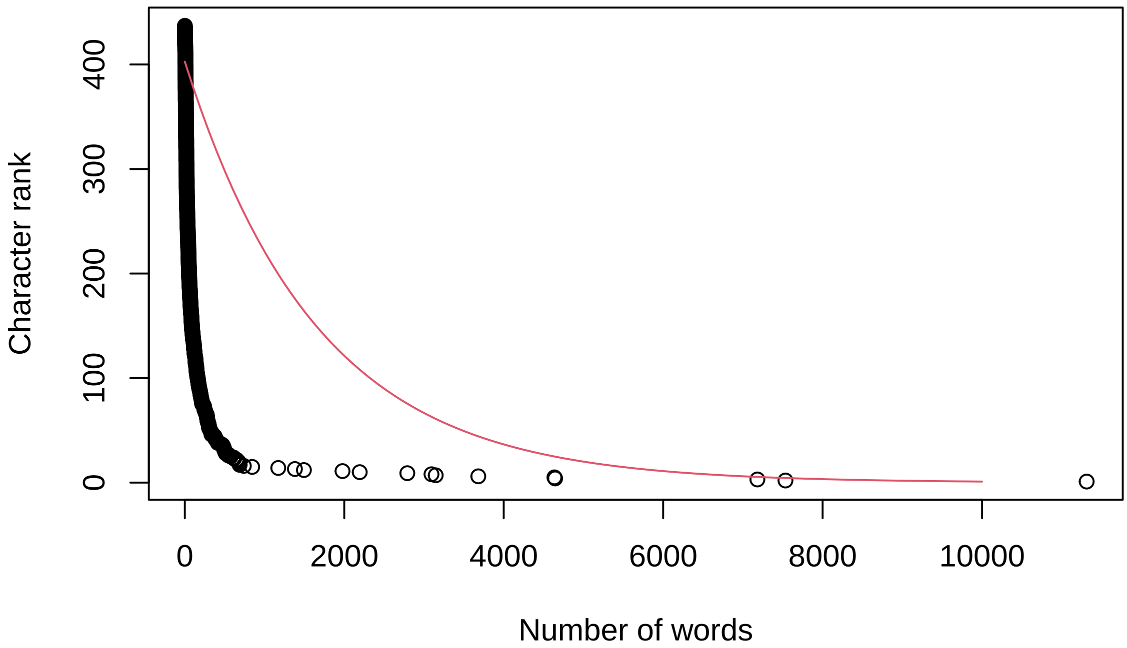
```
## data:  distr
## D = 0.9764, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

```
# plot a graph
#hist(distr, freq = FALSE, breaks = 100, xlim = c(0, quantile(distr, 0.99)))
#curve(dexp(x, rate = expFit$estimate), from = 0, col = "red", add = TRUE)
```

We see that the dialogue by character distribution does not fit any of these ideal distributions. This means that using standard tests such as the t-test is not appropriate in order to compare the distributions (e.g. for testing if the average number of words given to male characters is higher than the average number of words given to female characters).

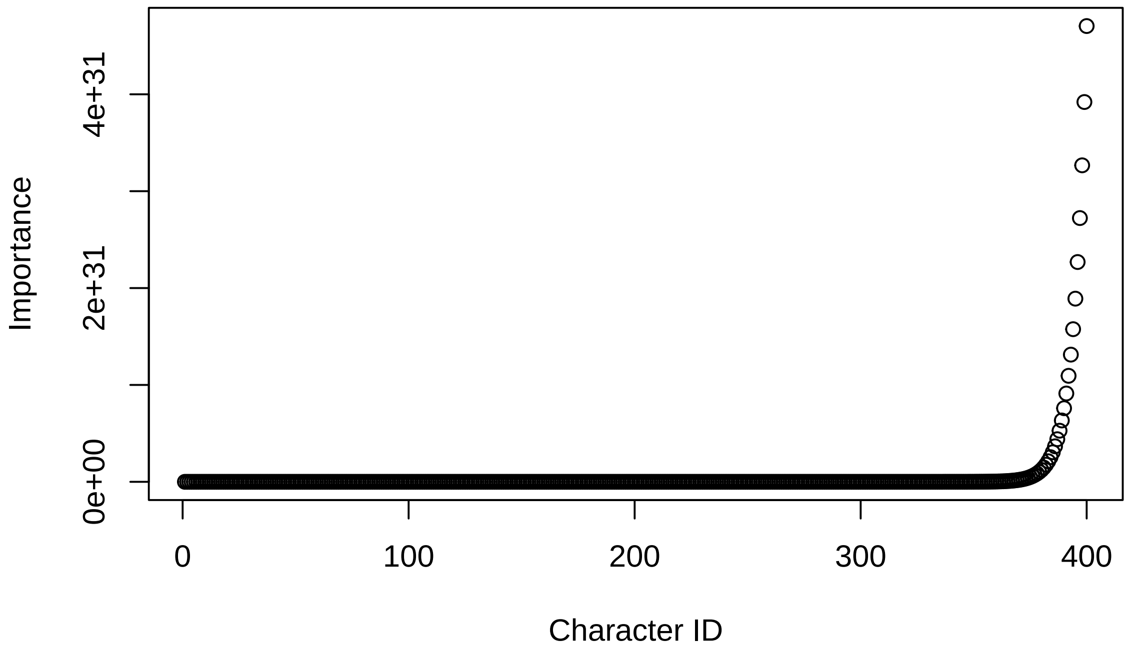
We can see the departure from an exponential curve by looking at the number of words spoken by a character compared to their rank (black circles): there is a very sharp decrease in characters, then a smaller set of 'main' characters. The turning point is a kind of "elbow" shape. The plot below includes a true exponential line in red for reference:

```
plot(1:nrow(d)~sort(d$words,decreasing = T),
     xlab="Number of words",ylab="Character rank")
# Exponential line
ys = seq(0,10000,length.out = 100)
points(ys,rev(1.0006^(ys)),type='l',col=2)
```



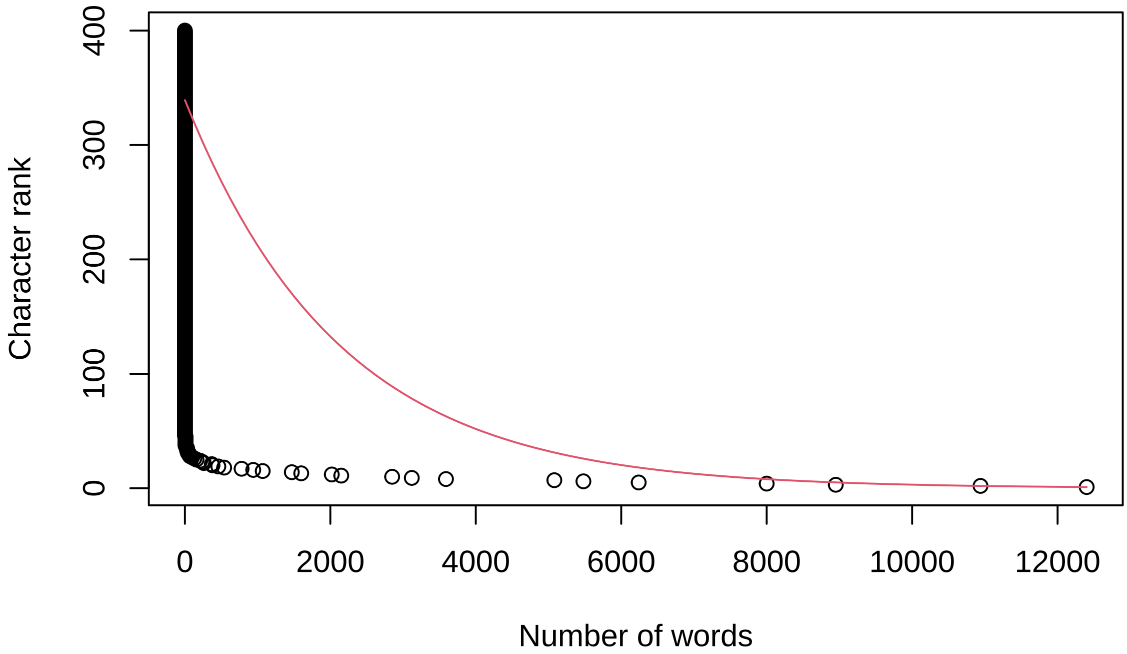
The "elbow" distribution we see in the data can be generated using the following method: Assume that we have 400 characters, and that the distribution of their importance to the plot is exponential (a few characters are much more likely to be chosen to speak than the majority)

```
nChar = 400
# Exponential character importance
charImportance = 1.2^(1:nChar)
plot(charImportance,xlab="Character ID",ylab="Importance")
```



Now we can simulate a conversation: We pick one character at random, but weighted in proportion to their importance (we're more likely to choose a more important character). Now pick one other character (that is not the first character) in the same way. They each say 10 words. We run this simple simulation many times, and show that it generates the same kind of "elbow" (black dots). A true exponential line is shown in red for comparison. With comparable settings for FFVII, it also generates roughly the same number of words that the most prolific character speaks.

```
simulateConversation = function(){
  chars = 1:nChar
  char1 = sample(chars,1,prob = charImportance)
  char2 = sample(chars[chars!=char1],1,prob = charImportance[chars!=char1])
  # 10 words each
  10* ((1:nChar) %in% c(char1,char2))
}

# 8000 lines in FFVII, which is 4000 pairs of lines
res = replicate(4000,simulateConversation())
distrSim = rowSums(res)

plot(1:length(distrSim)~sort(distrSim,decreasing = T),
     xlab="Number of words",ylab="Character rank")
# Exponential line
ys = seq(1,max(distrSim),length.out = 100)
points(ys,rev(1.00047^(ys)),type='l',col=2)
```



So, the distribution of words by characters may be a product of conversations being *interactive*: at least two speakers are required to have a conversation.

Difference in words per gender

Below we calculate the total number of words for male and female characters. There are other gender groups, but we'll focus just on the comparison between male and female.

```
dx = d[d$group %in% c("male","female"),]
totalWords = tapply(dx[,lengthVar],dx$group,sum)
cbind(words=totalWords,proprtion=totalWords/sum(totalWords))
```

```
##        words proprtion
## female 25102 0.2589437
## male   71838 0.7410563
```

Total number of male female characters.

```
totalCharacters = table(dx$group)
cbind(characters=totalCharacters, proportion = prop.table(totalCharacters))
```

```
##        characters proportion
## female         81  0.2177419
## male          291  0.7822581
```

It's clear that the proportion of male dialogue is to some extent affected by the proportion of male characters. In order to tell whether there is a bias for men to speak more than women independently of the number of male and female characters, we need to compare the true difference in the data to a "baseline". Baselines are a calculation of the range of differences we would expect in particular conditions, such as there being no bias in dialogue length by gender.

Therefore, we calculate the true difference in total number of words for male and female characters in our data. A positive number indicates more words for males, expressed in number of words:

```
true_diffInWords = diff(totalWords)
true_diffInWords
```

```
##  male
## 46736
```

Baseline 1: Randomly assigned gender

In order to contextualise the difference between the amount of male and female dialogue, we need to compare it to a baseline. We use two baselines in this study: In the first baseline, we hold fixed which lines belong to which characters, and randomly assign them a gender (creating roughly 50% female characters). In the second, we randomly shuffle the genders (preserving the proportion of characters). This is explained as follows:

We can simulate what would happen if the gender of each character was assigned randomly, with an even chance of each character being male or female. This can lead to a different number of male and female characters compared to the true game. A typical simulation would have roughly 50% male and 50% female. We know that this baseline is **unbiased**, at least in the sense that there is an equal probability of any character being male or female.

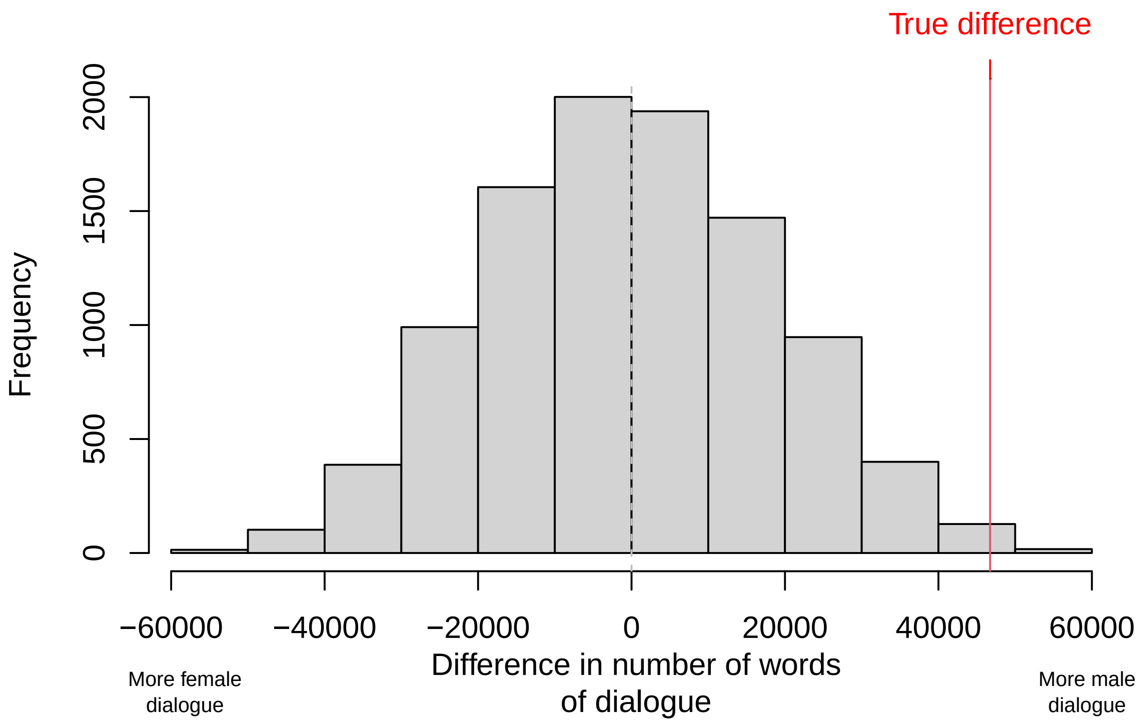
In the simulation, we assign gender randomly then work out the difference in the total number of words between male and female characters. We run this simulation 10,000 times to get lots of values for this measure (a "distribution").

```
randomGender = function(words,groups){
  diff(tapply(
        words,
        sample(unique(groups),length(words),replace = T),
        sum))
}
set.seed(451)
random_diffInWords = replicate(10000,
                         randomGender(dx[,lengthVar],dx$group))
```

So we now have two types of data: the true difference between the amount of male and female dialogue that we observed in the data. We'll call this the "true gender bias". And we also have lots of estimates of the difference between the amount of male and female dialogue if the gender was assigned randomly. This forms our baseline, and we'll call this the "random distribution".

We can now visualize these values: the random distribution is visualised below as a histogram, and the true gender bias is marked with a red line. If the true gender bias in dialogue is not different from the random distribution, then the true bias should lie in the middle of the random distribution. That is, it could have plausibly been generated by randomly assigning genders. However, in the graph below we see that the true gender bias lies at one extreme of the random distribution. This indicates that it is unlikely that the gender bias we see in the real data was generated by randomly assigning genders. That is, the true gender bias is very different from the baseline.

```
hist.compareDist = function(perm,trueVal){
  pmin = min(c(trueVal,perm,0))
  pmax = max(c(trueVal,perm,0))
  hist(perm, xlim=c(pmin,pmax),
       xlab="Difference in number of words\nof dialogue",
       main="")
  abline(v=0,col='gray',lty=2)
  axis(1,c(pmin,pmax),
       c("More female\ndialogue","More male\ndialogue"),
       tick=F,line=2,cex.axis=0.7)
  abline(v=trueVal,col=2)
  axis(3,trueVal,"True difference",col.axis="red",col.ticks="red")
}
hist.compareDist(random_diffInWords,true_diffInWords)
```



We can quantify the difference between the true bias and the random distribution. The first measure is the z-score, which indicates how far away the true difference is from the mean, expressed in the number of standard deviations. The second measure is a p-value, which indicates the proportion of simulations that resulted in a more extreme measure of difference than the true value.

```
zstats = function(true,dist){
  zscore = (true-mean(dist)) / sd(dist)
  pvalue = 1/length(dist)
  numAgainst = sum(true < dist)
  if(numAgainst>0){
    pvalue = numAgainst/ length(dist)
  }
  return(c(zscore=zscore,p=pvalue))
}
zstats(true_diffInWords,random_diffInWords)
```

```
## zscore.male           p
##    2.531894    0.003500
```

The p-value is less than 0.05, suggesting that there is less than 5% chance of observing the true bias in a simulation where gender is assigned randomly. That is, the true gender bias is unlikely to have been generated by an unbiased process. Put another way, the difference between male and female dialogue is significant in comparison to this baseline.

Baseline 2: Permuted gender

We can also specify a different baseline that takes into account the proportion of male and female characters. This can be done by randomly swapping the gender of characters. The difference from the first baseline is as follows: Imagine some real data where the characters were composed of 30% female characters and 70% male characters. The first baseline would change that proportion to roughly 50%/50%. In contrast, the simulation for the second baseline preserves the same proportion (30% female, 70% male), but with a different assignment of which characters are male and which are female.

There's another way to think about this second baseline that is an equivalent interpretation. Although the implementation below randomly permutes the gender values, it's effectively equivalent to permuting the values of the dialogue. That is, the process effectively randomly swaps all of one character's lines for all of another character's lines. For example, all of (main protagonist) Cloud's lines are now spoken by (party member) Aerith, and all of Aerith's lines are now spoken by (minor character) "Shopkeeper (Items Sector 7)".

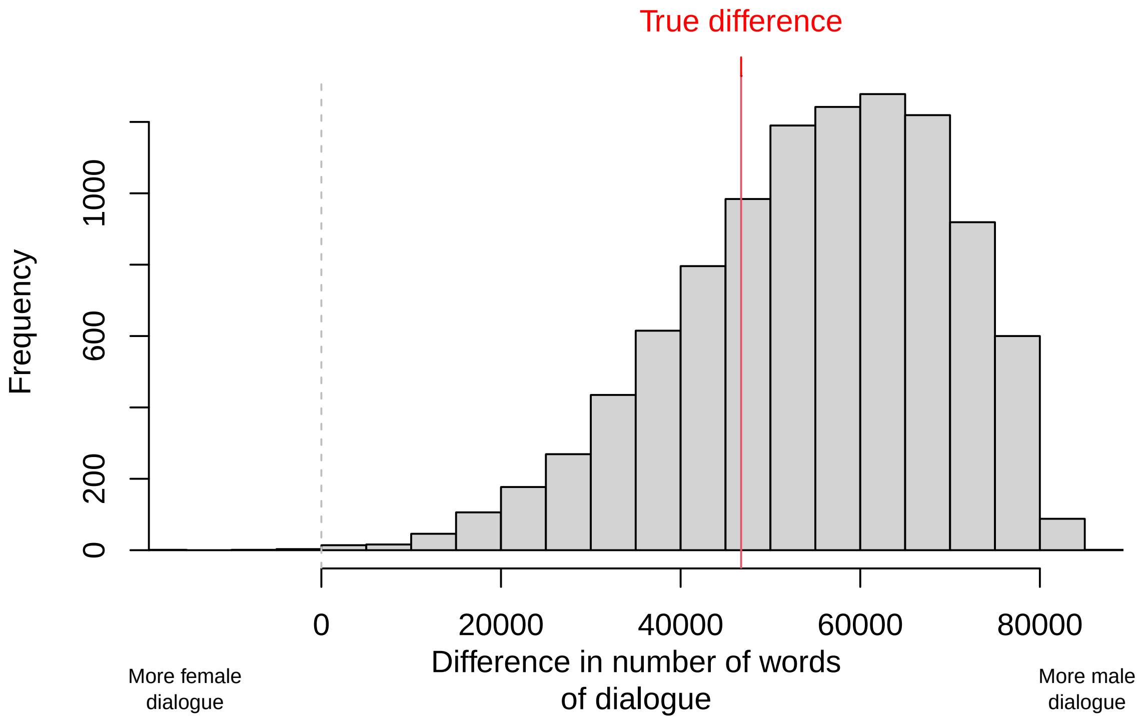
If there was no gender bias in the amount of dialogue written for each character, then this random swapping would not produce a big change in the calculation of the difference between genders. So the baseline is unbiased in the sense that we know that there's no bias in the amount of dialogue given to any particular character.

The function below works out the difference in the sum of words between the two gender groups, but where the assignment of characters to gender groups has been randomly permuted (sampled). This preserves the same number of male and female characters as in the true data.

```
permuteGender = function(words,group){
  diff(tapply(words,sample(group),sum))
}
set.seed(451)
permuted_diffInWords = replicate(10000,
                          permuteGender(dx[,lengthVar],dx$group))
```

We'll call this second baseline the "permuted distribution". We can now visualise the permuted distribution from these second baseline simulations, and compare it to the true gender bias This time, we see that the true gender bias lies in the middle of the distribution:

```
hist.compareDist(permuted_diffInWords,true_diffInWords)
```



And the stats also suggest that the true difference is reasonably likely to be generated by the permuted simulation. That is, it's plausible that the true data was generated by a process that had no bias for the amount of dialogue given to a character based on their gender.

```
zstats(true_diffInWords,permuted_diffInWords)
```

```
## zscore.male          p
##  -0.5441816   0.7197000
```

Summary

We discovered a few things about Final Fantasy VII:

- The distribution of dialogue by character does not approximate common distributions, preventing straightforward statistical tests.
- The difference in the number of words of dialogue between male and female characters is biased. There are more words spoken by male characters compared to a game where the gender of each character was randomly determined.
- However, there is no dialogue bias in Final Fantasy VII. That is, the proportion of dialogue reflects the proportion of male and female characters, and there's no evidence that the average female character is given fewer lines than the average male character.

In principle, the two types of test are independent: a game might have a small or large proportion of female characters, and those characters might have more or less to say than male characters.

Applying the baselines to all games

We can now apply the methods above to all the games in the corpus. These calculations were run offline in the script `compareToBaselines.R`.

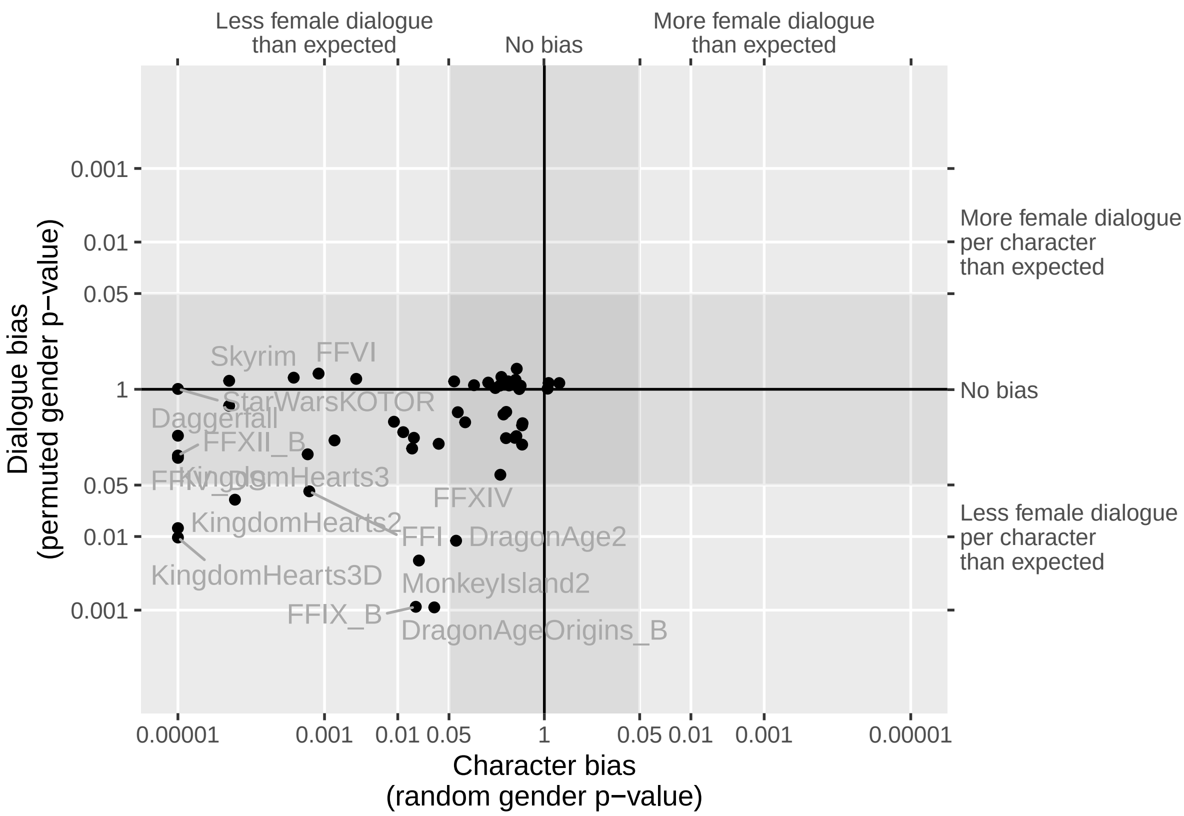
Some games have much larger scripts than others. So for the test that compares the entire corpus as a whole, the length of dialogue was transformed to words-per-thousand within a game.

```
all = read.csv("../results/compareToBaseline.csv",stringsAsFactors = F)
all = all[!is.na(all$p.perm),]
all = all[all$game!="Test",]
all = all[!all$alternativeMeasure,]
```

Create a bias plot for the publication:

```
all$p.random.log = log10(all$p.random)
all$p.perm.log = log10(all$p.perm)

all$p.random.log[all$z.random<0] = -all$p.random.log[all$z.random<0]
all$p.perm.log[all$z.perm<0] = -all$p.perm.log[all$z.perm<0]

threshold = log10(0.05)

options(scipen=999)
xs = c(0.05,0.01,0.001,0.00001)
xs2 = c(rev(xs),1,xs)
ls = c(log10(rev(xs)),0,-log10(xs))

gx = ggplot(all, aes(x=p.random.log,y=p.perm.log)) +
  annotate("rect", xmin = -1000, xmax = 1000, ymin = -threshold, ymax = threshold, alpha = .1) +
  annotate("rect", xmin = -threshold, xmax = threshold, ymin = -1000, ymax = 1000, alpha = .1) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  geom_text_repel(aes(label=shortName),color="dark gray",force = 1) +
  coord_cartesian(ylim=c(-4,4),xlim=c(-5,5)) +
  scale_x_continuous(breaks=ls,labels=xs2,
                     sec.axis = sec_axis(~.*1,
                        breaks = ls,
                        labels=c("","Less female dialogue\nthan expected","","",
                                 "No bias",
                                 "","",
                                 "More female dialogue\nthan expected",""))) +
  scale_y_continuous(breaks=ls,labels=xs2,
                     sec.axis = sec_axis(~.*1,
                        breaks = ls,
                        labels=c("","","Less female dialogue\nper character\nthan expected","",
                                 "No bias","",
                                 "More female dialogue\nper character\nthan expected","",""))) +
  theme(panel.grid.minor = element_blank()) +
  xlab("Character bias\n(random gender p-value)") +
  ylab("Dialogue bias\n(permuted gender p-value)")
gx
```

```
## Warning: ggrepel: 35 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



Write to file

```
pdf("../results/graphs/CompareToBaseline.pdf",width=9.5,height=7)
gx
```

```
## Warning: ggrepel: 28 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
dev.off()
```

```
## pdf
##   2
```

The graph above shows the p-values for the random baseline compared to the permuted baseline. Values nearer the center are not significantly different from the baseline. The gray rectangles show the threshold of $p < 0.05$.

Full results for the comparison to the random baseline:

```
knitr::kable(all[,c("game","z.random","p.random")])
```

	game	z.random	p.random
1	Chrono Trigger	0.4712166	0.49771
2	Dragon Age 2	1.0980156	0.06246
5	Dragon Age: Origins	2.8100350	0.03148
6	The Elder Scrolls II: Daggerfall	6.4962123	0.00001
7	The Elder Scrolls III: Morrowind	0.6609665	0.50423
8	The Elder Scrolls IV: Oblivion	1.5651459	0.05877
9	The Elder Scrolls V: Skyrim	3.2730719	0.00038
10	Final Fantasy	2.5829984	0.00062
11	Final Fantasy II	0.7891011	0.39550
14	Final Fantasy IV	3.3840605	0.00006
16	Final Fantasy IX	3.6726997	0.01761
17	Final Fantasy V	0.8898893	0.29979

	game	z.random	p.random
18	Final Fantasy VI	2.9433827	0.00083
21	Final Fantasy VII	2.5199998	0.00271
22	Final Fantasy VII Remake	1.2575180	0.10966
23	Final Fantasy VIII	0.6822423	0.41525
25	Final Fantasy X	2.1648299	0.00887
26	Final Fantasy X-2	0.1337843	0.45468
28	Final Fantasy XII	6.1437520	0.00001
29	Final Fantasy XIII	0.4819298	0.31968
30	Final Fantasy XIII-2	0.6035257	0.25938
31	Lightning Returns: Final Fantasy XIII	-0.2500463	0.62486
32	Final Fantasy XIV	0.9227224	0.25101
33	Final Fantasy XV	2.0950667	0.01570
34	Horizon Zero Dawn	0.4030046	0.40493
36	Kingdom Hearts	2.7962376	0.00059
37	Kingdom Hearts II	3.4460522	0.00001
38	Kingdom Hearts III	3.3153632	0.00001
39	Kingdom Hearts 3D: Dream Drop Distance	3.4228924	0.00001
40	King's Quest I: Quest for the Crown	2.0820536	0.01660
41	King's Quest II: Romancing the Throne	-1.2977746	0.87385
42	King's Quest III: To Heir Is Human	1.4005520	0.08311
43	King's Quest IV: The Perils of Rosella	-1.3980067	0.90203
44	King's Quest V	1.5237709	0.06591
45	King's Quest VI	1.3731348	0.00137
46	King's Quest VII: The Princeless Bride	0.0578734	0.47490
47	King's Quest VIII	1.4145117	0.01187
48	King's Quest Chapters	0.8082777	0.30192
50	Mass Effect	0.7143946	0.24692
51	Mass Effect 2	0.8067654	0.21425
54	Mass Effect 3	0.3878197	0.49747
56	Monkey Island 2: LeChuck's Revenge	2.5432044	0.01942
57	The Curse of Monkey Island	2.0836895	0.00005
58	The Secret of Monkey Island	2.4678511	0.00005
59	Persona 3	0.6137222	0.27697
60	Persona 4	0.2201780	0.42006
62	Persona 5	0.9781966	0.17144
63	Stardew Valley	0.4491905	0.32962
64	Star Wars: Knights of the Old Republic	3.5698348	0.00001
65	Super Mario RPG: Legend of the Seven Stars	1.2998554	0.03616

Full results for the comparison to the permuted baseline:

```
knitr::kable(all[,c("game","z.perm","p.perm")])
```

	game	z.perm	p.perm
1	Chrono Trigger	0.6102605	0.17736
2	Dragon Age 2	1.6960572	0.00877
5	Dragon Age: Origins	3.5510906	0.00109
6	The Elder Scrolls II: Daggerfall	0.7457174	0.23421
7	The Elder Scrolls III: Morrowind	0.7005772	0.34632
8	The Elder Scrolls IV: Oblivion	-0.7765375	0.78045
9	The Elder Scrolls V: Skyrim	-0.4950374	0.69546
10	Final Fantasy	1.4448200	0.04117
11	Final Fantasy II	0.9485726	0.21896
14	Final Fantasy IV	1.6481142	0.03165
16	Final Fantasy IX	4.3839283	0.00111
17	Final Fantasy V	1.0615415	0.21703

	game	z.perm	p.perm
18	Final Fantasy VI	-0.2238571	0.61135
21	Final Fantasy VII	-0.5420979	0.72154
22	Final Fantasy VII Remake	-1.2249782	0.87908
23	Final Fantasy VIII	0.8096734	0.23098
25	Final Fantasy X	0.4530733	0.36258
26	Final Fantasy X-2	-2.6913868	0.99228
28	Final Fantasy XII	1.1515818	0.12595
29	Final Fantasy XIII	-0.7842501	0.77992
30	Final Fantasy XIII-2	-0.4611162	0.67899
31	Lightning Returns: Final Fantasy XIII	-0.8750701	0.82245
32	Final Fantasy XIV	1.6103386	0.06900
33	Final Fantasy XV	1.1291897	0.15687
34	Horizon Zero Dawn	-0.8638365	0.74498
36	Kingdom Hearts	1.0852511	0.13089
37	Kingdom Hearts II	1.4546342	0.01302
38	Kingdom Hearts III	0.9764778	0.11716
39	Kingdom Hearts 3D: Dream Drop Distance	1.7027103	0.00970
40	King's Quest I: Quest for the Crown	0.7852188	0.21920
41	King's Quest II: Romancing the Throne	-1.1467688	0.82143
42	King's Quest III: To Heir Is Human	0.5537754	0.35642
43	King's Quest IV: The Perils of Rosella	-1.8757304	0.98008
44	King's Quest V	0.0712877	0.48851
45	King's Quest VI	0.7271370	0.20225
46	King's Quest VII: The Princeless Bride	-1.3483466	0.89441
47	King's Quest VIII	0.7016147	0.26181
48	King's Quest Chapters	0.5324388	0.49298
50	Mass Effect	-1.3162577	0.89131
51	Mass Effect 2	-1.8439488	0.96110
54	Mass Effect 3	0.4105362	0.32394
56	Monkey Island 2: LeChuck's Revenge	5.3020303	0.00472
57	The Curse of Monkey Island	0.2320127	0.59692
58	The Secret of Monkey Island	-0.0479110	0.76657
59	Persona 3	0.1318010	0.45482
60	Persona 4	-0.0602077	0.52554
62	Persona 5	-0.8941771	0.81080
63	Stardew Valley	-1.2221531	0.88629
64	Star Wars: Knights of the Old Republic	-2.6013485	0.98781
65	Super Mario RPG: Legend of the Seven Stars	1.6607465	0.18181

Mini graph for publication:

```
xs = c(0.05,0.001,0.00001)
xs2 = c(rev(xs),1,xs)
ls = c(log10(rev(xs)),0,-log10(xs))
gxMini = ggplot(all, aes(x=p.random.log,y=p.perm.log)) +
  annotate("rect", xmin = -1000, xmax = 1000, ymin = -threshold, ymax = threshold, alpha = .1) +
  annotate("rect", xmin = -threshold, xmax = threshold, ymin = -1000, ymax = 1000, alpha = .1) +
  geom_point() +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  coord_cartesian(ylim=c(-4,4),xlim=c(-5,5)) +
  scale_x_continuous(breaks=ls,labels=xs2,
                     sec.axis = sec_axis(~.*1,
                        breaks = ls,
                        labels=c("","Less female\ndialogue\nthan expected","",
                                 "No bias",
```

```
                                  "",
                                  "More female\ndialogue\nthan expected",""))) +
  scale_y_continuous(breaks=ls,labels=xs2,
                     sec.axis = sec_axis(~.*1,
                        breaks = ls,
                        labels=c("","Less female\ndialogue\nper character\nthan expected","",
                                  "No bias","",
                                  "More female\ndialogue\nper character\nthan expected", ""))) +
  theme(panel.grid.minor = element_blank()) +
  xlab("Character bias\n(random gender p-value)") +
  ylab("Dialogue bias\n(permuted gender p-value)")
pdf("../results/graphs/CompareToBaseline_Mini.pdf",height=4,width=5)
gxMini
dev.off()
```

```
## pdf
##   2
```

Big plot for paper

```r
# pdf("../results/graphs/Big3.pdf",width=10,height=6)
# ggarrange(ggarrange(gxMini,
#                      wordsVCharacters+ geom_point(),
#                      nrow = 2, labels = c("A", "B")),
#           changeOverTime,
#           ncol = 2,
#           labels = "C",widths = c(1,1.5))
# dev.off()

pdf("../results/graphs/Big2.pdf",width=8,height=4)
ggarrange(gxMini,
          wordsVCharacters+ geom_point(),
          ncol = 2, labels = c("A", "B"),widths=c(1.2,1))
dev.off()
```

```
## pdf
##   2
```

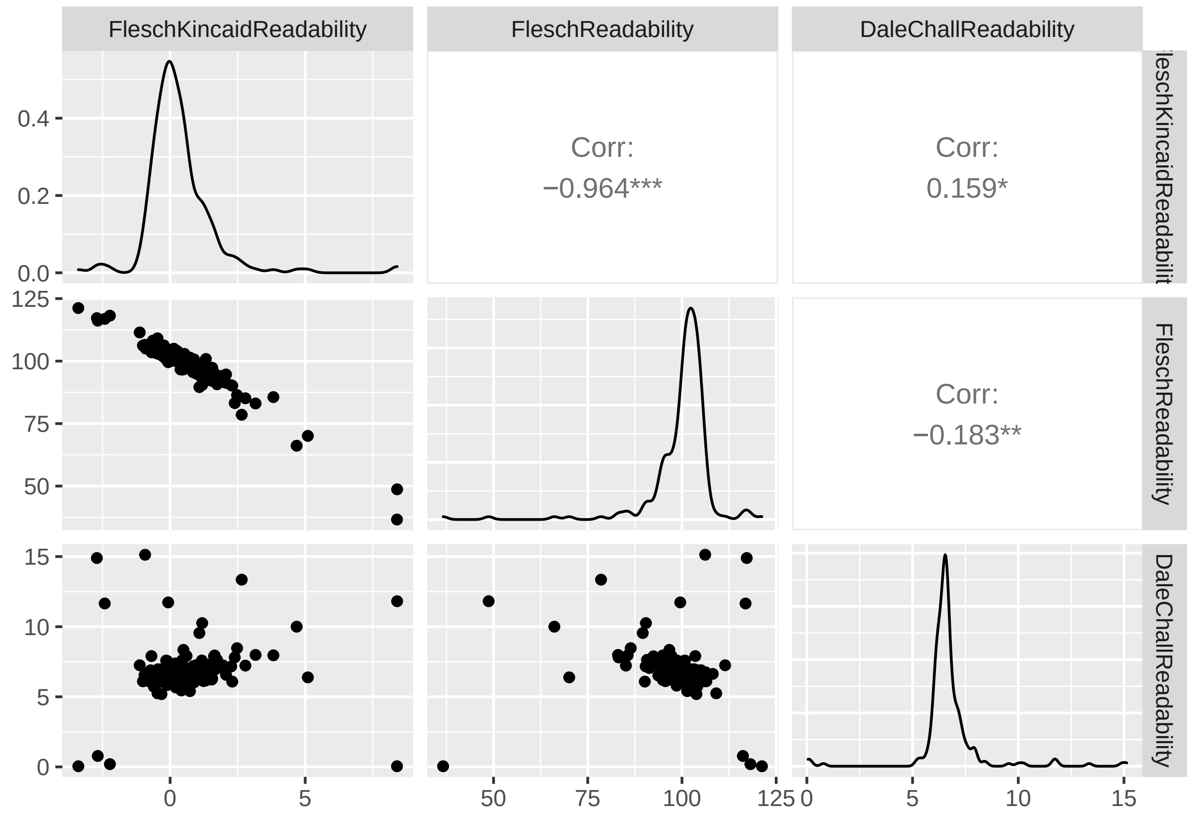
Readability

There are multiple measures of readability calculated by Textatistic:

- Flesch score
- Flesh Kincaid score
- Dale-Chall score

The Flesch and Dale-Chall measures are not strongly correlated with each other:

```
readMeasures = c("FleschKincaidReadability","FleschReadability","DaleChallReadability")
ggpairs(stats[,readMeasures])
```




```
getReadabilityTestText = function(readability,rt){
  p = rt$p.value
  if(p<0.0001){
    p  = "< 0.0001"
  } else{
    p = round(p,3)
  }
  paste0("game-level mean readability for male characters = ",
         round(mean(readability[1,],na.rm=T),2),
         ", sd = ",
         round(sd(readability[1,],na.rm=T),2),
         ", for female characters = ",
         round(mean(readability[2,],na.rm=T),2),
         ", sd = ",
         round(sd(readability[2,],na.rm=T),2),
         ", paired t-test t = ",
         round(rt$statistic,2),
         ", n = ",
         sum(!is.na(readability[1,])),
         ", p ",p
```

```
  )
}

games = unique(stats$folder)
grpPerGame =tapply(stats$group,stats$folder,length)
games = games[games %in% names(grpPerGame[grpPerGame>1])]

readability.DC = sapply(games, function(g){
  c(stats[stats$folder==g & stats$group=="male",]$DaleChallReadability,
    stats[stats$folder==g & stats$group=="female",]$DaleChallReadability)})
readability.DC.t = t.test(readability.DC[1,], readability.DC[2,], paired = T)
readability.DC.text = getReadabilityTestText(readability.DC,readability.DC.t)
cat(readability.DC.text,file="../results/latexStats/readability-DC-TTest.tex")

readability.F = sapply(games, function(g){
  c(stats[stats$folder==g & stats$group=="male",]$FleschReadability,
    stats[stats$folder==g & stats$group=="female",]$FleschReadability)})
readability.F.t = t.test(readability.F[1,], readability.F[2,], paired = T)
readability.F.text = getReadabilityTestText(readability.F,readability.F.t)
cat(readability.F.text,file="../results/latexStats/readability-F-TTest.tex")

readability.FK = sapply(games, function(g){
  c(stats[stats$folder==g & stats$group=="male",]$FleschKincaidReadability,
    stats[stats$folder==g & stats$group=="female",]$FleschKincaidReadability)})
readability.FK.t = t.test(readability.FK[1,], readability.FK[2,], paired = T)
readability.FK.text = getReadabilityTestText(readability.FK,readability.FK.t)
cat(readability.FK.text,file="../results/latexStats/readability-FK-TTest.tex")
```

The statistical results are as follows:

Dale-Chall: game-level mean readability for male characters = 6.58, sd = 0.37, for female characters = 6.42, sd = 0.41, paired t-test $t = 4.58$, $n = 50$, $p < 0.0001$

Flesch: game-level mean readability for male characters = 101.5, sd = 3.39, for female characters = 101.34, sd = 3.81, paired t-test $t = 0.4$, $n = 50$, p 0.692

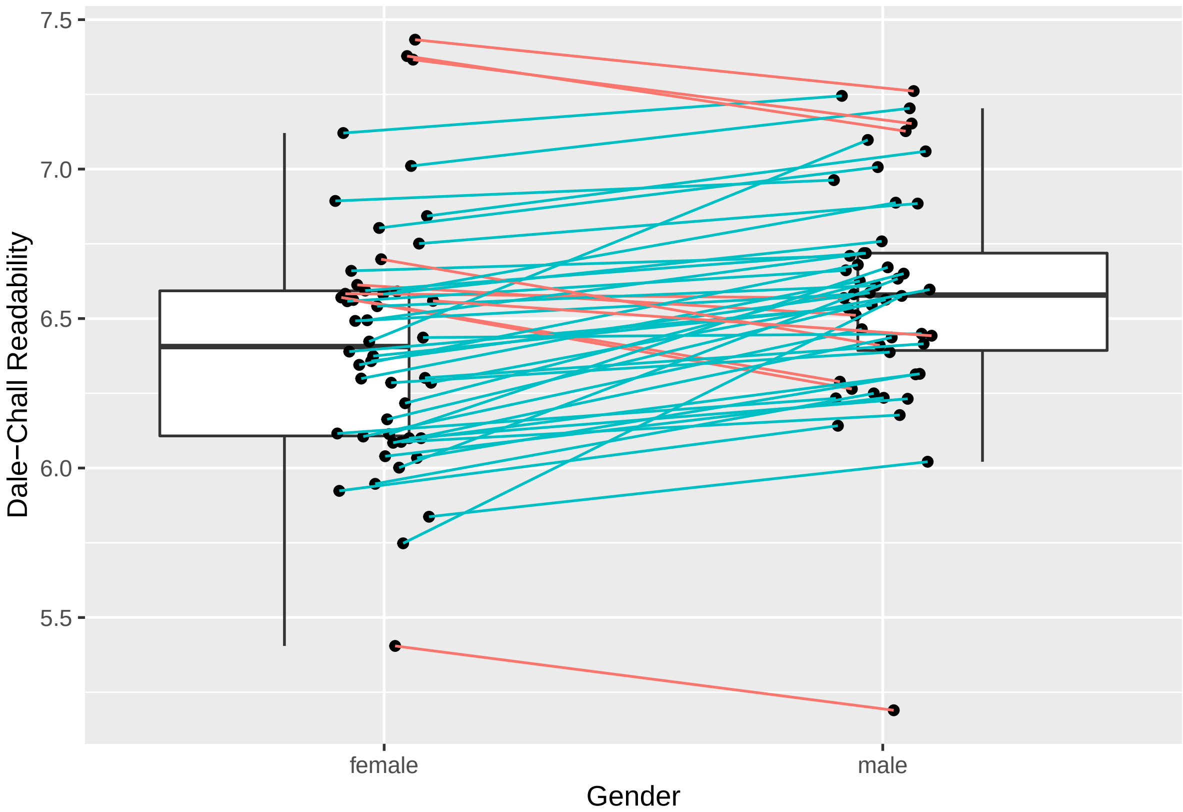
Flesch-Kinkaid: game-level mean readability for male characters = 0.16, sd = 0.68, for female characters = 0.18, sd = 0.77, paired t-test $t = -0.36$, $n = 50$, p 0.719

For the Dale-Chall scores, male dialogue has significantly higher values than female dialogue. This suggests that the required reading grade for male text is higher, or in other words male dialogue includes a smaller proportion of high-frequency words. The effect size is roughly the same as for the difference in reading ease between male and female academic journal publications (Hengel, 2022, table 3).

Below, we plot Dale-Chall scores, with values from the same game being connected by a line. The lines are coloured by blue = male > female, red = female > male :

```
stats = stats[order(stats$folder,stats$group),]
dir = tapply(stats[stats$group %in% c("male","female"),]$DaleChallReadability,
             stats[stats$group %in% c("male","female"),]$folder,
             function(X){X[1]<X[2]})
stats$dir = dir[stats$folder]
dcBox = stats[stats$group %in% c("male","female"),] %>%
  ggplot(aes(x=group,y=DaleChallReadability)) +
  geom_boxplot(outlier.alpha = 0, width = 0.5,
               position = position_nudge(x=c(-0.2,0.2))) +
  geom_point(aes(group=folder), position = position_dodge(0.2)) +
  geom_line(aes(group=folder,colour=dir),position = position_dodge(0.2))+
  xlab("Gender") +
  ylab("Dale-Chall Readability")+
  theme(legend.position = "none")
```

```r
dcBox
```




```r
pdf("../results/graphs/Readability.pdf",width=6,height=4)
dcBox
dev.off()
```

```
## pdf
##   2
```

However, the Flesch measures are not significant. In contrast, in Hengel (2022, table 3), the Flesch measures have stronger effect sizes than the Dale-Chall scores. Taken together with the lack of correlation between the measures, this casts some doubt on the robustness of the results for readability.

Discussion

This report demonstrated several things:

- There is about twice as much male dialogue than female dialogue in video games.
- There is high agreement between different measures of the amount of dialogue (number of words, lines, syllables).
- The proportion of female words is increasing over time.
- There is a significant correlation between the proportion of female characters and the proportion of female dialogue.
- The average male chararacter does not say significantly more than the average female character.
- There is considerable variation between games on the gender difference in number of words per sentence.
- There are some differences in readability between male and female dialogue, though the results are not robust over different measures of readability.

References

Hengel, E. (2022) Publishing while female: Are women held to higher standards? Evidence from peer review.https://www.erinhengel.com/research/publishing_female.pdf

5 Major and Minor characters

Gender balance between major and minor characters

TODO: Sort either beta regression or GAM

Introduction

This report explores whether the gender bias in the data as a whole is also observed when looking at either only major characters or only minor characters. Major characters may be playable characters, party characters, characters more central to the plot, or characters that speak more. One concern is that dialogue data for major characters is more complete and that gender is easier to code. This might bias estimates of the proportion of dialogue by female characters.

The analyses below demonstrate that conclusions about the proportion of female dialogue are unlikely to be affected by such concerns in our data.

Gender bias in the dialogue of major and minor characters

We first look at major and minor characters as discrete groups.

Load libraries:

```
library(rjson)
library(ggplot2)
library(ggstance)
library(mgcv)
library(knitr)
library(betareg)
```

Load data (for games with coded main player characters):

```
folders = list.dirs("../data", recursive = T)
folders = folders[sapply(folders,function(X){
    "stats_by_character.csv" %in% list.files(X)
  })]


allGames = NULL
for(folder in folders){
  shortName = tail(strsplit(folder,"/")[[1]],1)
  js = fromJSON(file = paste0(folder,"/meta.json"))
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  if(!alternativeMeasure){
    statsByChar = read.csv(paste0(folder,"/stats_by_character.csv"),stringsAsFactors = F)
    statsByChar = statsByChar[!is.na(statsByChar$words),]
    statsByChar = statsByChar[statsByChar$words>0,]
```

```
    if(nrow(statsByChar)>0 && !is.null(js$mainPlayerCharacters)){
      majc = statsByChar$charName %in% js$mainPlayerCharacters
      minc = (!statsByChar$charName %in% js$mainPlayerCharacters) &
        (statsByChar$group %in% c("male","female"))
      majc.Female = statsByChar$group=="female" & majc
      majc.Male = statsByChar$group=="male" & majc
      minc.Female =  statsByChar$group=="female" & minc
      minc.Male =  statsByChar$group=="male" & minc

      # Only include games with coded major characters
      if((sum(majc.Female) + sum(majc.Male))> 0 ){
        print(folder)
        majc.Female.words = sum(statsByChar[majc.Female,]$words)
        propFemaleDialogue.mainChar = 0
        if(majc.Female.words>0){
         propFemaleDialogue.mainChar =
           sum(statsByChar[majc.Female,]$words) /
             (sum(statsByChar[majc.Female,]$words) +
              sum(statsByChar[majc.Male,]$words))
        }
        propFemaleDialogue.minorChar =
          sum(statsByChar[minc.Female,]$words) /
            (sum(statsByChar[minc.Female,]$words) +
             sum(statsByChar[minc.Male,]$words))

        ret = data.frame(
          folder = folder,
          game = js$game,
          shortName = shortName,
          group = c("major","minor"),
          numFemaleWords = c(
            sum(statsByChar[majc.Female,]$words),
            sum(statsByChar[minc.Female,]$words)),
          numMaleWords = c(
            sum(statsByChar[majc.Male,]$words),
            sum(statsByChar[minc.Male,]$words)),
          propFemaleDialogue = c(
            propFemaleDialogue.mainChar,
            propFemaleDialogue.minorChar
          ),
          numFemaleCharacters = c(sum(majc.Female),sum(minc.Female)),
          numMaleCharacters = c(sum(majc.Male),sum(minc.Male))
        )
        allGames = rbind(allGames, ret)
      }
    }
  }
}
```

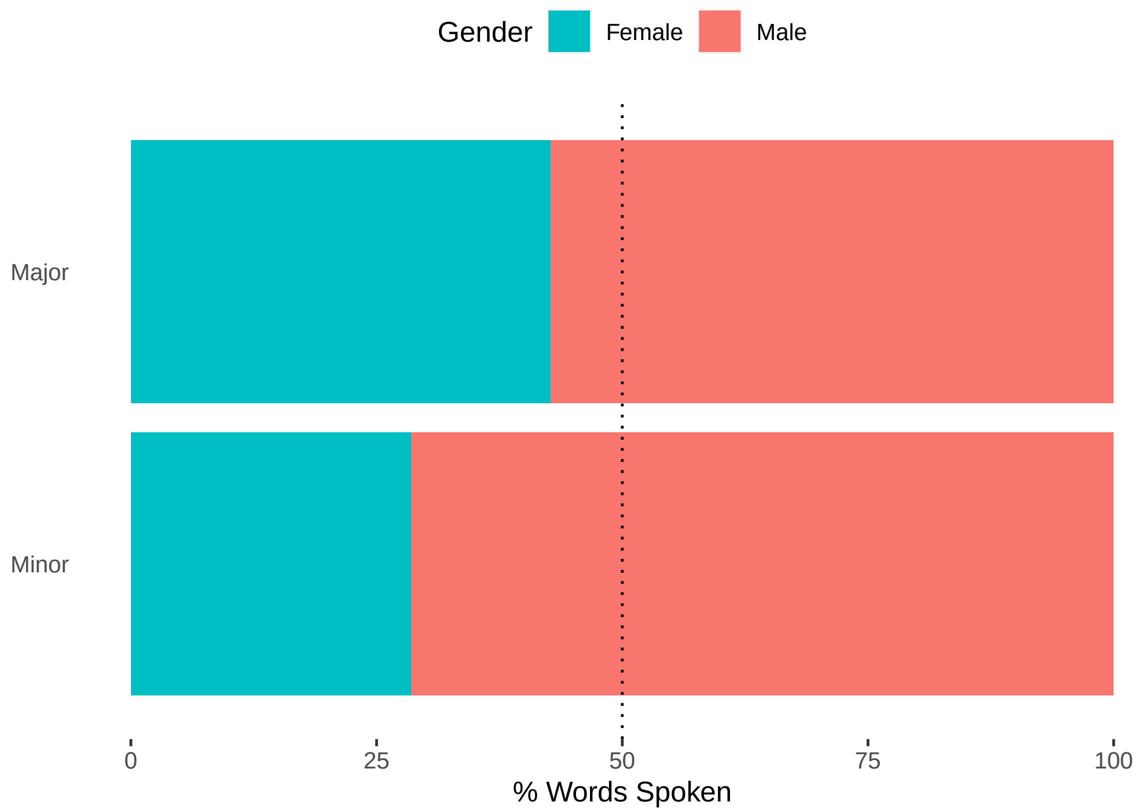
```
## [1] "../data/ChronoTrigger/ChronoTrigger"
## [1] "../data/DragonAge/DragonAgeOrigins_B"
## [1] "../data/FinalFantasy/FFII"
## [1] "../data/FinalFantasy/FFIV_DS"
## [1] "../data/FinalFantasy/FFIX_B"
```

```
## [1] "../data/FinalFantasy/FFV"
## [1] "../data/FinalFantasy/FFVI"
## [1] "../data/FinalFantasy/FFVII"
## [1] "../data/FinalFantasy/FFVII_Remake"
## [1] "../data/FinalFantasy/FFVIII"
## [1] "../data/FinalFantasy/FFX_B"
## [1] "../data/FinalFantasy/FFX2"
## [1] "../data/FinalFantasy/FFXII_B"
## [1] "../data/FinalFantasy/FFXIII"
## [1] "../data/FinalFantasy/FFXIII-2"
## [1] "../data/FinalFantasy/FFXIII-LR"
## [1] "../data/FinalFantasy/FFXV"
## [1] "../data/Horizon/HorizonZeroDawn"
## [1] "../data/KingdomHearts/KingdomHearts_B"
## [1] "../data/KingdomHearts/KingdomHearts2"
## [1] "../data/KingdomHearts/KingdomHearts3"
## [1] "../data/KingdomHearts/KingdomHearts3D"
## [1] "../data/KingsQuest/KingsQuest1"
## [1] "../data/KingsQuest/KingsQuest2"
## [1] "../data/KingsQuest/KingsQuest3"
## [1] "../data/KingsQuest/KingsQuest4"
## [1] "../data/KingsQuest/KingsQuest5"
## [1] "../data/KingsQuest/KingsQuest6"
## [1] "../data/KingsQuest/KingsQuest7"
## [1] "../data/KingsQuest/KingsQuest8"
## [1] "../data/KingsQuest/KingsQuestChapters"
## [1] "../data/MassEffect/MassEffect1B"
## [1] "../data/MassEffect/MassEffect2"
## [1] "../data/MassEffect/MassEffect3C"
## [1] "../data/MonkeyIsland/MonkeyIsland2"
## [1] "../data/MonkeyIsland/TheCurseOfMonkeyIsland"
## [1] "../data/MonkeyIsland/TheSecretOfMonkeyIsland"
## [1] "../data/Persona/Persona3"
## [1] "../data/Persona/Persona4"
## [1] "../data/Persona/Persona5B"
## [1] "../data/StarWarsKOTOR/StarWarsKOTOR"
## [1] "../data/SuperMarioRPG/SuperMarioRPG"
```

Visualise total amount of female vs. male dialogue for major and minor characters:

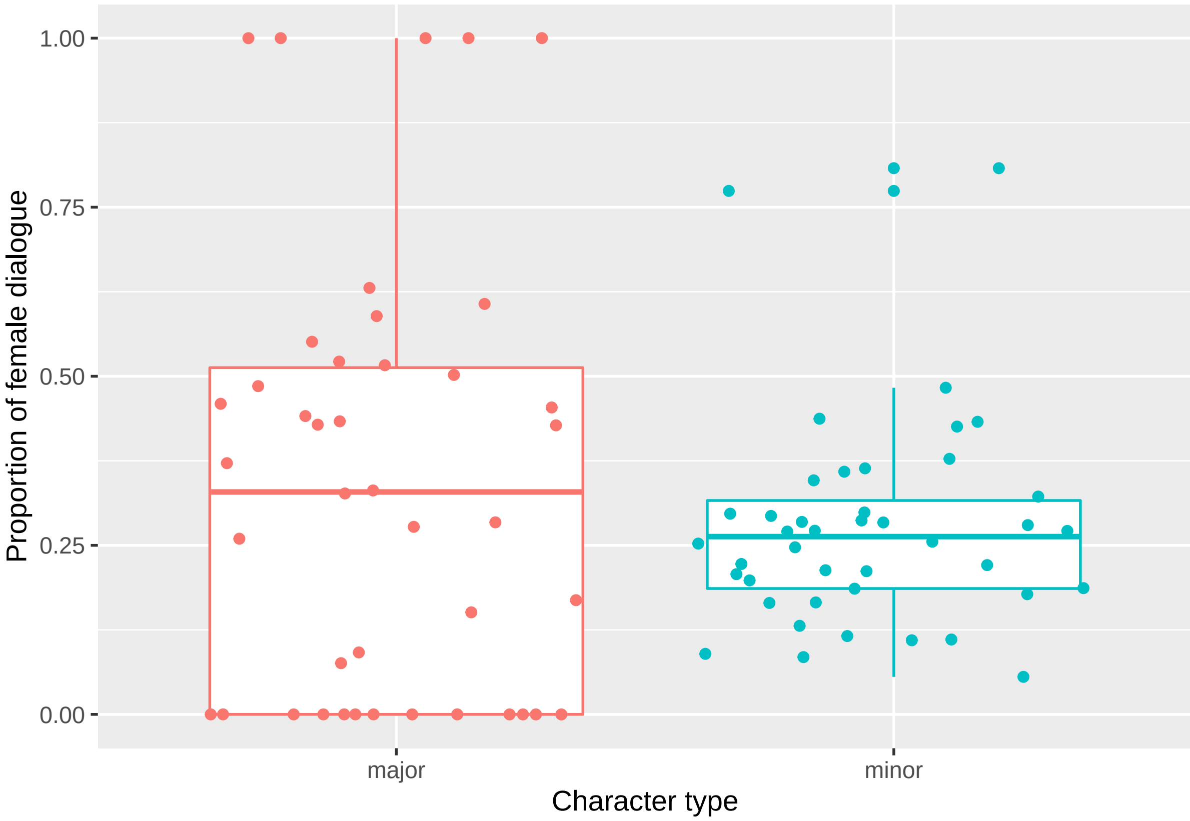
```
allGames.Maj.PercentFemale =
  100 * (
  sum(allGames[allGames$group=="major",]$numFemaleWords) /
   (sum(allGames[allGames$group=="major",]$numFemaleWords) +
      sum(allGames[allGames$group=="major",]$numMaleWords)))
allGames.Maj.PercentMale = 100 - allGames.Maj.PercentFemale

allGames.Min.PercentFemale =
  100 * (
  sum(allGames[allGames$group=="minor",]$numFemaleWords) /
   (sum(allGames[allGames$group=="minor",]$numFemaleWords) +
      sum(allGames[allGames$group=="minor",]$numMaleWords)))
allGames.Min.PercentMale = 100 - allGames.Min.PercentFemale

allGames$propFemaleCharacters=
```

```
  allGames$numFemaleCharacters /
    (allGames$numFemaleCharacters + allGames$numMaleCharacters)

dx = data.frame(
  Gender=factor(c("Male","Female","Male","Female"),
                levels=c("Male","Female")),
  Group = factor(c("Major","Major","Minor","Minor"),
                 levels=c("Minor","Major")),
  percentageWords=
    c(allGames.Maj.PercentMale,allGames.Maj.PercentFemale,
      allGames.Min.PercentMale,allGames.Min.PercentFemale))

ggplot(dx,aes(x=Group,y=percentageWords,fill=Gender))+ geom_bar(stat='identity')+
  geom_hline(yintercept=50,linetype="dotted") +
  coord_flip(ylim = c(0,100)) +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = "top") +
  scale_fill_discrete(breaks=c("Female","Male"),name="Gender")+
  ylab("% Words Spoken") +
  xlab("")
```



Below is a boxplot of the distribution of individual games. It's clear that the minor character group has a higher mean proportion of female dialogue, though the range for major characters is higher.

```
ggplot(allGames, aes(y=propFemaleDialogue,x=group,colour=group)) +
  geom_boxplot() +
  geom_jitter() +
  theme(legend.position = "none")+
  xlab("Character type") +
  ylab("Proportion of female dialogue")
```



Statistical test of average proportion of female dialogue in each game, comparing major and minor characters.

```
t.test(allGames$propFemaleDialogue~allGames$group)
```

```
##
##  Welch Two Sample t-test
##
## data:  allGames$propFemaleDialogue by allGames$group
## t = 1.2057, df = 58.644, p-value = 0.2328
## alternative hypothesis: true difference in means between group major and group minor is not equal to
## 95 percent confidence interval:
##  -0.04416159  0.17801532
## sample estimates:
## mean in group major mean in group minor
##           0.3424845           0.2755577
```

The test is significant, suggesting that the female dialogue is lower in major characters than in minor characters.

However, we also know that the proportion of female dialogue is predicted by the proportion of female characters. The regression below tests whether the character groups (major or minor) predict the proportion of female dialogue over and above the proportion of female characters within the group:

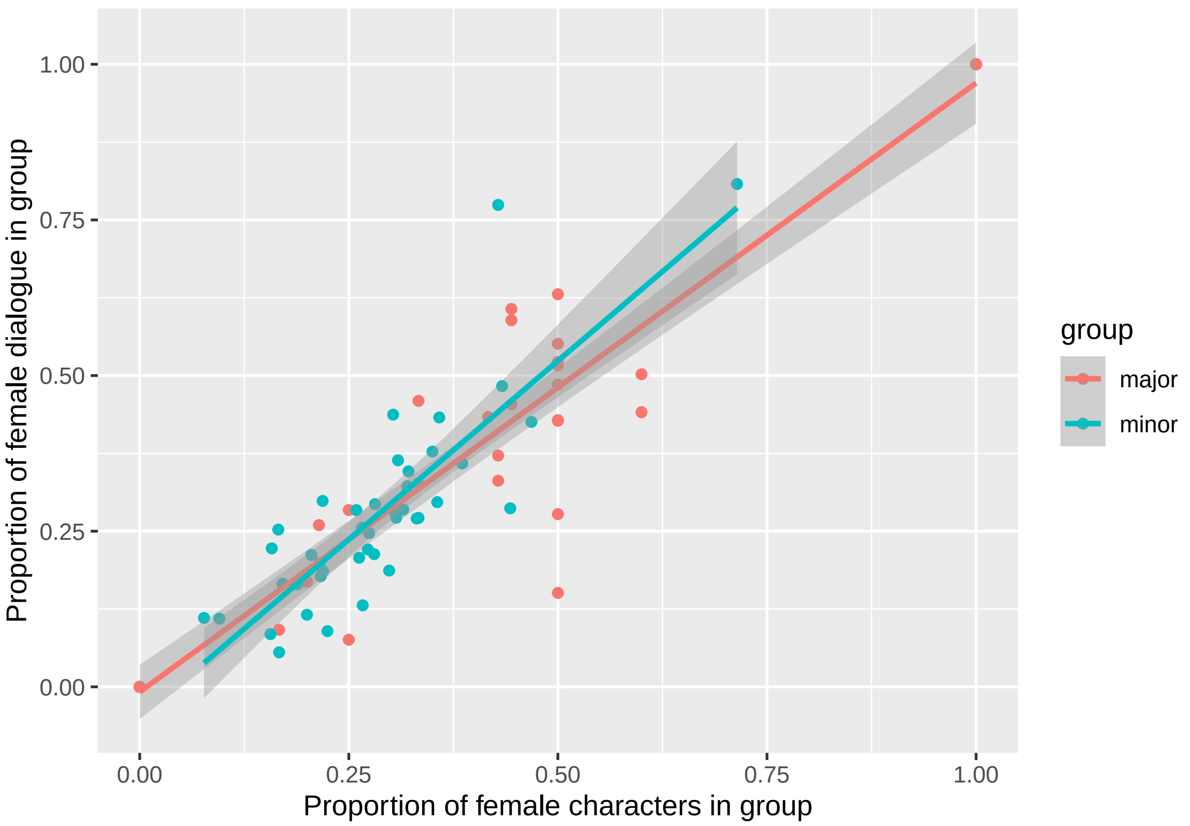
```
mInt = lm(propFemaleDialogue~ group * propFemaleCharacters, data = allGames)
summary(mInt)
```

```
##
## Call:
## lm(formula = propFemaleDialogue ~ group * propFemaleCharacters,
##     data = allGames)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33000 -0.04380  0.00805  0.03433  0.33245
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -0.008052   0.020819  -0.387    0.700
## groupminor                      -0.042076   0.043018  -0.978    0.331
## propFemaleCharacters             0.977932   0.043627  22.416   <2e-16 ***
## groupminor:propFemaleCharacters  0.169571   0.130958   1.295    0.199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08907 on 80 degrees of freedom
## Multiple R-squared:  0.8825, Adjusted R-squared:  0.8781
## F-statistic: 200.2 on 3 and 80 DF,  p-value: < 2.2e-16
```

The effect of the proprotion of female characters is significant, but the effect of group is not, nor is the interaction.

```
ggplot(allGames,
       aes(x=propFemaleCharacters,
           y=propFemaleDialogue,
           colour=group)) +
  geom_point() +
  stat_smooth(method='lm') +
  xlab("Proportion of female characters in group") +
  ylab("Proportion of female dialogue in group")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



In summary, the gender bias against female dialogue is evident for both major and minor characters. This bias is exaggerated for major characters. However, this can be mostly explained by the number of female characters, rather than a systematic difference between major and minor character groups.

The full list of results:

```
kable(allGames[,c("game","group","propFemaleDialogue")])
```

game	group	propFemaleDialogue
Chrono Trigger	major	0.6306403
Chrono Trigger	minor	0.2526138
Dragon Age: Origins	major	0.4538605
Dragon Age: Origins	minor	0.2554227
Final Fantasy II	major	0.2839721
Final Fantasy II	minor	0.2986326
Final Fantasy IV	major	0.0916418
Final Fantasy IV	minor	0.1309963
Final Fantasy IX	major	0.3311388
Final Fantasy IX	minor	0.2073560
Final Fantasy V	major	0.4411320
Final Fantasy V	minor	0.0555346
Final Fantasy VI	major	0.2598039
Final Fantasy VI	minor	0.0847276
Final Fantasy VII	major	0.3266936
Final Fantasy VII	minor	0.1981802
Final Fantasy VII Remake	major	0.5020579
Final Fantasy VII Remake	minor	0.2838857
Final Fantasy VIII	major	0.4283841
Final Fantasy VIII	minor	0.1859402
Final Fantasy X	major	0.3715065
Final Fantasy X	minor	0.2967687
Final Fantasy X-2	major	1.0000000
Final Fantasy X-2	minor	0.2117596
Final Fantasy XII	major	0.2773916
Final Fantasy XII	minor	0.2716304
Final Fantasy XIII	major	0.4854317
Final Fantasy XIII	minor	0.2207313
Final Fantasy XIII-2	major	0.4593027
Final Fantasy XIII-2	minor	0.3778930
Lightning Returns: Final Fantasy XIII	major	1.0000000
Lightning Returns: Final Fantasy XIII	minor	0.3588774
Final Fantasy XV	major	0.0000000
Final Fantasy XV	minor	0.4326998
Horizon Zero Dawn	major	1.0000000
Horizon Zero Dawn	minor	0.2704224
Kingdom Hearts	major	0.0000000
Kingdom Hearts	minor	0.2846654
Kingdom Hearts II	major	0.0000000
Kingdom Hearts II	minor	0.2131808
Kingdom Hearts III	major	0.0756906
Kingdom Hearts III	minor	0.1777299
Kingdom Hearts 3D: Dream Drop Distance	major	0.0000000
Kingdom Hearts 3D: Dream Drop Distance	minor	0.0895143
King's Quest I: Quest for the Crown	major	0.0000000
King's Quest I: Quest for the Crown	minor	0.1158129
King's Quest II: Romancing the Throne	major	0.0000000
King's Quest II: Romancing the Throne	minor	0.8076923
King's Quest III: To Heir Is Human	major	0.0000000

game	group	propFemaleDialogue
King's Quest III: To Heir Is Human	minor	0.2713463
King's Quest IV: The Perils of Rosella	major	1.0000000
King's Quest IV: The Perils of Rosella	minor	0.7741100
King's Quest V	major	0.0000000
King's Quest V	minor	0.3461666
King's Quest VI	major	0.0000000
King's Quest VI	minor	0.2471298
King's Quest VII: The Princeless Bride	major	1.0000000
King's Quest VII: The Princeless Bride	minor	0.1867603
King's Quest VIII	major	0.0000000
King's Quest VIII	minor	0.2799134
King's Quest Chapters	major	0.1509165
King's Quest Chapters	minor	0.4830259
Mass Effect	major	0.5510669
Mass Effect	minor	0.2935248
Mass Effect 2	major	0.4334063
Mass Effect 2	minor	0.4372264
Mass Effect 3	major	0.5162871
Mass Effect 3	minor	0.3221426
Monkey Island 2: LeChuck's Revenge	major	0.0000000
Monkey Island 2: LeChuck's Revenge	minor	0.2224024
The Curse of Monkey Island	major	0.0000000
The Curse of Monkey Island	minor	0.1095802
The Secret of Monkey Island	major	0.0000000
The Secret of Monkey Island	minor	0.1107168
Persona 3	major	0.5889274
Persona 3	minor	0.2867931
Persona 4	major	0.5217268
Persona 4	minor	0.4256015
Persona 5	major	0.4274823
Persona 5	minor	0.3639303
Star Wars: Knights of the Old Republic	major	0.6070153
Star Wars: Knights of the Old Republic	minor	0.1655871
Super Mario RPG: Legend of the Seven Stars	major	0.1688733
Super Mario RPG: Legend of the Seven Stars	minor	0.1647965

Bias in different quantiles of character dialgoue

Does the gender bias differ for characters that speak a lot compared to characters that don't?

For each game, we divide characters into four groups based on the amount of dialogue they speak. Characters are ranked by the proportion of dialogue that they speak within the game. Then the characters are split into four groups of even number (the 'quantiles' of the dialogue proportions). Across all games, the total number of words is calculated for each gender for each quantile.

```
folders = list.dirs("../data", recursive = T)
folders = folders[sapply(folders,function(X){
    "stats_by_character.csv" %in% list.files(X)
  })]


allChars = NULL
for(folder in folders){
  shortName = tail(strsplit(folder,"/")[[1]],1)
  js = fromJSON(file = paste0(folder,"/meta.json"))
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  if(!alternativeMeasure){
    statsByChar = read.csv(paste0(folder,"/stats_by_character.csv"),stringsAsFactors = F)
    statsByChar = statsByChar[!is.na(statsByChar$words),]
    statsByChar = statsByChar[statsByChar$words>0,]
    if(nrow(statsByChar)>0){
        statsByChar = statsByChar[statsByChar$group %in% c("male","female"),]
        statsByChar$dialogProp = statsByChar$words/sum(statsByChar$words)
        allChars = rbind(allChars,statsByChar)
      }
    }
}
```

```
numQuantiles = 4
allChars$Quantile = cut(allChars$dialogProp,
      breaks= quantile(allChars$dialogProp,
          probs = seq(0,1,length.out=numQuantiles+1)))
q = data.frame(
  Quantile = 1:numQuantiles,
  femaleWords = tapply(
    allChars[allChars$group=="female",]$words,
    allChars[allChars$group=="female",]$Quantile,sum),
  maleWords = tapply(
    allChars[allChars$group=="male",]$words,
    allChars[allChars$group=="male",]$Quantile,sum),
  femaleChars = tapply(
    allChars[allChars$group=="female",]$words,
    allChars[allChars$group=="female",]$Quantile,length),
  maleChars = tapply(
    allChars[allChars$group=="male",]$words,
    allChars[allChars$group=="male",]$Quantile,length)
)
q$propFemale = q$femaleWords/ (q$femaleWords+q$maleWords)
q$propMale = q$maleWords/ (q$femaleWords+q$maleWords)
```

```r
q$propFemaleChar = q$femaleChar/ (q$femaleChar+q$maleChar)
q$propMaleChar = q$maleChar/ (q$femaleChar+q$maleChar)
```

There are small but significant differences in the proportion of words in each of the four quantiles:

```r
q[,c("femaleWords","maleWords","propFemale","propMale")]
```

```
##                      femaleWords maleWords propFemale  propMale
## (1.29e-06,0.000151]        27980     60779  0.3152356 0.6847644
## (0.000151,0.00051]         79830    190099  0.2957444 0.7042556
## (0.00051,0.00188]         221795    520444  0.2988188 0.7011812
## (0.00188,0.62]           1667455   2910938  0.3642009 0.6357991
```

```r
chisq = chisq.test(q[,c("femaleWords","maleWords")])
chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  q[, c("femaleWords", "maleWords")]
## X-squared = 16467, df = 3, p-value < 2.2e-16
```

```r
pv = chisq$p.value
if(pv < 0.0001){
  pv = paste0("p = 0.0001")
  } else{
  pv = paste0("p = ",round(pv,3))
}
chisqOut = paste0("$\\chi^2$ = ", round(chisq$statistic,2),", ",pv)
cat(chisqOut,file="../results/latexStats/quantileWordsChiSq.tex")
# compare just first and last quantile
chisq.test(q[c(1,4),c("femaleWords","maleWords")])
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  q[c(1, 4), c("femaleWords", "maleWords")]
## X-squared = 902.33, df = 1, p-value < 2.2e-16
```

However, the proportion of characters is not significantly different (the p-value is marginal at the 0.05 level, but very different from the result above).

```r
q[,c("femaleChars","maleChars","propFemaleChar","propMaleChar")]
```

```
##                      femaleChars maleChars propFemaleChar propMaleChar
## (1.29e-06,0.000151]          871      2176      0.2858549    0.7141451
## (0.000151,0.00051]           902      2141      0.2964180    0.7035820
## (0.00051,0.00188]            926      2118      0.3042050    0.6957950
## (0.00188,0.62]               874      2171      0.2870279    0.7129721
```

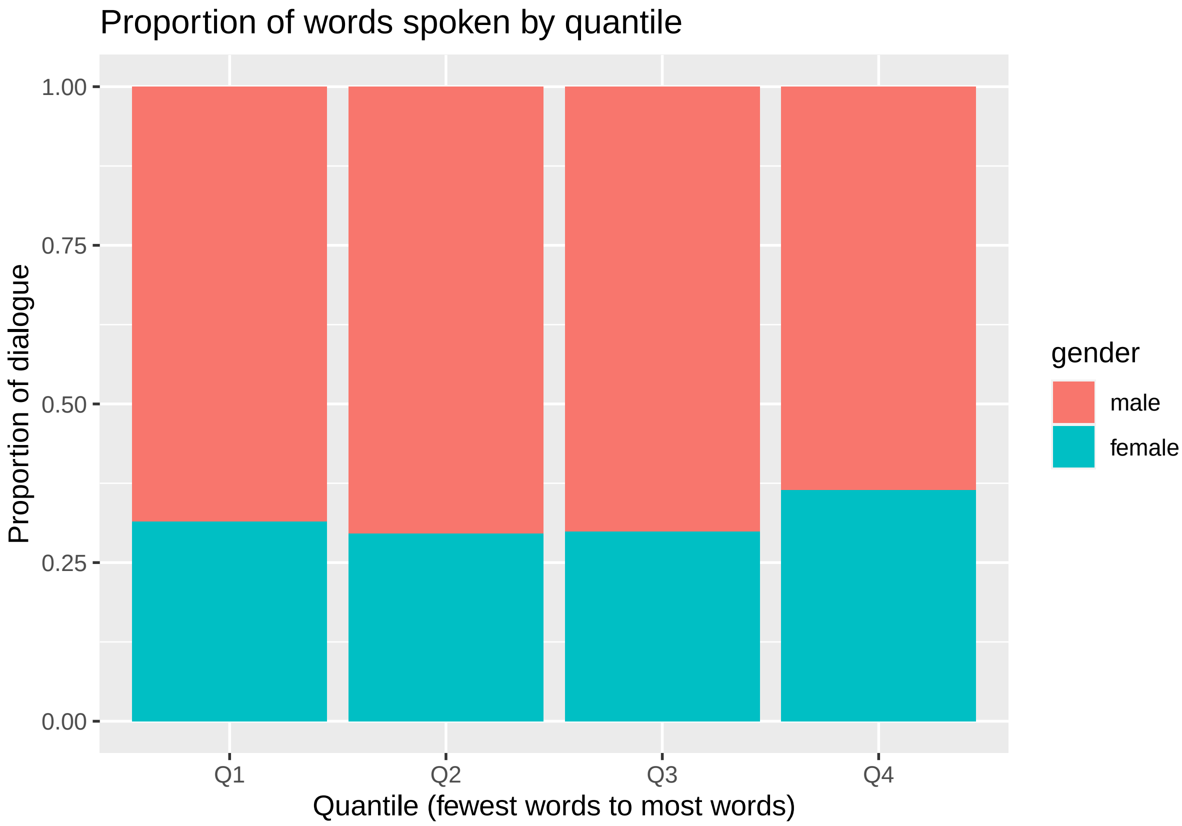
```r
chisq = chisq.test(q[,c("femaleChars","maleChars")])
pv = chisq$p.value
if(pv < 0.0001){
  pv = paste0("p = 0.0001")
  } else{
  pv = paste0("p = ",round(pv,3))
}
```

```r
chisqOut = paste0("$\\chi^2$ = ", round(chisq$statistic,2),", ",pv)
cat(chisqOut,file="../results/latexStats/quantileCharChiSq.tex")
# compare just first and last quantile
chisq.test(q[c(1,4),c("femaleChars","maleChars")])
```

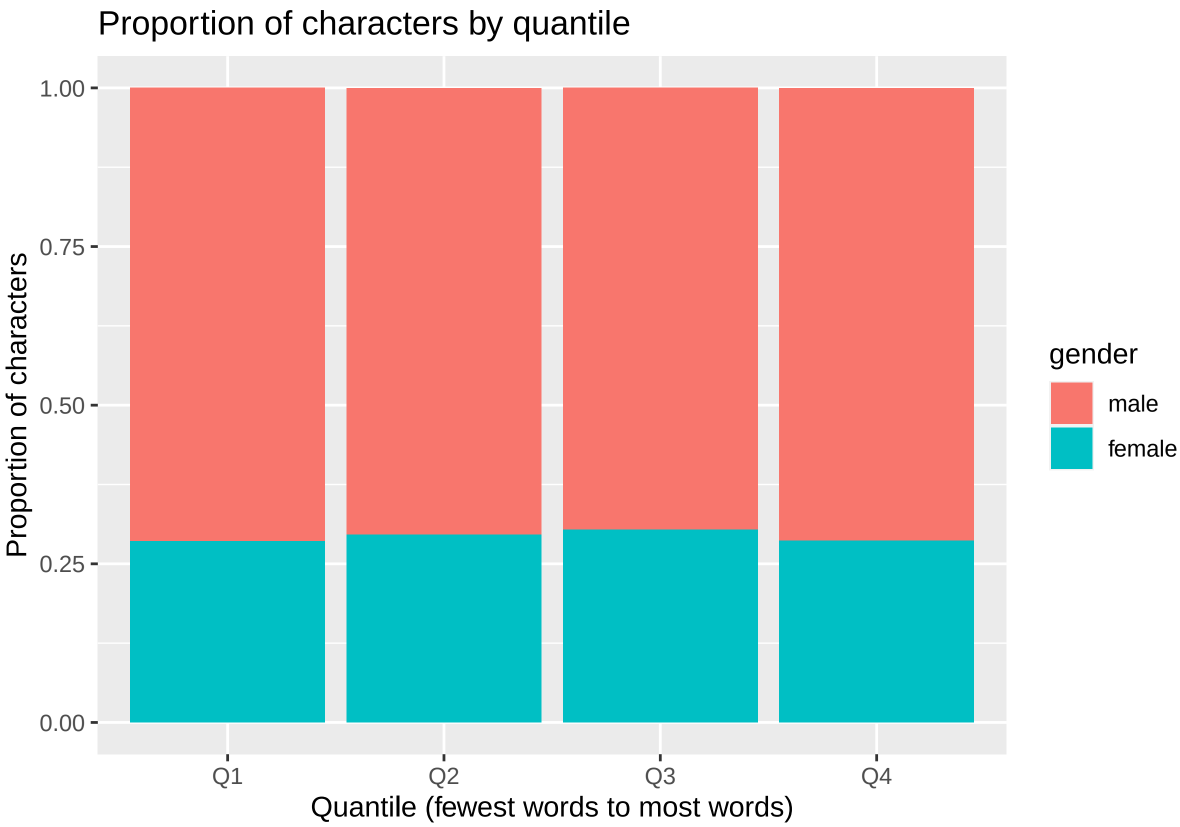
```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  q[c(1, 4), c("femaleChars", "maleChars")]
## X-squared = 0.0053164, df = 1, p-value = 0.9419
```

Plot results:

```r
q2 = data.frame(
  Quantile = paste0("Q",rep(1:numQuantiles,2)),
  gender = factor(rep(c("female","male"),each=numQuantiles),
                  levels=c("male","female")),
  prop = c(q$propFemale,q$propMale),
  propChar =c(q$propFemaleChar,q$propMaleChar)
)

ggplot(q2,aes(x=Quantile,fill=gender,y=prop)) +
  geom_bar(stat = "identity",position="stack") +
  ylab("Proportion of dialogue") +
  xlab("Quantile (fewest words to most words)") +
  ggtitle("Proportion of words spoken by quantile")
```



```
ggplot(q2,aes(x=Quantile,fill=gender,y=propChar)) +
  geom_bar(stat = "identity",position="stack") +
  ylab("Proportion of characters") +
  xlab("Quantile (fewest words to most words)")+
  ggtitle("Proportion of characters by quantile")
```



In summary, there are small differences in the proportion of female dialogue for talkative and less talkative characters. But the overall

Other estimates of bias

One concern is that the coding for main characters may be more complete or more accurate than coding for minor characters, because minor characters are harder to find in videos, have less documentation on wikis and less direct linguistic cues.

Method 1: Complete vs. incomplete scripts

Each game is coded for whether it's complete, partial or a sample of the dialogue in the full game. If the estimate was biased by the completeness of the coding, then we might expect to see a difference in the estimated proportion of female dialogue between these types.

Below is a boxplot showing how the proportion of female dialogue differs between the game types.

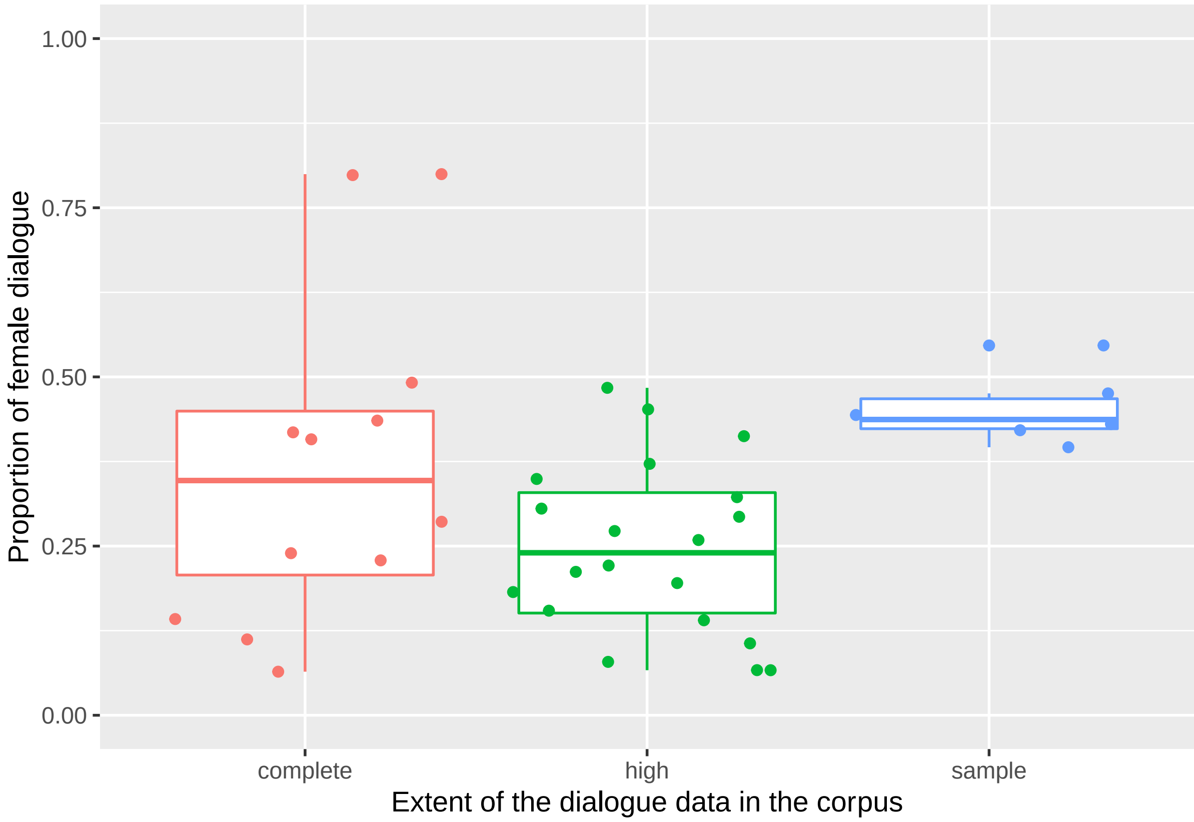
First, we load the data:

```
allGames.completeness = NULL
for(folder in folders){
  js = fromJSON(file = paste0(folder,"/meta.json"))
  completeness = NA
  if(!is.null(js$sourceFeatures)){
    if(!is.null(js$sourceFeatures$completeness)){
      completeness = js$sourceFeatures$completeness
    }
  }
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  if(!alternativeMeasure){
    stats = read.csv(paste0(folder,"/stats_by_character.csv"),stringsAsFactors = F)
    propFemaleDialogue = sum(stats[stats$group=="female",]$words) /
      sum(stats[stats$group %in% c("male","female"),]$words)
    propFemaleCharacters = sum(stats$group=="female")/
      sum(stats$group %in% c("male","female"))
    ret = data.frame(
      folder = folder,
      game = js$game,
      completeness = completeness,
      propFemaleDialogue = propFemaleDialogue,
      propFemaleCharacters = propFemaleCharacters,
      year = js$year
    )
    allGames.completeness = rbind(allGames.completeness, ret)
  }
}
allGames.completeness = allGames.completeness[
  is.finite(allGames.completeness$propFemaleDialogue),]
```

Below we plot the proportion of female dialogue according to the extent of the data coded:

```
ggplot(allGames.completeness[!is.na(allGames.completeness$completeness),],
       aes(x=completeness, y =propFemaleDialogue, colour=completeness)) +
  geom_boxplot() +
  geom_jitter() +
  theme(legend.position = "none")+
```

```
  xlab("Extent of the dialogue data in the corpus") +
  ylab("Proportion of female dialogue") +
  coord_cartesian(ylim=c(0,1))
```



A t-test comparing proportion of female dialogue in 'complete' and 'high' sources:

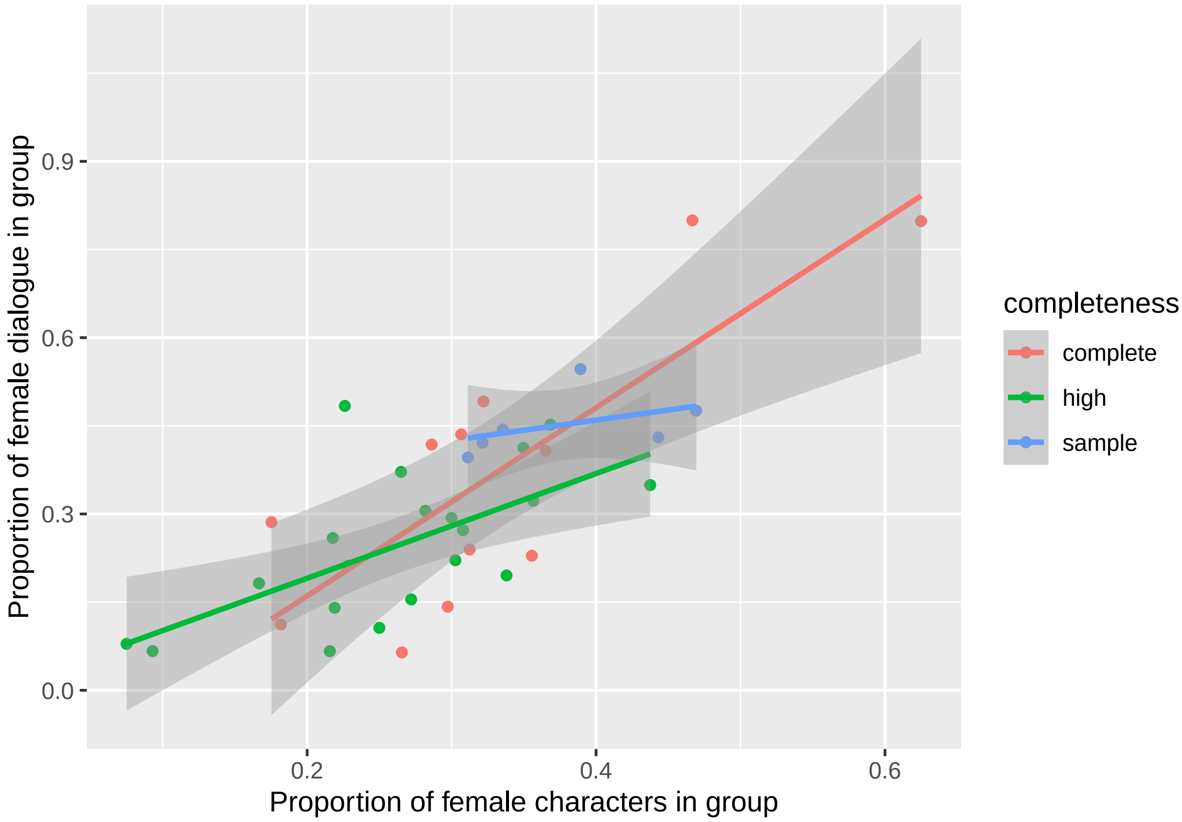
```
t.test(propFemaleDialogue~completeness,
       data = allGames.completeness[
         allGames.completeness$completeness %in%
           c("complete", "high"),])
```

```
##
##  Welch Two Sample t-test
##
## data:  propFemaleDialogue by completeness
## t = 1.6076, df = 14.676, p-value = 0.1292
## alternative hypothesis: true difference in means between group complete and group high is not equal t
## 95 percent confidence interval:
##  -0.03986572  0.28264781
## sample estimates:
## mean in group complete     mean in group high
##              0.3686546              0.2472635
```

The proportion of female dialogue is slightly higher for more complete data. This might suggest that the estimate of female dialogue is underestimated in the corpus. However, we also know that the proportion of female dialogue is predicted by the proportion of female characters:

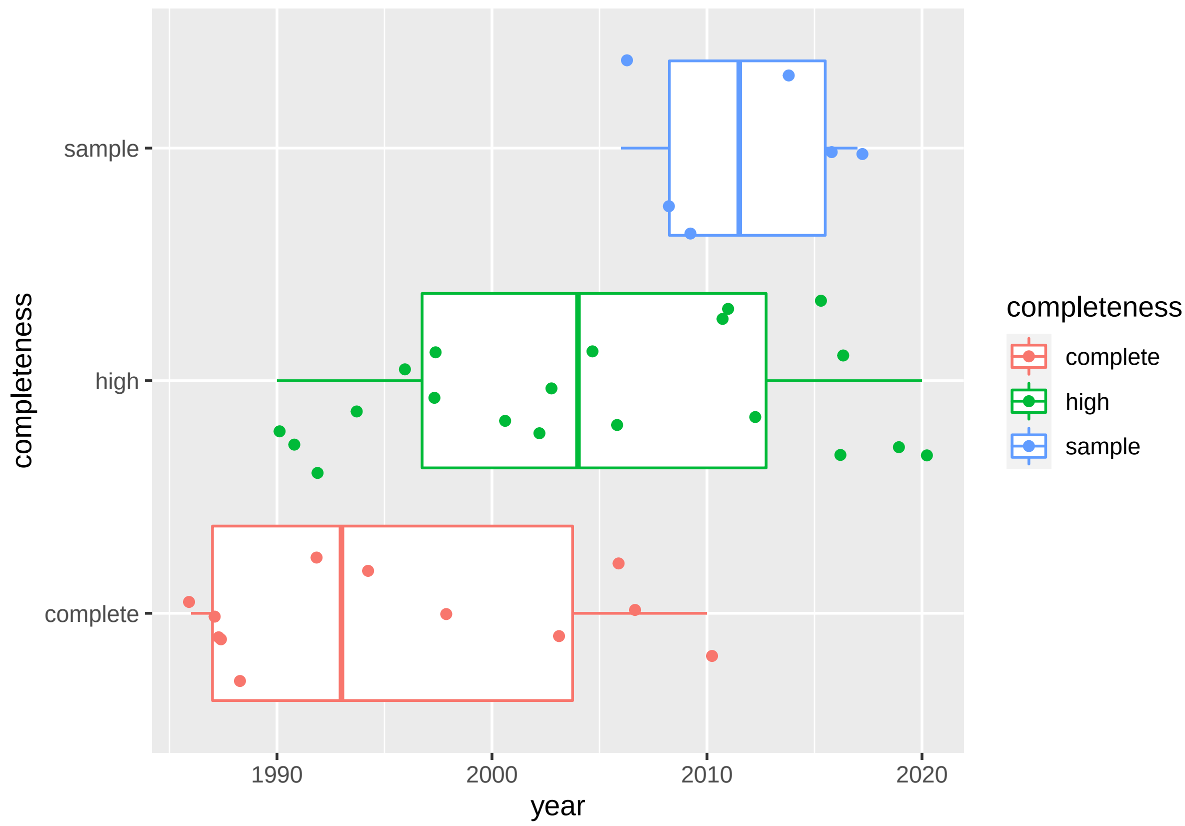
```
ggplot(allGames.completeness[!is.na(allGames.completeness$completeness),],
       aes(x=propFemaleCharacters,
           y =propFemaleDialogue,
           colour=completeness)) +
  geom_point() +
  stat_smooth(method='lm') +
  xlab("Proportion of female characters in group") +
  ylab("Proportion of female dialogue in group")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



It could also be the case that the year of publication is related to how complete the script is (since earlier games have less content, and are more likely to have accessible data). This appears to be the case:

```
ggplot(allGames.completeness[!is.na(allGames.completeness$completeness),],
       aes(x=year, y=completeness,colour=completeness)) +
  geom_boxplot() +
  geom_jitter()
```



This makes it seem like the the relationship with completeness may be partially driven by the proprotion of female characters and/or by the date of release. Here is a regression, predicting the proportion of female dialogue by completeness and by the proportion of female characters:

```
mComp = lm(propFemaleDialogue ~ completeness + propFemaleCharacters + year,
   data = allGames.completeness)
summary(mComp)
```

```
##
## Call:
## lm(formula = propFemaleDialogue ~ completeness + propFemaleCharacters +
##     year, data = allGames.completeness)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.22375 -0.08847  0.01843  0.07518  0.28310
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -2.574439   4.555489  -0.565    0.576
## completenesshigh     -0.054719   0.051601  -1.060    0.297
## completenesssample    0.005606   0.070947   0.079    0.937
## propFemaleCharacters  1.182882   0.210203   5.627 2.89e-06 ***
## year                  0.001279   0.002287   0.559    0.580
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.1207 on 33 degrees of freedom
## Multiple R-squared:  0.6009, Adjusted R-squared:  0.5525
## F-statistic: 12.42 on 4 and 33 DF,  p-value: 2.858e-06
```

It seems there is no effect of completeness once the proprotion of female charaters is taken into account.

Method 2: Cumulative gender balance

We can try to estimate how our measuring of gender balance is affected by the coding of minor characters. This can be done by looking at how the estimate varies as we observe more and more minor characters. That is, what would the estimate be like if we had only coded the top 10% of most prolific characters (those that speak most), or the top 20%, or top 30% etc. At some point, the estimate of gender balance would converge on the final estimate for all coded characters.

Example

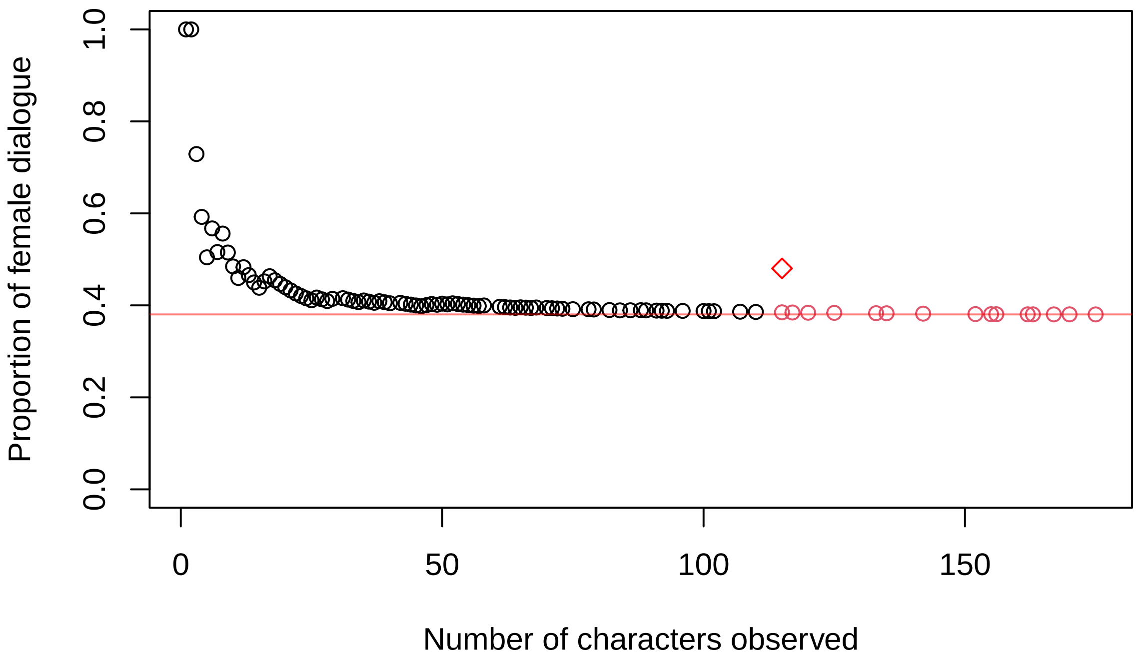
Here's a function that works out the gender balance in dialogue over the range of characters. It ranks all characters from most words to least words, then works out the gender balance taking into account just the top character, the top two characters, the top three characters etc.

```
getBalanceOverCumulativeCharacterRange = function(statsByChar){
  # Remove other groups except male and female
  statsByChar = statsByChar[statsByChar$group %in% c("male","female"),]
  statsByChar = statsByChar[!is.na(statsByChar$words),]
  statsByChar = statsByChar[statsByChar$words>0,]

  # Sorted list of number of words for each character (most to least)
  sortedUniqueNumOfWords = sort(unique(statsByChar$words),decreasing = TRUE)

  # Table of total number of words observed for each gender
  #  as the number of characters observed increases from
  #  character with most dialogue to least dialogue
  wordsByGender.Cumulative =
    sapply(sortedUniqueNumOfWords,
      function(minNumWords){
        x = statsByChar[statsByChar$words>=minNumWords,]
        femaleWords = sum(x[x$group=="female",]$words)
        maleWords = sum(x[x$group=="male",]$words)
        femaleProp = femaleWords / (femaleWords+maleWords)
        return(c(femaleWords,maleWords))
      })
  # Convert to proportion of female dialogue
  femalePropCumulative = wordsByGender.Cumulative[1,] /
    colSums(wordsByGender.Cumulative)
  totalEstimate = femalePropCumulative[length(femalePropCumulative)]
  # Binomial test at each point: is it significantly different
  #  from the total?
  sigDifferentFromTotal.p = apply(wordsByGender.Cumulative,2,
    function(wbg){
      x = binom.test(wbg,p = totalEstimate)$p.value
  })

  sigDifferentFromTotal = sigDifferentFromTotal.p>0.05
  firstNonSignif = which(sigDifferentFromTotal)[1]
  # stable non-signif (point after last significant result)
  firstStableNonSignif = length(sigDifferentFromTotal) - which(!rev(sigDifferentFromTotal))[1] + 2

  numCharCumulative = sapply(sortedUniqueNumOfWords,
      function(minNumWords){
        sum(statsByChar$words>=minNumWords)
      })
```

```
  numCharBeforeEstimateFirstNotSigDiff = numCharCumulative[firstNonSignif]
  numCharBeforeEstimateStabilises = numCharCumulative[firstStableNonSignif]

  return(list(femalePropCumulative = femalePropCumulative,
      numCharCumulative = numCharCumulative,
      p = sigDifferentFromTotal.p,
      numCharBeforeEstimateFirstNotSigDiff = numCharBeforeEstimateFirstNotSigDiff,
      numCharBeforeEstimateStabilises = numCharBeforeEstimateStabilises
              ))
}

# Function for visualising the information
plotCumulativeCharRange = function(stats){
  plot(stats$femalePropCumulative~
       stats$numCharCumulative,
     xlab = "Number of characters observed",
     ylab = "Proportion of female dialogue",
     ylim=c(0,1),
     col = 1 + (stats$p>0.05))
abline(h=tail(stats$femalePropCumulative,n=1),
       col=rgb(1,0,0,0.5))
points(stats$numCharBeforeEstimateStabilises,
       tail(stats$femalePropCumulative,n=1) + 0.1,
       pch=5, col="red")
}
```

We can apply this to Chrono Trigger:

```
chrono =  read.csv("../data/ChronoTrigger/ChronoTrigger/stats_by_character.csv",
                   stringsAsFactors = F)
chrono = chrono[!is.na(chrono$words),]
chrono = chrono[chrono$words>0,]
chrono.stats = getBalanceOverCumulativeCharacterRange(chrono)
plotCumulativeCharRange(chrono.stats)
```

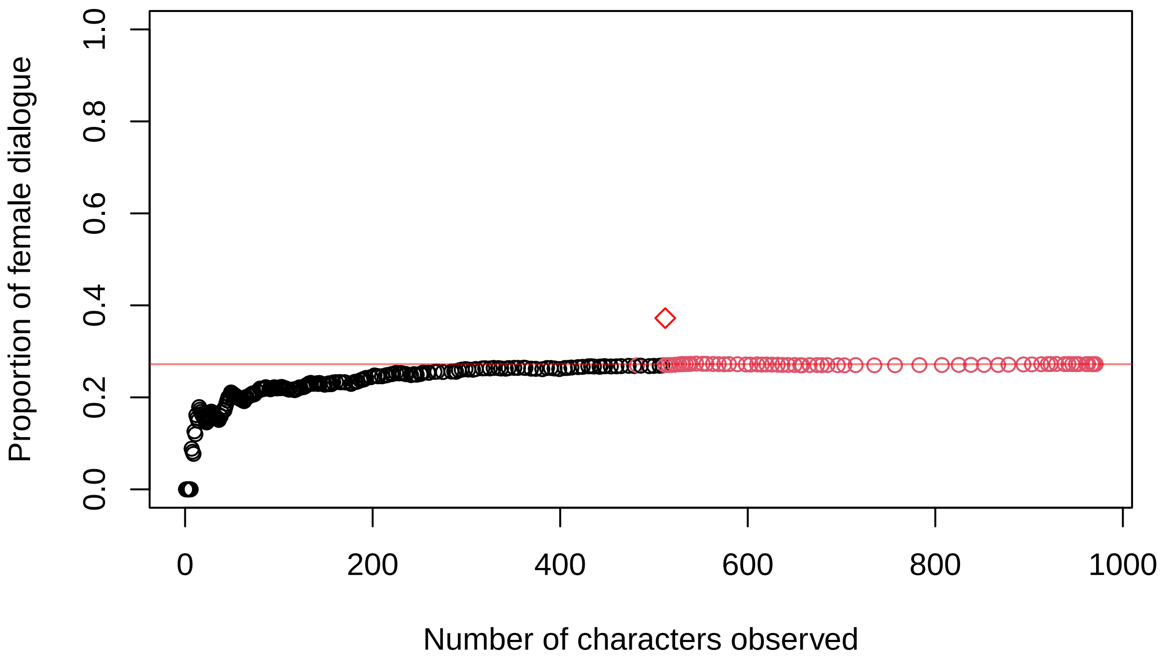


The points in the plot above shows the gender balance when taking into account various numbers of characters. Note that there is not an estiamte for every possible number of characters, since several characters are tied in the number of words they speak. The red line shows the gender balance in the full game. The points are coloured red if they are not significantly different from the gender balance in the full game.

It's clear that the proportion of female dialogue is higher for characters with more dialogue. The first time that the estimate is not signficantly different from the total estimate is after seeing 115 characters, or after seeing 66% of the characters. After this, the estimate does not change significantly. This is indicated with a red diamond on the plot. This also happens to be the same as the point at which the estimate stabilises (the point after which the estimate is never again significantly different from the total estimate).

Things are a little different for Final Fantasy XII, which has many more characters:

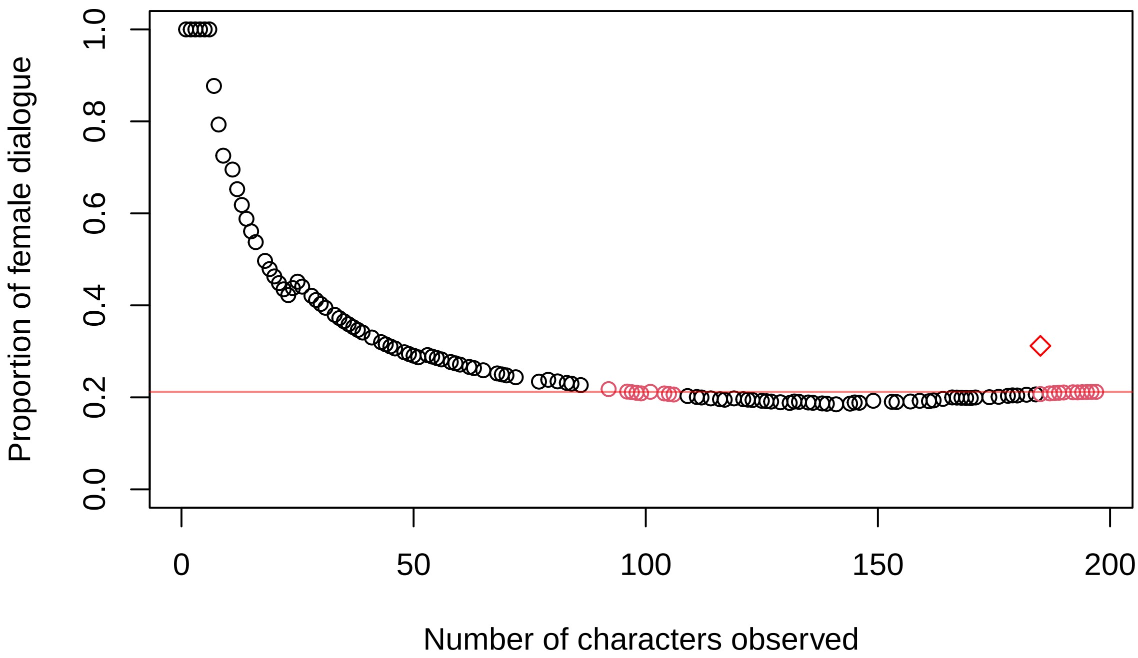
```
FFXII =  read.csv("../data/FinalFantasy/FFXII_B/stats_by_character.csv",
                  stringsAsFactors = F)
FFXII.stats = getBalanceOverCumulativeCharacterRange(FFXII)
plotCumulativeCharRange(FFXII.stats)
```



Here, it's clear that the proportion of female dialogue is *lower* for characters with more dialogue. The first time that the estimate is not signficantly different from the total estimate is after seeing 482 characters, or after seeing 50% of the characters.

Things are different again for Daggerfall:

```
Daggerfall =  read.csv("../data/ElderScrolls/Daggerfall/stats_by_character.csv",
                  stringsAsFactors = F)
Daggerfall.stats = getBalanceOverCumulativeCharacterRange(Daggerfall)
plotCumulativeCharRange(Daggerfall.stats)
```



Here, the female dialogue is over-estimated for characters with lots of dialogue, but then under-estimated for mid-range characters. It is only after seeing 47% of characters that the estimate stabilises.

Cumulative gender balance over all games

We can now estiamte the statistics above for all games:

```
folders = list.dirs("../data", recursive = T)
folders = folders[sapply(folders,function(X){
  "stats_by_character.csv" %in% list.files(X)
})]


allGames.Cum = NULL
for(folder in folders){
  shortName = tail(strsplit(folder,"/")[[1]],1)
  js = fromJSON(file = paste0(folder,"/meta.json"))
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  if(!alternativeMeasure){
    print(folder)
    statsByChar = read.csv(paste0(folder,"/stats_by_character.csv"),stringsAsFactors = F)
    statsByChar = statsByChar[!is.na(statsByChar$words),]
    statsByChar = statsByChar[statsByChar$words>0,]

    if(nrow(statsByChar)>0){
      gameStats = getBalanceOverCumulativeCharacterRange(statsByChar)
      totalNumChar = tail(gameStats$numCharCumulative,n=1)
      percentCharBeforeEstimateStabilises=
        100 * (gameStats$numCharBeforeEstimateStabilises / totalNumChar)
      ret = data.frame(
        folder = folder,
        game = js$game,
        shortName = shortName,
        totalNumChar = totalNumChar,
        femalePercentCumulative = 100*gameStats$femalePropCumulative,
        percentCharCumulative = 100*(gameStats$numCharCumulative/totalNumChar),
        percentCharBeforeEstimateStabilises = percentCharBeforeEstimateStabilises
      )
      ret$femalePercentTotal = tail(ret$femalePercentCumulative,n=1)
      allGames.Cum = rbind(allGames.Cum,ret)
    }
  }
}
```

```
## [1] "../data/ChronoTrigger/ChronoTrigger"
## [1] "../data/DragonAge/DragonAge2"
## [1] "../data/DragonAge/DragonAgeOrigins_B"
## [1] "../data/ElderScrolls/Daggerfall"
## [1] "../data/ElderScrolls/Morrowind"
## [1] "../data/ElderScrolls/Oblivion"
## [1] "../data/ElderScrolls/Skyrim"
## [1] "../data/FinalFantasy/FFI"
## [1] "../data/FinalFantasy/FFII"
## [1] "../data/FinalFantasy/FFIV_DS"
## [1] "../data/FinalFantasy/FFIX_B"
```

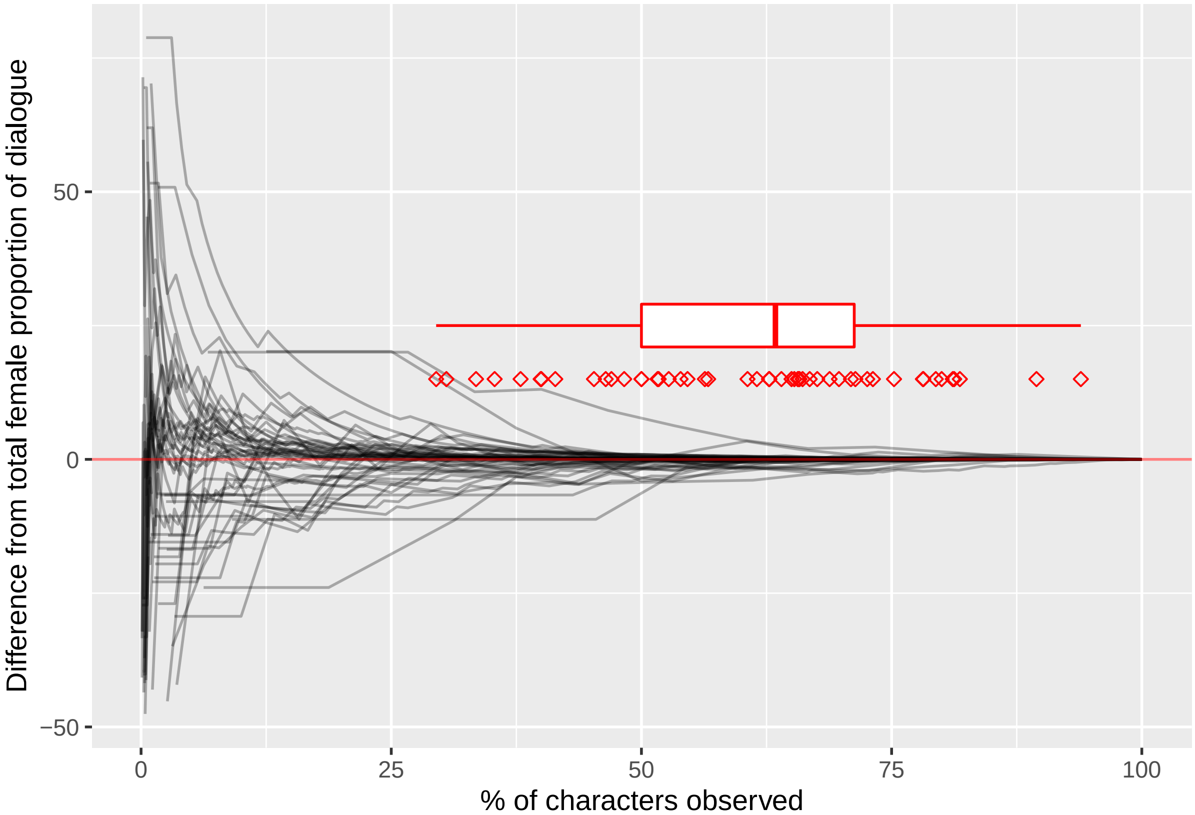
```
## [1] "../data/FinalFantasy/FFV"
## [1] "../data/FinalFantasy/FFVI"
## [1] "../data/FinalFantasy/FFVII"
## [1] "../data/FinalFantasy/FFVII_Remake"
## [1] "../data/FinalFantasy/FFVIII"
## [1] "../data/FinalFantasy/FFX_B"
## [1] "../data/FinalFantasy/FFX2"
## [1] "../data/FinalFantasy/FFXII_B"
## [1] "../data/FinalFantasy/FFXIII"
## [1] "../data/FinalFantasy/FFXIII-2"
## [1] "../data/FinalFantasy/FFXIII-LR"
## [1] "../data/FinalFantasy/FFXIV"
## [1] "../data/FinalFantasy/FFXV"
## [1] "../data/Horizon/HorizonZeroDawn"
## [1] "../data/KingdomHearts/KingdomHearts_B"
## [1] "../data/KingdomHearts/KingdomHearts2"
## [1] "../data/KingdomHearts/KingdomHearts3"
## [1] "../data/KingdomHearts/KingdomHearts3D"
## [1] "../data/KingsQuest/KingsQuest1"
## [1] "../data/KingsQuest/KingsQuest2"
## [1] "../data/KingsQuest/KingsQuest3"
## [1] "../data/KingsQuest/KingsQuest4"
## [1] "../data/KingsQuest/KingsQuest5"
## [1] "../data/KingsQuest/KingsQuest6"
## [1] "../data/KingsQuest/KingsQuest7"
## [1] "../data/KingsQuest/KingsQuest8"
## [1] "../data/KingsQuest/KingsQuestChapters"
## [1] "../data/MassEffect/MassEffect1B"
## [1] "../data/MassEffect/MassEffect2"
## [1] "../data/MassEffect/MassEffect3C"
## [1] "../data/MonkeyIsland/MonkeyIsland2"
## [1] "../data/MonkeyIsland/TheCurseOfMonkeyIsland"
## [1] "../data/MonkeyIsland/TheSecretOfMonkeyIsland"
## [1] "../data/Persona/Persona3"
## [1] "../data/Persona/Persona4"
## [1] "../data/Persona/Persona5B"
## [1] "../data/StardewValley/StardewValley"
## [1] "../data/StarWarsKOTOR/StarWarsKOTOR"
## [1] "../data/SuperMarioRPG/SuperMarioRPG"
```

```
allGames.Cum$diffFromTotalEstimate =
  allGames.Cum$femalePercentCumulative - allGames.Cum$femalePercentTotal
```

The graph below visualises the data for all games. The vertical axis plots the distance from the final total estiamte for the specific game, so that all games converge on zero (no difference from the total estimate). The horizontal axis shows the percentage of characters seen, so that each game is normalised for the number of characters. Red diamonds show the points at which the estimate for each game stabilises. The boxplot shows the distribution of the stabilisation points over the horizontal axis.

```
ggplot(allGames.Cum,
       aes(x=percentCharCumulative,
           y=diffFromTotalEstimate,
           group=folder)) +
  geom_line(alpha=0.3) +
geom_hline(yintercept = 0, alpha=0.5,colour='red') +
```

```
  geom_point(data=allGames.Cum[!duplicated(allGames.Cum$folder),],
    aes(x=percentCharBeforeEstimateStabilises,
        y=15),
    colour="red",shape =5) +
  geom_boxploth(data=allGames.Cum[!duplicated(allGames.Cum$folder),],
                aes(x=percentCharBeforeEstimateStabilises,
                    y=25,group=NULL), colour="red", width=8) +
  ylab("Difference from total female proportion of dialogue") +
  xlab("% of characters observed")
```



Estimates for all games stabilised before seeing 95% of their characters, suggesting that the estimates for the specific games in the corpus would not change much with the addition of data from more minor characters.

The mean stabilisation point is after seeing 60.7437992% of characters (95% quantile = [31.17468 ,87.7511962].

Is the estimate biased?

The graphs above show that, when considering only the most prolific characters, there are games which both over- and under- estimate the proprotion of female dialogue. Additionally, we want to formally test whether the estimate of the estimate of the proportion of female dialogue is biased *in the corpus as a whole*. We can do this by looking at the data after seeing 10% of the characters in a game. The t-test below tests whether there is a bias one way or the other:

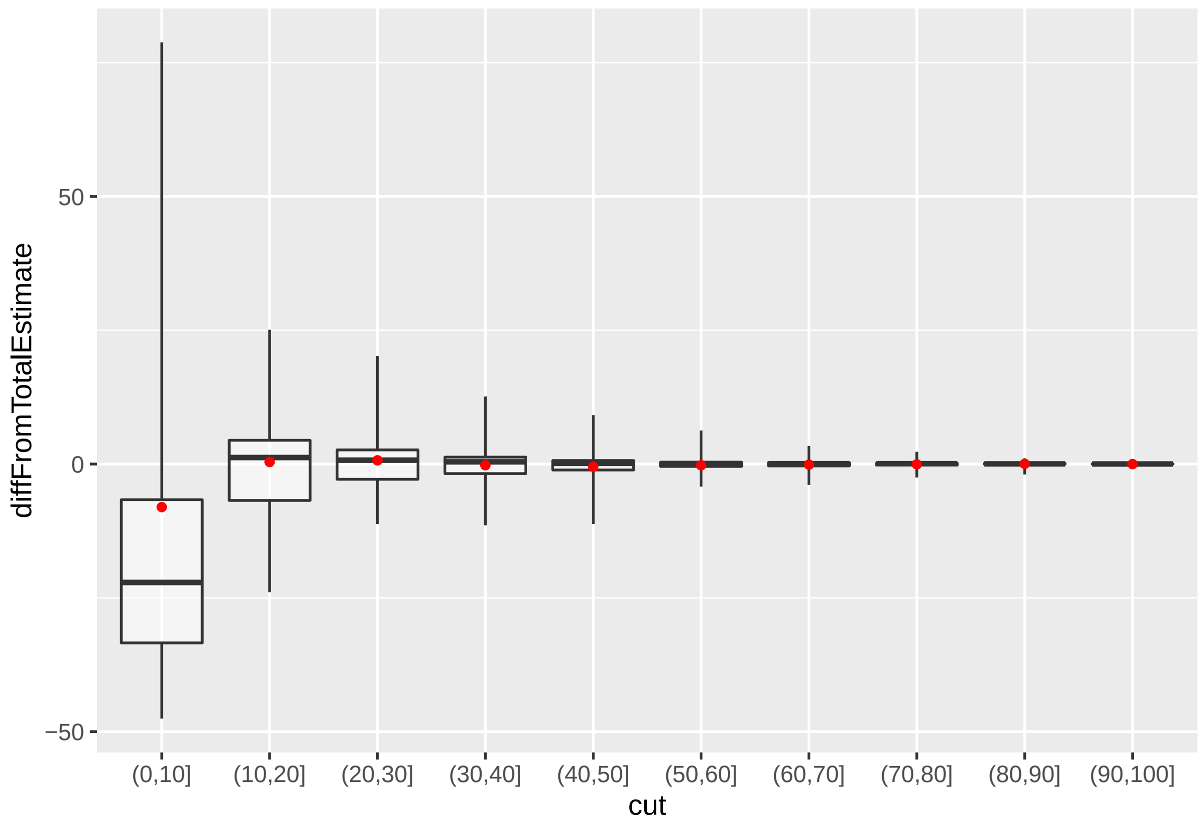
```
firstTenPercent = allGames.Cum[allGames.Cum$percentCharCumulative>=10.00,]
firstTenPercent = firstTenPercent[!duplicated(firstTenPercent$folder),]
tBias = t.test(firstTenPercent$diffFromTotalEstimate)
tBias
```

```
##
##  One Sample t-test
##
## data:  firstTenPercent$diffFromTotalEstimate
## t = -0.008225, df = 49, p-value = 0.9935
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -2.967659  2.943466
## sample estimates:
##  mean of x
## -0.0120968
```

The mean difference from the final estimate is -0.0120968, which indicates that the estimate based on prolific characters may be under-estimating the proportion of female speech on average. However, this is not significantly different from 0.

A similar point is made when drawing boxplots for the mean estimate of female dialogue % for each 10% of the range of characters seen (the whiskers show the full range of the data). Seeing only 10% of the characters biases the estimates (the proportion of female dialogue is under-estimated, though this isn't significant as shown by the t-test above). After seeing 10% of the characters, the estimates are very close to the final estimates after seeing all the characters.

```
allGames.Cum$cut = cut(allGames.Cum$percentCharCumulative,
                       breaks = seq(0,100,by=10))
ggplot(allGames.Cum[!duplicated(paste(allGames.Cum$folder, allGames.Cum$cut)),],
       aes(y=diffFromTotalEstimate,
           x = cut)) +
  geom_boxplot(alpha=0.5, coef = Inf) +
  stat_summary(fun=mean, geom="point", shape=20, size=2, color="red", fill="red")
```



However, cutting the data in 10% chunks is arbitrary, and the data within each chunk are not necessarily normally distributed.

A more generalised answer can be give by fitting a general additive model (GAM), predicting the difference from the total estimate based on the percentage of characters seen, with random effects to capture the dependency of datapoints that belong to the same game.

```
mBeta = betareg(I(0.3 + (diffFromTotalEstimate/100)) ~
                percentCharCumulative | percentCharCumulative,
                data = allGames.Cum[allGames.Cum$percentCharCumulative>5,])

summary(mBeta)
```

```
##
## Call:
## betareg(formula = I(0.3 + (diffFromTotalEstimate/100)) ~ percentCharCumulative |
##     percentCharCumulative, data = allGames.Cum[allGames.Cum$percentCharCumulative >
##     5, ])
##
## Standardized weighted residuals 2:
##      Min       1Q   Median       3Q      Max
## -14.2673  -0.0195   0.1715   0.3686   7.3104
##
## Coefficients (mean model with logit link):
##                         Estimate Std. Error  z value Pr(>|z|)
## (Intercept)           -8.401e-01  1.743e-03 -482.060  < 2e-16 ***
## percentCharCumulative -8.464e-05  1.949e-05   -4.344  1.4e-05 ***
```

```
##
## Phi coefficients (precision model with log link):
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)           3.1811550  0.0330767   96.17   <2e-16 ***
## percentCharCumulative 0.0873846  0.0006701  130.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 1.741e+04 on 4 Df
## Pseudo R-squared: 0.009122
## Number of iterations: 17 (BFGS) + 2 (Fisher scoring)
```

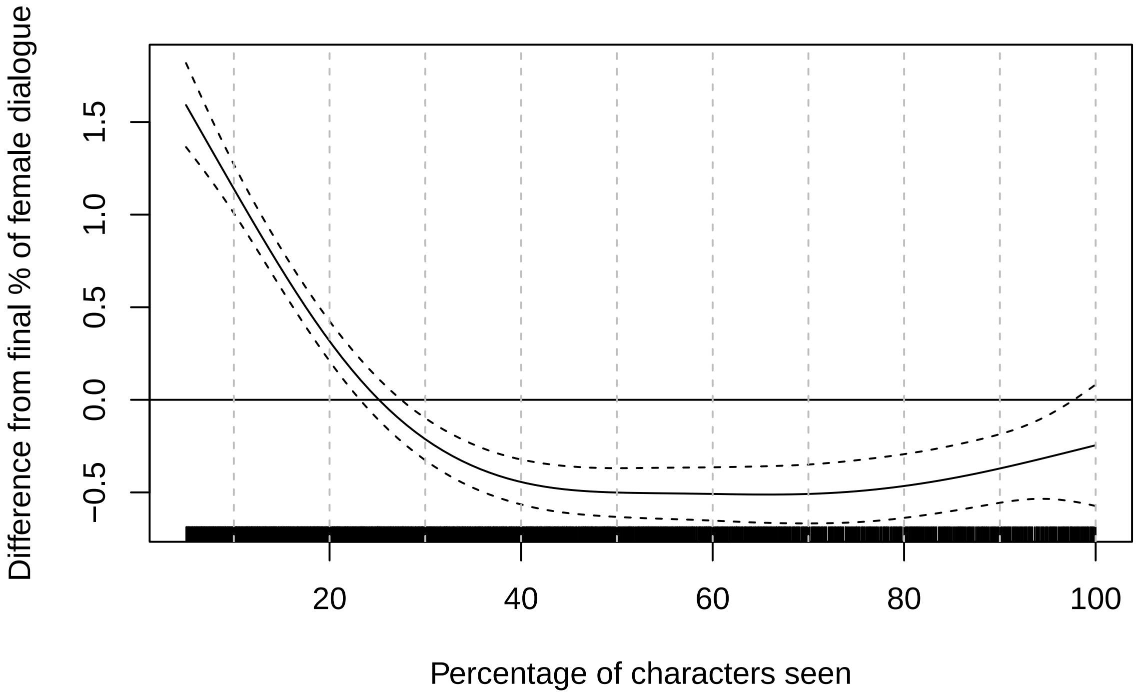
```
px = predict(mBeta, newdata = data.frame(
  percentCharCumulative = seq(0,100,by=5)
))
px.var = predict(mBeta, type="variance", newdata = data.frame(
  percentCharCumulative = seq(0,100,by=5)
))

allGames.Cum$folder.factor = factor(allGames.Cum$folder)
mGAM = gam(diffFromTotalEstimate~
             s(percentCharCumulative) +
             s(folder.factor, bs = 're'),
           data = allGames.Cum[allGames.Cum$percentCharCumulative>5,])
summary(mGAM)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## diffFromTotalEstimate ~ s(percentCharCumulative) + s(folder.factor,
##     bs = "re")
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1051     0.6181    0.17    0.865
##
## Approximate significance of smooth terms:
##                             edf Ref.df     F p-value
## s(percentCharCumulative)  4.354  5.355 60.05  <2e-16 ***
## s(folder.factor)         48.616 49.000 62.21  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.347   Deviance explained = 35.3%
## GCV = 7.6505  Scale est. = 7.5847    n = 6276
```

The model suggests that there is a significant relationship overall. We can visualise the curve to get an idea of the trends:

```
plot.gam(mGAM, xlab="Percentage of characters seen",
     ylab = "Difference from final % of female dialogue",
     select=1)
abline(h=0,col=1)
```

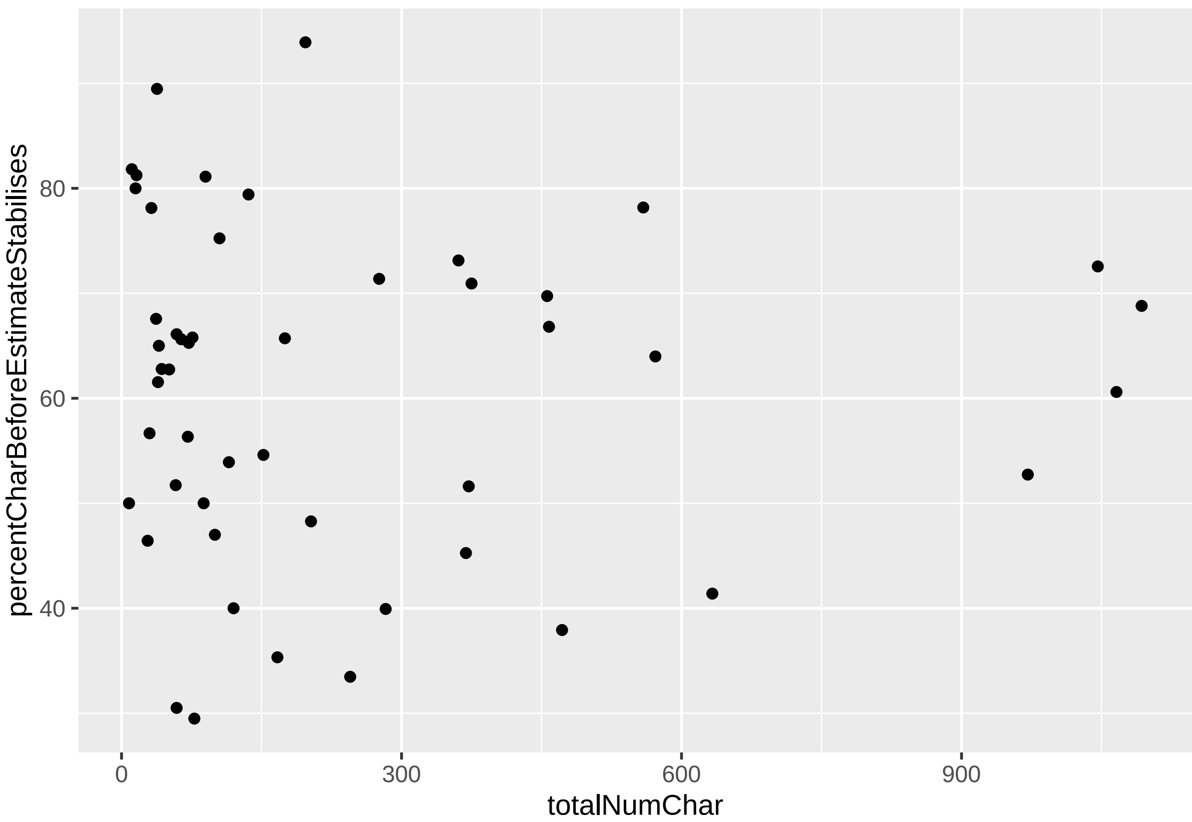
```
abline(v=seq(0,100,by=10), lty=2,col="gray")
```



In contrast to the boxplot, the visualisation suggets that female dialogue tends to be *over-estimated* for prolific characters, then slightly under-estimated after seeing about between 20-50% of characters. After about 60% of characters, the estimate stabilises. Note that the range is very small - predicting biases of around 1% at most.

The plot below shows how the stabilisation point varies with the total number of characters in the game. There's no strong relationship between the two.

```
ggplot(data=allGames.Cum[!duplicated(allGames.Cum$folder),],
       aes(y=percentCharBeforeEstimateStabilises,
           x =totalNumChar)) +
  geom_point()
```




```
cor.test(allGames.Cum[!duplicated(allGames.Cum$folder),]$percentCharBeforeEstimateStabilises,
         allGames.Cum[!duplicated(allGames.Cum$folder),]$totalNumChar)
```

```
##
##  Pearson's product-moment correlation
##
## data:  allGames.Cum[!duplicated(allGames.Cum$folder), ]$percentCharBeforeEstimateStabilises and allGa
## t = -0.091829, df = 48, p-value = 0.9272
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.2905291  0.2660761
## sample estimates:
##         cor
## -0.01325316
```

Conclusion

The analyses above suggest that the gender bias in video game dialogue is present in major and minor characters. Although the bias was stronger for major characters, this can be mostly explained by the low number of female characters in this group, rather than a systematic difference between major and minor character groups *per-se*.

Furthermore, the estimate of the proportion of female characters is not systematically biased across games due to being able to get more complete or accurate data on more prolific characters. Individual games were biased in different directions, but with the majority converging after seeing 90% of the characters. The estimate of female dialogue across all games was not different from the final estimate after seeing the top 60% prolific characters in each game. For some ranges of the data, the average estimate of female dialogue across all games was an *under-estimate*, rather than an over-estimate. In any case, the bias in the estimate after seeing 20% of the characters are in the range of a few percentage points. This is much smaller than the overall gender bias in the corpus.

In conclusion, it's unlikely that the gender biases reported in the main data are affected by a tendency to have more complete or accurate data on more central characters.

6 Player perceptions of gender bias

Player perceptions of gender balance

Introduction

The aim of this survey is to identify what people's perceptions are regarding the balance of dialogue between male and female characters in role-playing video games. This is part of a larger project that involves collecting a corpus of video game dialogue and making objective measures of gender bias. We've found that almost all games have more male dialogue than female dialogue, even when the main protagonists are female.

The purpose of the current study is to see whether the objective measures of gender bias align with gamer's subjective intuitions. This will help us frame the added value of objective corpus-based measures.

We have objective findings from 50 video games that demonstrate a clear bias against female dialogue. The main findings are:

- Games have an average of around 30% female dialogue.
- Very few games have more than 50% female dialogue.
- The proportion of female dialogue is increasing over time.

The first two statistics surprised us, suggesting that the severity of the bias isn't well known. But we want to be able to demonstrate this objectively.

We now want to compare this objective measure with people's subjective perceptions. The survey asks participants to guess what the main findings are. However, people may be aware of an imbalance without thinking it is necessarily a problem. So there is also a section that asks whether the participant thinks that male and female characters are under represented, and whether the representation needs to be improved.

We expect that people's intuitions may be influenced by their experience with role-playing games and identity (age category, gender). Therefore, we collected basic information about these aspects to control for these influences.

This study gained ethical approval from Cardiff University's ENCAP Research Ethics Committee (SREC reference: ENCAP/Roberts/17-01-2022).

Methodology

188 participants were recruited via online platforms including twitter, facebook, reddit (r/JRPG, r/rpg_gamers, r/FinalFantasy), Cardiff University yammer, Cardiff University Gaming Society Discord, and University of Glasgow Games and Gaming Lab. They followed a link that redirected them randomly to one of two surveys. The surveys were identical, except one asked questions framed in terms of the proportion of female dialogue, and one asked questions in terms of the proportion of male dialogue. Below, text in square brackets show how these two options varied.

Participants confirmed they were older than 18 years of age, then answered the following questions:

1. Age category
2. What is your gender? (free text)
3. How much experience do you have playing Role-Playing video games (RPGs)? (I've never played an RPG, I occasionally play RPGs, I play an RPG at least once a month, I play an RPG at least once a week)

4. What do you think? Let's say we counted all the words spoken by male and female characters (playable characters or NPCs) in a typical game. What percentage of words would be spoken by [female/male] characters? (number between 0 and 100)
5. What percentage of RPG games do you think have more [female/male] dialogue than [male/female] dialogue? (number between 0 and 100)
6. This question takes a bit more thought. Imagine we picked 10 popular RPG games. How many games do you think would have ...

- ... between 0% and 10% [female/male] dialogue?
- ... between 10% and 30% [female/male] dialogue?
- ... between 10% and 30% [female/male] dialogue?
- ... between 50% and 70% [female/male] dialogue?
- ... between 70% and 90% [female/male] dialogue?
- ... between 90% and 100% [female/male] dialogue?

5 point scales for strongly disagree, disagree, neither agree nor disagree, agree, strongly agree:

7. Do you disagree or agree with the following statements?

- [female/male] characters are under-represented in video games
- [male/female] characters are under-represented in video games
- The representation of [female/male] characters should be improved
- The representation of [male/female] characters should be improved

8. How do you think the proportion of [female/male] dialogue has changed over the last 20 years? (5 point scale: Increased a lot (more [female/male] dialogue now than before), Increased a bit, Stayed the same, Decreased a bit, Decreased a lot (less [female/male] dialogue now than before))

Responses for questions 4, 5 and 6 for the male-framed surveys were flipped to express the answers in terms of amount of female dialogue.

Below, we analyse the responses.

Analysis

Load survey data

```
library(ggplot2)
library(ggpubr)
library(cowplot)
library(viridis)
library(party)
library(rjson)
```

The survey data is in two files: one for each framing of the questions.

```
df = read.csv("surveyFemaleFraming.csv",stringsAsFactors = F)
dm = read.csv("surveyMaleFraming.csv", stringsAsFactors = F)

varNames = read.csv("varNames.csv",stringsAsFactors = F)
names(df) = varNames[match(names(df),varNames$SurveyQ),]$Label
names(dm) = varNames[match(names(dm),varNames$SurveyQ),]$Label
df = df[,names(dm)]

df$framing = "female"
dm$framing = "male"
```

```
d = rbind(df,dm)
```

Flip the responses for the male-framed condition:

```
d[d$framing=="male",]$percentFemaleDialogue = 100-d[d$framing=="male",]$percentFemaleDialogue
d[d$framing=="male",]$percentGamesMoreFemale = 100-d[d$framing=="male",]$percentGamesMoreFemale

d[d$framing=="female",]$FemaleChange20y = as.character(
  factor(d[d$framing=="female",]$FemaleChange20y,
         levels = c("Decreased a lot (less female dialogue now than before)",
                    "Decreased a bit",
                    "Stayed the same",
                    "Increased a bit",
                    "Increased a lot (more female dialogue now than before)"),
         labels = c("Decreased a lot",
                    "Decreased a bit",
                    "Stayed the same",
                    "Increased a bit",
                    "Increased a lot")))

d[d$framing=="male",]$FemaleChange20y = as.character(
  factor(d[d$framing=="male",]$FemaleChange20y,
         levels = c("Decreased a lot (less male dialogue now than before)",
                    "Decreased a bit",
                    "Stayed the same",
                    "Increased a bit",
                    "Increased a lot (more male dialogue now than before)"),
         labels = c("Increased a lot",
                    "Increased a bit",
                    "Stayed the same",
                    "Decreased a bit",
                    "Decreased a lot")))

d$FemaleChange20y = factor(d$FemaleChange20y,
  levels = c("Decreased a lot",
                    "Decreased a bit",
                    "Stayed the same",
                    "Increased a bit",
                    "Increased a lot"))
```

Quickly check the variation with framing using a t-test and a permutation test (since the distributions are skewed). The guesses of the percentage of female dialogue do not vary by framing, but the guesses of the percentage of games with more than 50% female dialogue do vary, which we address below.

```
t.test(d$percentFemaleDialogue~d$framing)
```

```
##
##  Welch Two Sample t-test
##
## data:  d$percentFemaleDialogue by d$framing
## t = 0.58012, df = 184.05, p-value = 0.5625
## alternative hypothesis: true difference in means between group female and group male is not equal to
## 95 percent confidence interval:
##  -2.429237  4.452820
## sample estimates:
```

```
## mean in group female   mean in group male 
##             31.36735             30.35556
```

```
trueDiff = diff(tapply(d$percentFemaleDialogue,d$framing,mean))
perm = function(){
  diff(tapply(sample(d$percentFemaleDialogue),d$framing,mean))
}
permDiff = replicate(1000,perm())
sum(permDiff>trueDiff)/1000
```

```
## [1] 0.728
```

```
t.test(d$percentGamesMoreFemale~d$framing)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  d$percentGamesMoreFemale by d$framing
## t = -3.0485, df = 172.11, p-value = 0.002663
## alternative hypothesis: true difference in means between group female and group male is not equal to
## 95 percent confidence interval:
##  -17.130426  -3.665492
## sample estimates:
## mean in group female   mean in group male 
##             19.10204             29.50000
```

```
trueDiff2 = diff(tapply(d$percentGamesMoreFemale,d$framing,mean))
perm2 = function(){
  diff(tapply(sample(d$percentGamesMoreFemale),d$framing,mean))
}
permDiff2 = replicate(1000,perm2())
sum(permDiff2>trueDiff2)/1000
```

```
## [1] 0.001
```

Flip the responses for question 6. We assume that the participants have an accurate, symmetrical understanding of the scale.

```
fcats = c("f10","f30","f50","f70","f90","f100")
d[d$framing=="male",fcats] = d[d$framing=="male",rev(fcats)]
```

Participants were given a free text box to describe their gender. We re-code categories that are obviously equivalent to 'male' and 'female' below (at least in terms of gender, rather than sex). There were only 4 participants who did not identify as either male or female. This is too few to analyses statistically. While their responses are analysed in general, they are not part of the analyses that focus on responses from male and female participants.

```
genderCoding = read.csv("genderCodes.csv",stringsAsFactors = F)
d$gender = tolower(d$genderString)
d[d$gender %in% genderCoding$string,]$gender = genderCoding[
      match(d[d$gender %in% genderCoding$string,]$gender,
      genderCoding$string),]$code
d$gender[d$gender==""] = NA
d$gender2 = d$gender
d[!is.na(d$gender2) & !(d$gender2 %in% c("female","male")),]$gender2 = "other"
```

Overall means:

```
percentFemaleDialogueGuessMean = mean(d$percentFemaleDialogue)
percentFemaleDialogueGuessSD = sd(d$percentFemaleDialogue)
percentGamesMore50FemaleMean = mean(d$percentGamesMoreFemale)
percentGamesMore50FemaleSD = sd(d$percentGamesMoreFemale)
```

The mean guess for the average percentage of female dialogue was 30.88% (sd = 12.02).

The mean guess for the percentage of games with more than 50% female dialogue was 24.08% (sd = 23.68).

```
cat(nrow(d),file="../../results/latexStats/Survey_N.tex")
cat(round(percentFemaleDialogueGuessMean),
    file="../../results/latexStats/Survey_percentFemaleDialogueGuessMean.tex")
cat(round(percentFemaleDialogueGuessSD,2),
    file="../../results/latexStats/Survey_percentFemaleDialogueGuessSD.tex")
cat(round(percentGamesMore50FemaleMean),
    file="../../results/latexStats/Survey_percentGamesMore50FemaleMean.tex")
cat(round(percentGamesMore50FemaleSD,2),
    file="../../results/latexStats/Survey_percentGamesMore50FemaleSD.tex")
```

Load game data

Obtain the true values from the corpus

```
folders = list.dirs("../../data", recursive = T)
folders = folders[sapply(folders,function(X){
  "stats_by_character.csv" %in% list.files(X)
})]


trueFemaleDialogue = c()
for(folder in folders){
  alternativeMeasure = FALSE
  js = fromJSON(file = paste0(folder,"/meta.json"))
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  if(!alternativeMeasure){
    stats = read.csv(paste0(folder,"/stats.csv"),stringsAsFactors = F)
    fprop = stats[stats$group=="female",]$words/
      (stats[stats$group=="female",]$words +
         stats[stats$group=="male",]$words)
    trueFemaleDialogue = c(trueFemaleDialogue,fprop)
  }
}
trueFemaleDialogueDist = trueFemaleDialogue[!is.na(trueFemaleDialogue)]
truePercentFemaleDialogue = mean(trueFemaleDialogueDist)
truePercentGamesMoreThan50 = sum(trueFemaleDialogueDist>=0.5)/length(trueFemaleDialogueDist)
```

Statistical test: is the mean guess significantly different from the true value?

```
mainTestPercentFemaleDialouge = t.test(d$percentFemaleDialogue,
                                       mu=truePercentFemaleDialogue*100)
mpx = "p < 0.0001"
if(mainTestPercentFemaleDialouge$p.value>0.0001){
  mpx = paste0("p = ",round(mainTestPercentFemaleDialouge$p.value,4))
}
```

```r
cat(paste0("t = ",round(mainTestPercentFemaleDialouge$statistic,2),", ",mpx),
    file = "../../results/latexStats/Survey_percentFemaleDialogue_Stat.tex")
mainTestPercentFemaleDialouge
```

```
##
##  One Sample t-test
##
## data:  d$percentFemaleDialogue
## t = -0.68151, df = 187, p-value = 0.4964
## alternative hypothesis: true mean is not equal to 31.48047
## 95 percent confidence interval:
##  29.15347 32.61249
## sample estimates:
## mean of x
##  30.88298
```

```r
mainTestPercentGamesMoreThan50 = t.test(d$percentGamesMoreFemale,
                                        mu=truePercentGamesMoreThan50*100)
mpx = "p < 0.0001"
if(mainTestPercentGamesMoreThan50$p.value>0.0001){
  mpx = paste0("p = ",round(mainTestPercentGamesMoreThan50$p.value,4))
}
cat(paste0("t = ",round(mainTestPercentGamesMoreThan50$statistic,2),", ",mpx),
    file = "../../results/latexStats/Survey_percentMoreThan50_Stat.tex")
mainTestPercentGamesMoreThan50
```

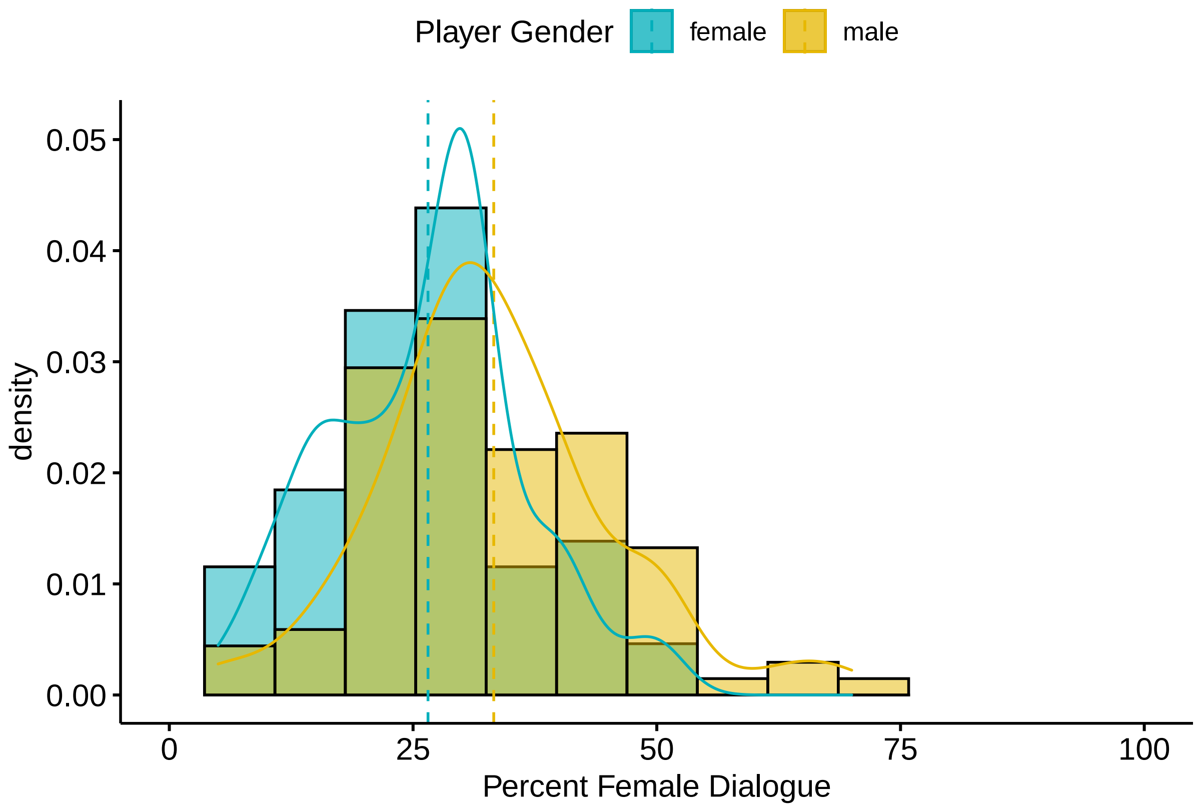
```
##
##  One Sample t-test
##
## data:  d$percentGamesMoreFemale
## t = 10.468, df = 187, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 6
## 95 percent confidence interval:
##  20.67269 27.48689
## sample estimates:
## mean of x
##  24.07979
```

Variation by gender

Below we visualise the data in various ways. In particular, we are interested in whether the estimates vary by the gender of the participant.

Distribution of guesses for the percentage of female dialogue:

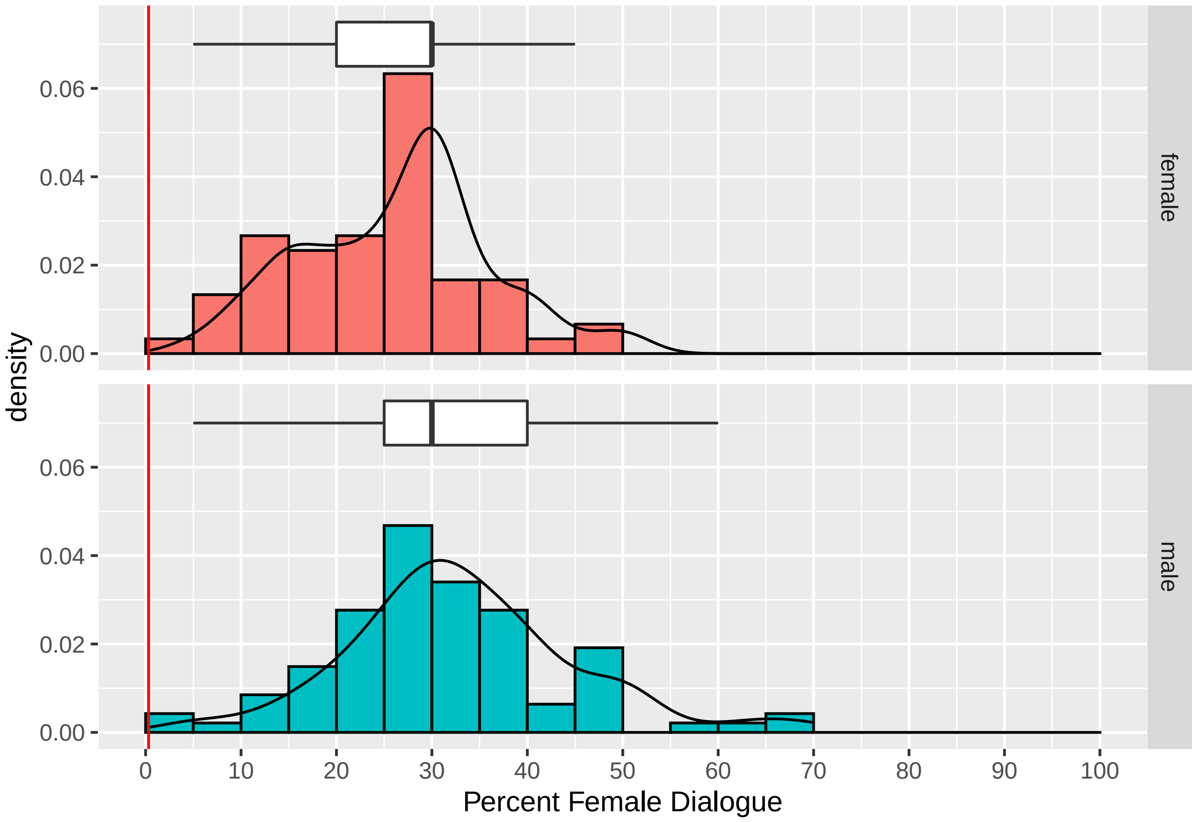
```
gghistogram(
  d[d$gender %in% c("female","male"),],
  bins = 10,
  x = "percentFemaleDialogue",
  y = "..density..",
  add = "mean", #rug = TRUE,
  fill = "gender", palette = c("#00AFBB", "#E7B800"),
  add_density = TRUE,
  xlim=c(0,100),
  legend.title="Player Gender") +
  xlab("Percent Female Dialogue")
```




```
ggplot(d[d$gender %in% c("female","male"),],
       aes(x=percentFemaleDialogue)) +
  geom_histogram(aes(y = ..density.., fill=gender),breaks=seq(0,100,5), color='black')+
  geom_density() +
  xlim(0,100) +
  facet_grid(rows=vars(gender)) +
  scale_x_continuous(breaks = seq(0,100,10)) +
  geom_boxplot(width=0.01,position=position_nudge(y=0.07),outlier.alpha = 0) +
  xlab("Percent Female Dialogue") +
```

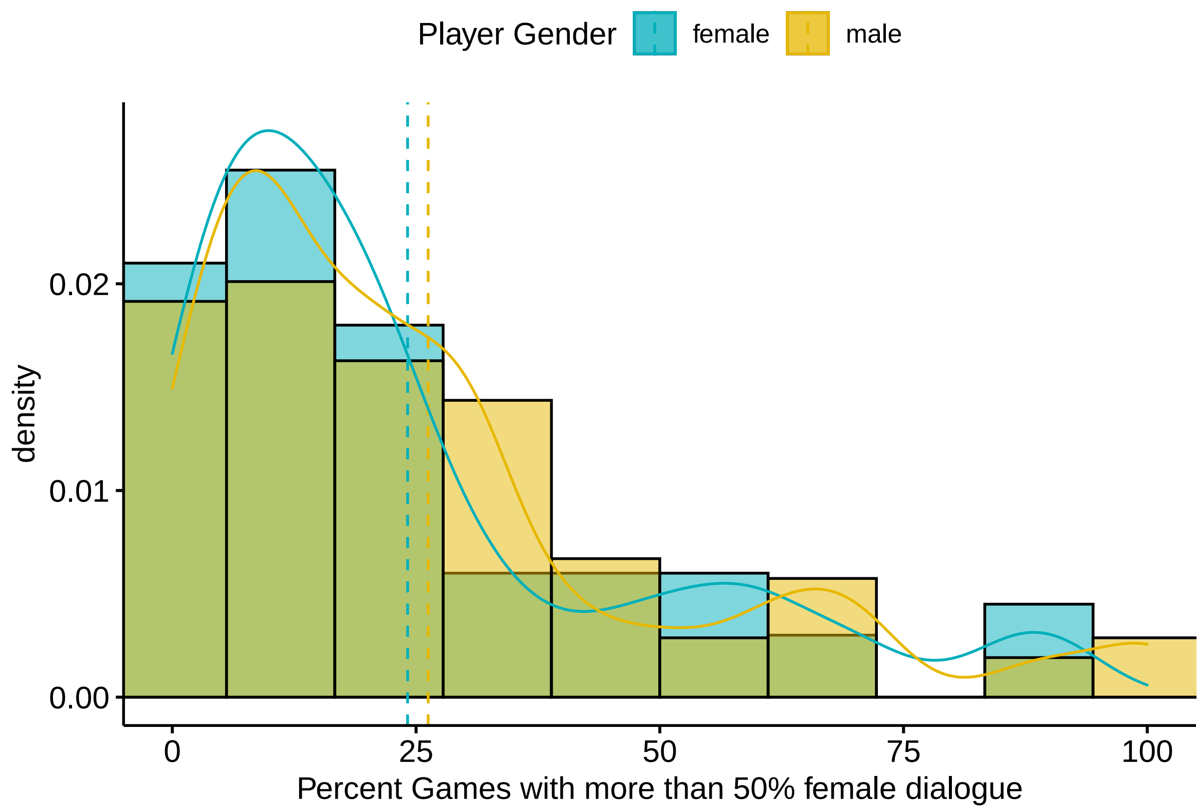
```
  geom_vline(xintercept = truePercentFemaleDialogue,color="red") +
  theme(legend.position = "none")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```



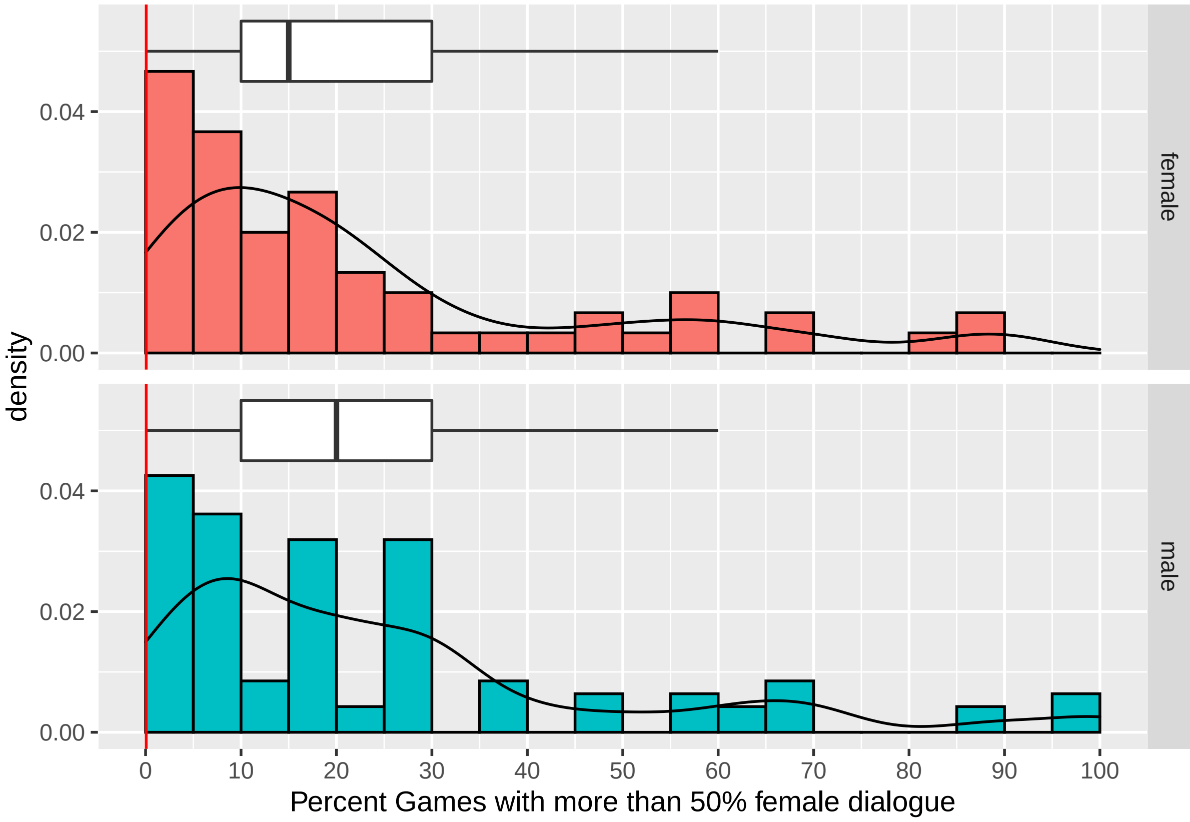
Distribution of guesses for the percentage of games with more than 50% female dialogue:

```
gghistogram(
  d[d$gender %in% c("female","male"),],
  bins = 10,
  x = "percentGamesMoreFemale",
  y = "..density..",
  add = "mean", #rug = TRUE,
  fill = "gender", palette = c("#00AFBB", "#E7B800"),
  add_density = TRUE,
  xlim=c(0,100),
  legend.title="Player Gender") +
  xlab("Percent Games with more than 50% female dialogue")
```




```
ggplot(d[d$gender %in% c("female","male"),],
       aes(x=percentGamesMoreFemale)) +
  geom_histogram(aes(y = ..density.., fill=gender),breaks=seq(0,100,5), color='black')+
  geom_density() +
  xlim(0,100) +
  facet_grid(rows=vars(gender)) +
  scale_x_continuous(breaks = seq(0,100,10)) +
  geom_boxplot(width=0.01,position=position_nudge(y=0.05),outlier.alpha = 0) +
  xlab("Percent Games with more than 50% female dialogue") +
  geom_vline(xintercept = truePercentGamesMoreThan50,color="red") +
  theme(legend.position = "none")
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```


0.04
0.02
0.00
female
density
0.04
0.02
0.00
male
0
10
20
30
40
50
60
70
80
90
100
Percent Games with more than 50% female dialogue

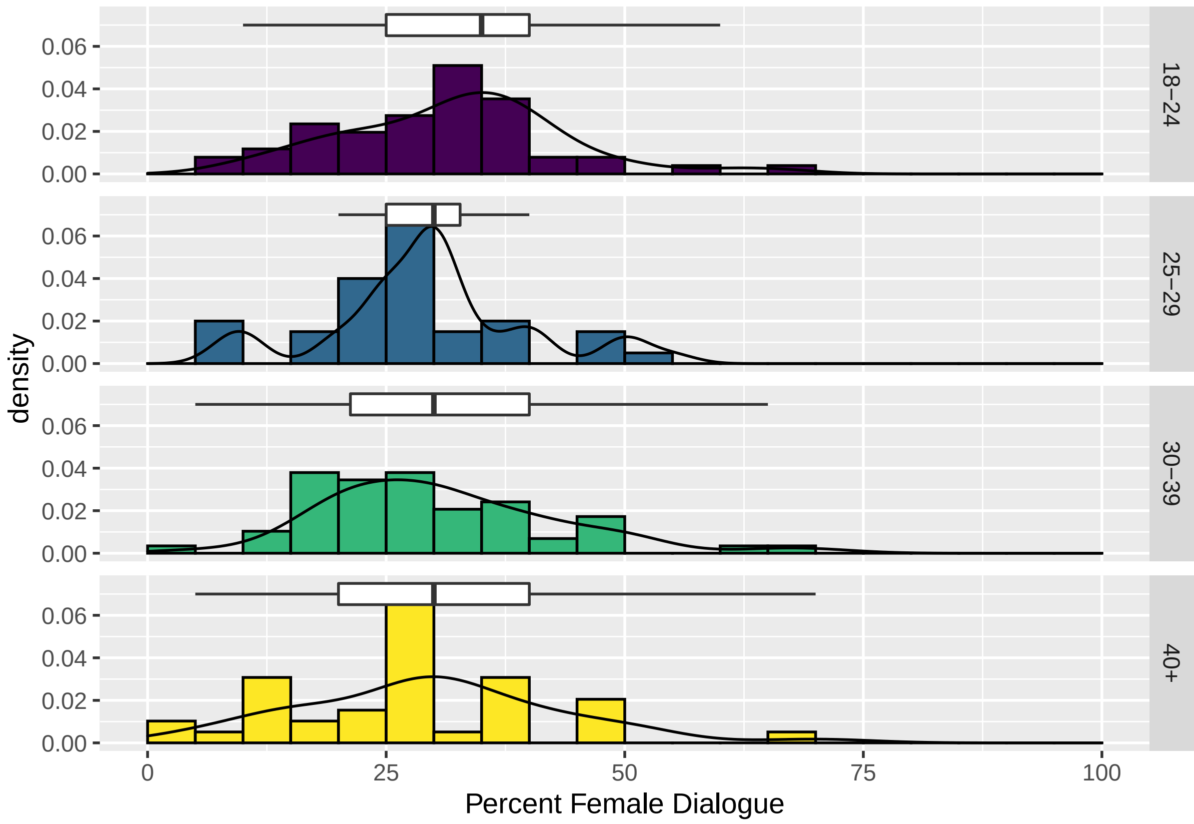
Variation by Age

Below we plot the distribution of guesses by the age of the respondent. One hypothesis is that older players may be more likely to have played earlier games compared to younger players (or less likely to have played newer games). Given that earlier games have less female dialogue, guesses of older players may be lower than younger players. Having said this, accessing games has become much easier in the last 10 years, so we may not expect a strong result.

There were relatively few participants over 40, so these are collapsed into a "+40" category.

The red dotted line shows the actual proportion in the corpus.

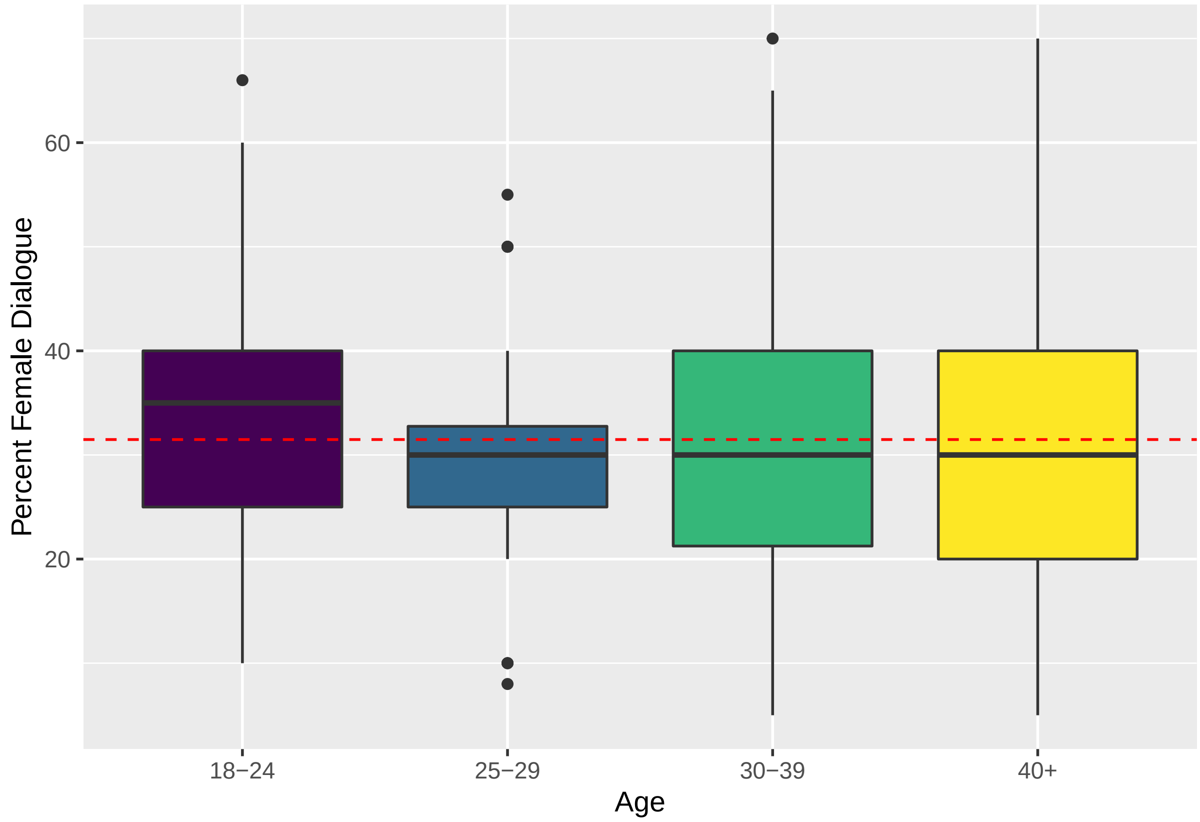
```
d$age2 = d$age
d$age2[d$age2 %in% c("40-49","50-59","60-69","70+")] = "40+"
d$age2 = factor(d$age2,ordered = T)

ggplot(d, aes(x=percentFemaleDialogue)) +
  geom_histogram(aes(y = ..density.., fill=age2),breaks=seq(0,100,5), color='black')+
  geom_density() +
  xlim(0,100) +
  facet_grid(rows=vars(age2)) +
  geom_boxplot(width=0.01,position=position_nudge(y=0.07),outlier.alpha = 0) +
  xlab("Percent Female Dialogue") +
  theme(legend.position = "none")
```




```
ggplot(d,aes(y=percentFemaleDialogue,x=age2,fill=age2)) +
  geom_boxplot() +
  xlab("Age") + ylab("Percent Female Dialogue") +
```

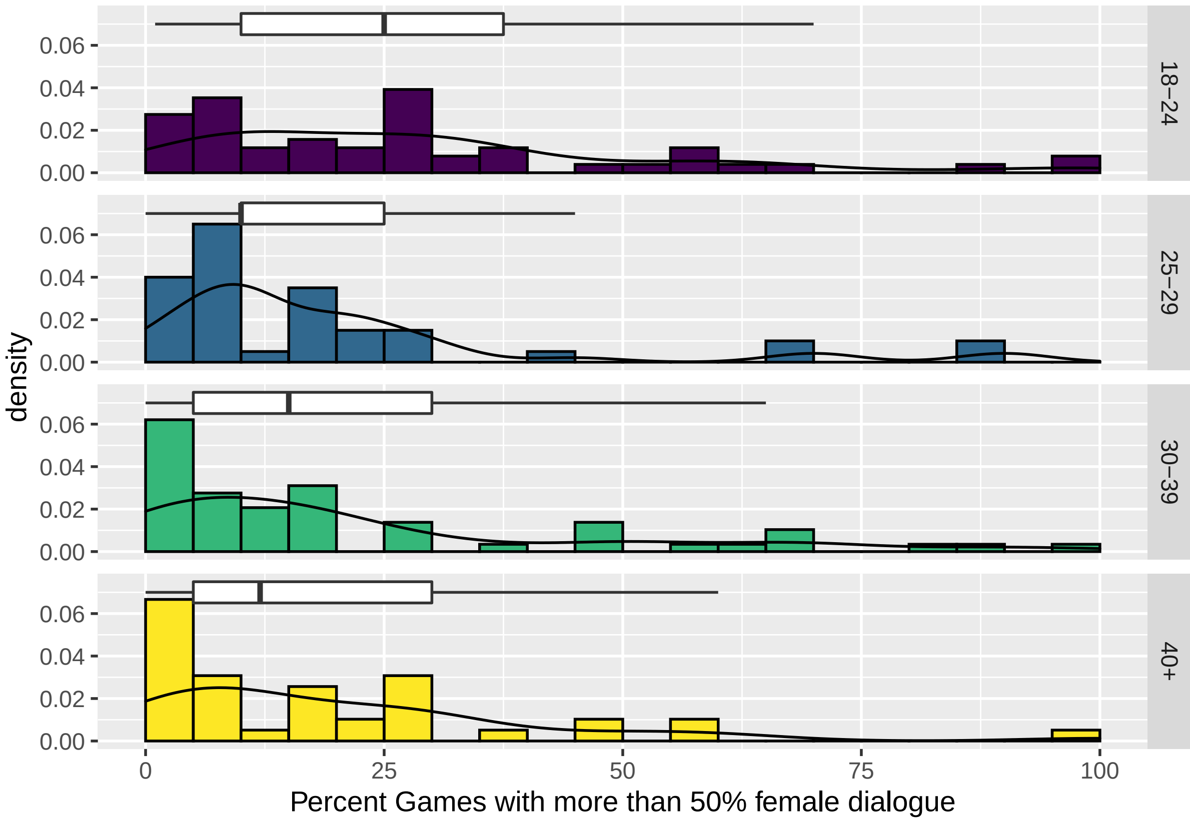
11

```
  theme(legend.position = "none") +
  geom_hline(yintercept = truePercentFemaleDialogue*100, color="red",linetype="dashed")
```



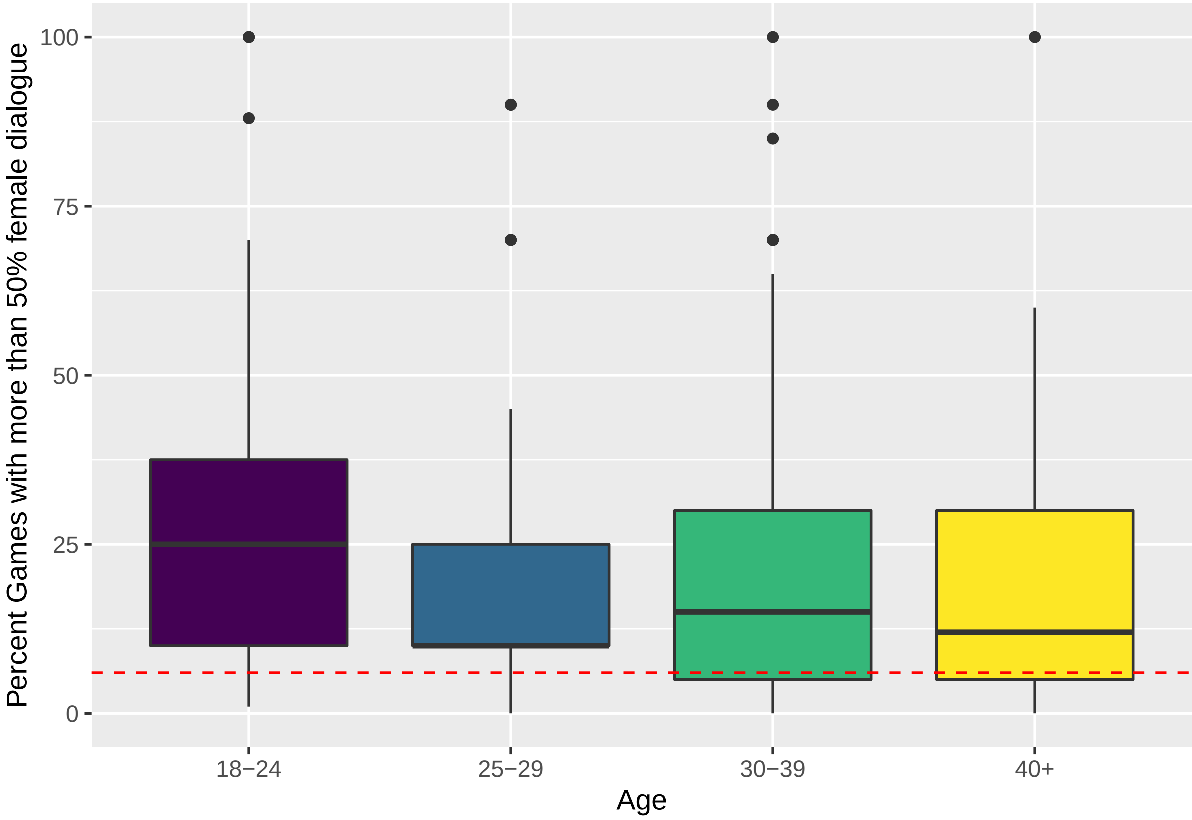
Similar plots for the proportion of games with more than 50% dialogue (red dotted line shows .

```
ggplot(d, aes(x=percentGamesMoreFemale)) +
  geom_histogram(aes(y = ..density.., fill=age2),breaks=seq(0,100,5), color='black')+
  geom_density() +
  xlim(0,100) +
  facet_grid(rows=vars(age2)) +
  geom_boxplot(width=0.01,position=position_nudge(y=0.07),outlier.alpha = 0) +
  xlab("Percent Games with more than 50% female dialogue") +
  theme(legend.position = "none")
```




```
ggplot(d,aes(y=percentGamesMoreFemale,x=age2,fill=age2)) +
  geom_boxplot() +
  xlab("Age") + ylab("Percent Games with more than 50% female dialogue") +
  theme(legend.position = "none") +
  geom_hline(yintercept = truePercentGamesMoreThan50*100, color="red",linetype="dashed")
```

13

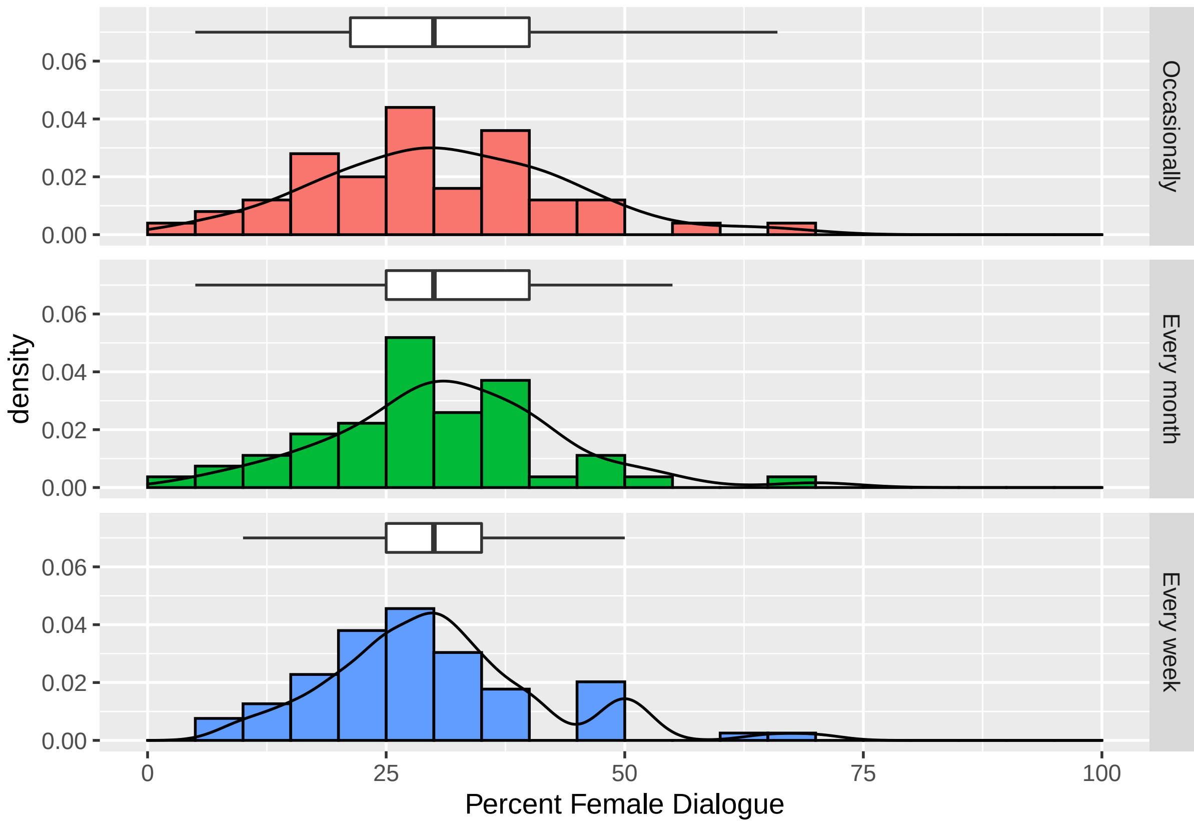

Percent Games with more than 50% female dialogue
100
75
50
25
0
18–24
25–29
30–39
40+
Age

Variation by experience

Below we plot the data by experience. One hypothesis is that more experienced players will have more accurate guesses. Since many players over-estimate the measures, this might lead to more experienced players having lower guesses. It might also predict a smaller variation for more experienced players

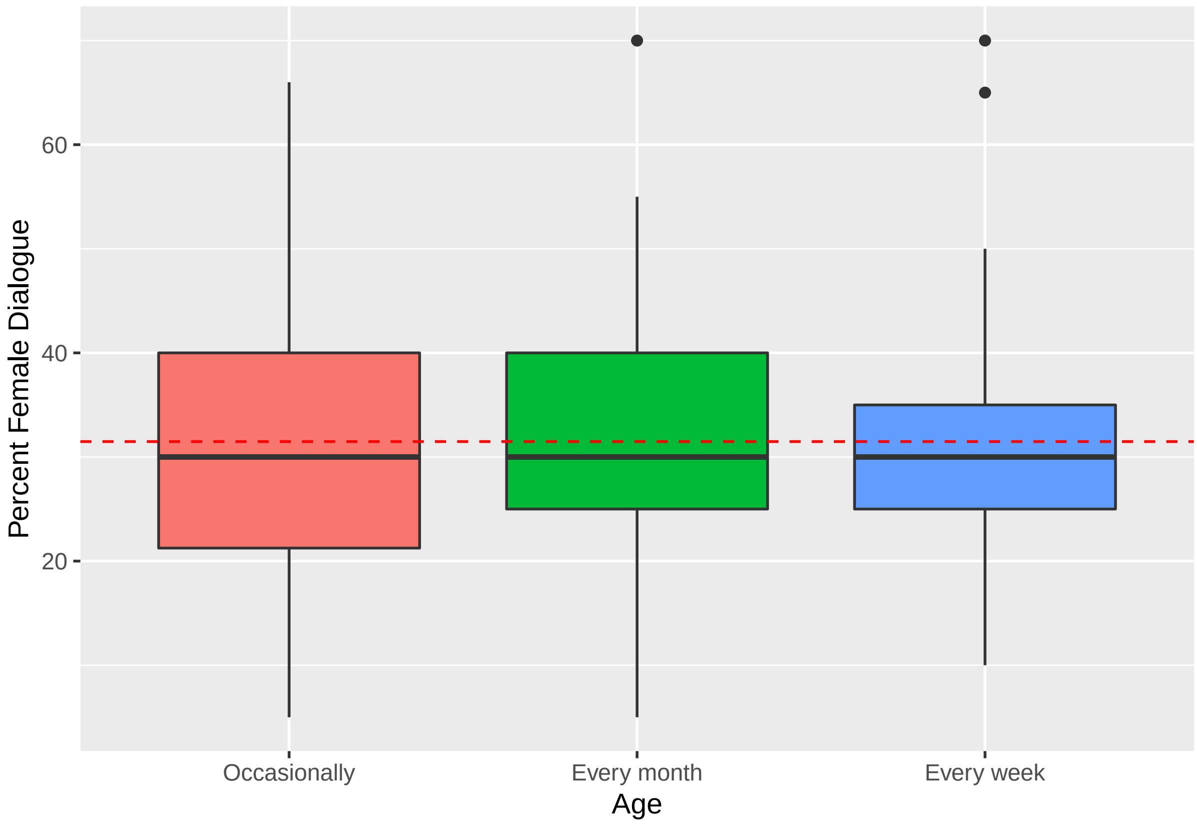
Female dialogue

```
d$rpgExperience = factor(d$rpgExperience,
    levels = c("I've never played an RPG", "I occasionally play RPGs",
        'I play an RPG at least once a month',"I play an RPG at least once a week"),
        labels = c("Never","Occasionally","Every month","Every week"))

ggplot(d[d$rpgExperience!="Never",],
    aes(x=percentFemaleDialogue)) +
  geom_histogram(aes(y = ..density.., fill=rpgExperience),breaks=seq(0,100,5), color='black')+
  geom_density() +
  xlim(0,100) +
  facet_grid(rows=vars(rpgExperience)) +
  geom_boxplot(width=0.01,position=position_nudge(y=0.07),outlier.alpha = 0) +
  xlab("Percent Female Dialogue") +
  theme(legend.position = "none")
```



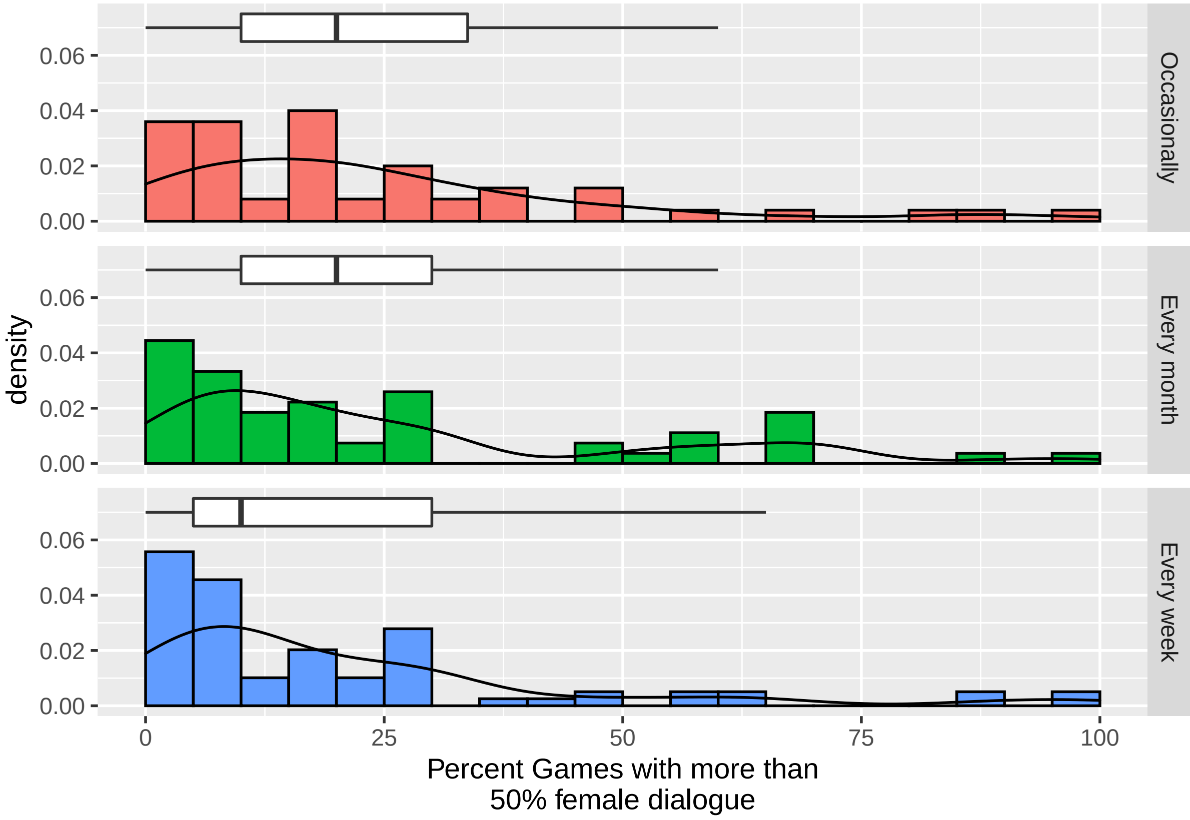

```
ggplot(d[d$rpgExperience!="Never",],
    aes(y=percentFemaleDialogue,x=rpgExperience,fill=rpgExperience)) +
  geom_boxplot() +
```

```
  xlab("Age") + ylab("Percent Female Dialogue") +
  theme(legend.position = "none") +
  geom_hline(yintercept = truePercentFemaleDialogue*100, color="red",linetype="dashed")
```

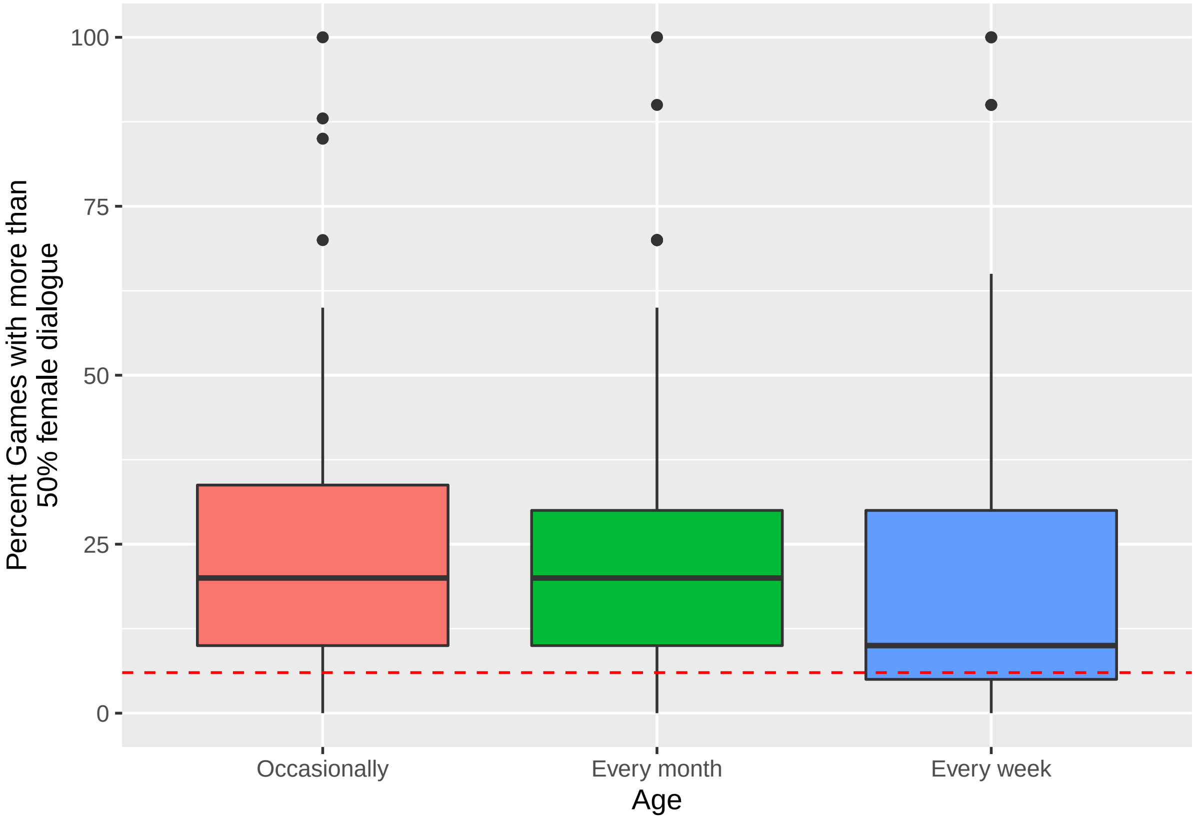


Games with more than 50% female dialogue

```
ggplot(d[d$rpgExperience!="Never",],
    aes(x=percentGamesMoreFemale)) +
  geom_histogram(aes(y = ..density.., fill=rpgExperience),breaks=seq(0,100,5), color='black')+
  geom_density() +
  xlim(0,100) +
  facet_grid(rows=vars(rpgExperience)) +
  geom_boxplot(width=0.01,position=position_nudge(y=0.07),outlier.alpha = 0) +
  xlab("Percent Games with more than\n50% female dialogue") +
  theme(legend.position = "none")
```




```
ggplot(d[d$rpgExperience!="Never",],
    aes(y=percentGamesMoreFemale,x=rpgExperience,fill=rpgExperience)) +
  geom_boxplot() +
  xlab("Age") + ylab("Percent Games with more than\n50% female dialogue") +
  theme(legend.position = "none") +
  geom_hline(yintercept = truePercentGamesMoreThan50*100, color="red",linetype="dashed")
```



Percent Games with more than 50% female dialogue
100
75
50
25
0
Occasionally
Every month
Every week
Age

Overall predictions

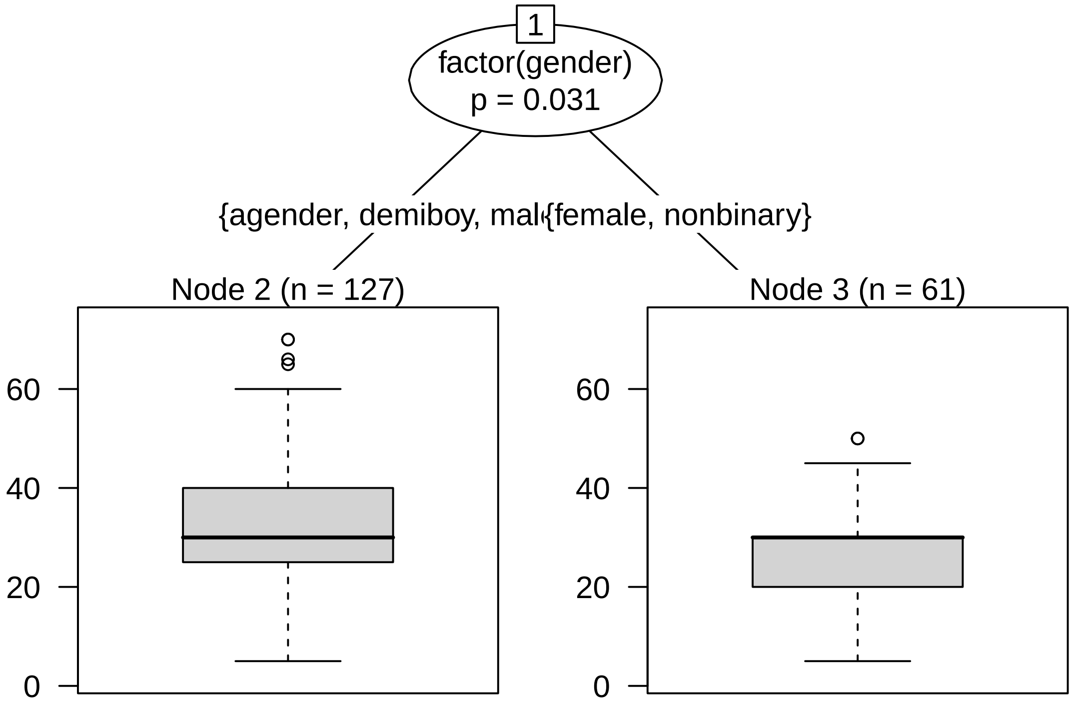
We use a regression model to identify what factors predict player guesses. For guesses of the percentage of female dialogue, player gender is the only siginficant predicting factor:

```
m0 = lm(percentFemaleDialogue ~
          age2 + framing + gender2 + rpgExperience,
        data=d)
summary(m0)
```

```
##
## Call:
## lm(formula = percentFemaleDialogue ~ age2 + framing + gender2 +
##     rpgExperience, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.667  -7.250  -0.685   4.915  38.285
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               17.12566    6.05862   2.827 0.005354 **
## age2.L                    -2.32361    1.89664  -1.225 0.222476
## age2.Q                     0.91678    1.86233   0.492 0.623255
## age2.C                    -1.56501    1.83536  -0.853 0.395202
## framingmale                0.07414    1.89685   0.039 0.968875
## gender2male                6.62937    1.90247   3.485 0.000649 ***
## gender2other               2.54725    6.00804   0.424 0.672201
## rpgExperienceOccasionally  9.83990    6.11429   1.609 0.109675
## rpgExperienceEvery month   9.56655    6.05323   1.580 0.116150
## rpgExperienceEvery week    9.40978    5.98262   1.573 0.117888
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.27 on 148 degrees of freedom
##   (30 observations deleted due to missingness)
## Multiple R-squared:  0.1145, Adjusted R-squared:  0.06062
## F-statistic: 2.126 on 9 and 148 DF,  p-value: 0.0307
```

We can use a decision tree that takes all the gender categories into account. This splits the data into two groups. The first is ‘female’ and ‘nonbinary’, who on average guess lower than the second group, ‘male’, ‘agender’ and ‘demiboy’.

```
tree = ctree(percentFemaleDialogue ~
          age2 + factor(framing) + factor(gender) + rpgExperience,
          data=d)
plot(tree)
```



Question framing seems to be the main predictor for percentage of games with more than 50% female dialogue. The guesses for the proportion of games with more than 50% dialogue are higher for participants asked to estimate the proportion of games with more *male* dialogue than female (their guesses were then inverted for the analyses above and below). This suggests that there is a framing bias.

```
mOB = lm(percentGamesMoreFemale ~
          framing,
        data=d)
summary(mOB)
```

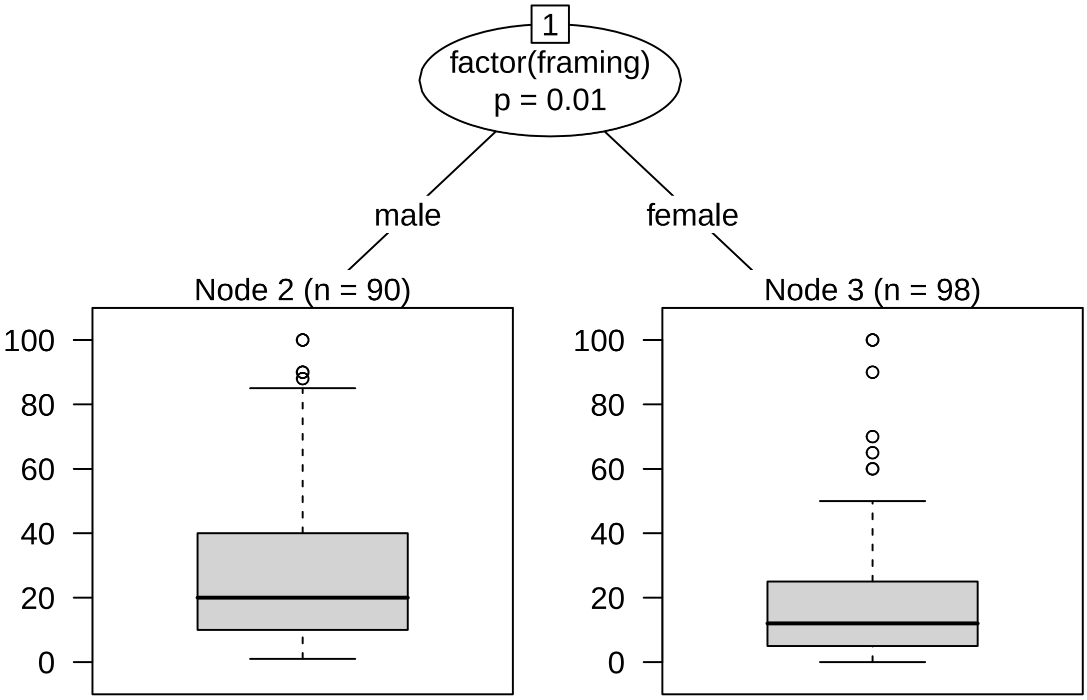
```
##
## Call:
## lm(formula = percentGamesMoreFemale ~ framing, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.500 -17.352  -9.102   7.048  80.898
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.102      2.340   8.164 4.79e-14 ***
## framingmale   10.398      3.382   3.075  0.00242 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.16 on 186 degrees of freedom
## Multiple R-squared:  0.04837,    Adjusted R-squared:  0.04325
```

```
## F-statistic: 9.454 on 1 and 186 DF,  p-value: 0.002424
```

```
m1B = lm(percentGamesMoreFemale ~
          age2 + framing + gender2 + rpgExperience,
        data=d)
summary(m1B)
```

```
##
## Call:
## lm(formula = percentGamesMoreFemale ~ age2 + framing + gender2 +
##     rpgExperience, data = d)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.939 -15.347  -6.780   7.942  77.728
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   4.604     12.599   0.365  0.71528
## age2.L                       -2.417      3.944  -0.613  0.54097
## age2.Q                        1.743      3.873   0.450  0.65331
## age2.C                       -5.579      3.817  -1.462  0.14592
## framingmale                  12.667      3.944   3.211  0.00162 **
## gender2male                   3.916      3.956   0.990  0.32381
## gender2other                 -8.105     12.493  -0.649  0.51750
## rpgExperienceOccasionally    11.136     12.714   0.876  0.38253
## rpgExperienceEvery month     15.748     12.587   1.251  0.21287
## rpgExperienceEvery week       9.846     12.441   0.791  0.42994
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.44 on 148 degrees of freedom
##   (30 observations deleted due to missingness)
## Multiple R-squared:  0.1129, Adjusted R-squared:  0.05899
## F-statistic: 2.094 on 9 and 148 DF,  p-value: 0.03349
```

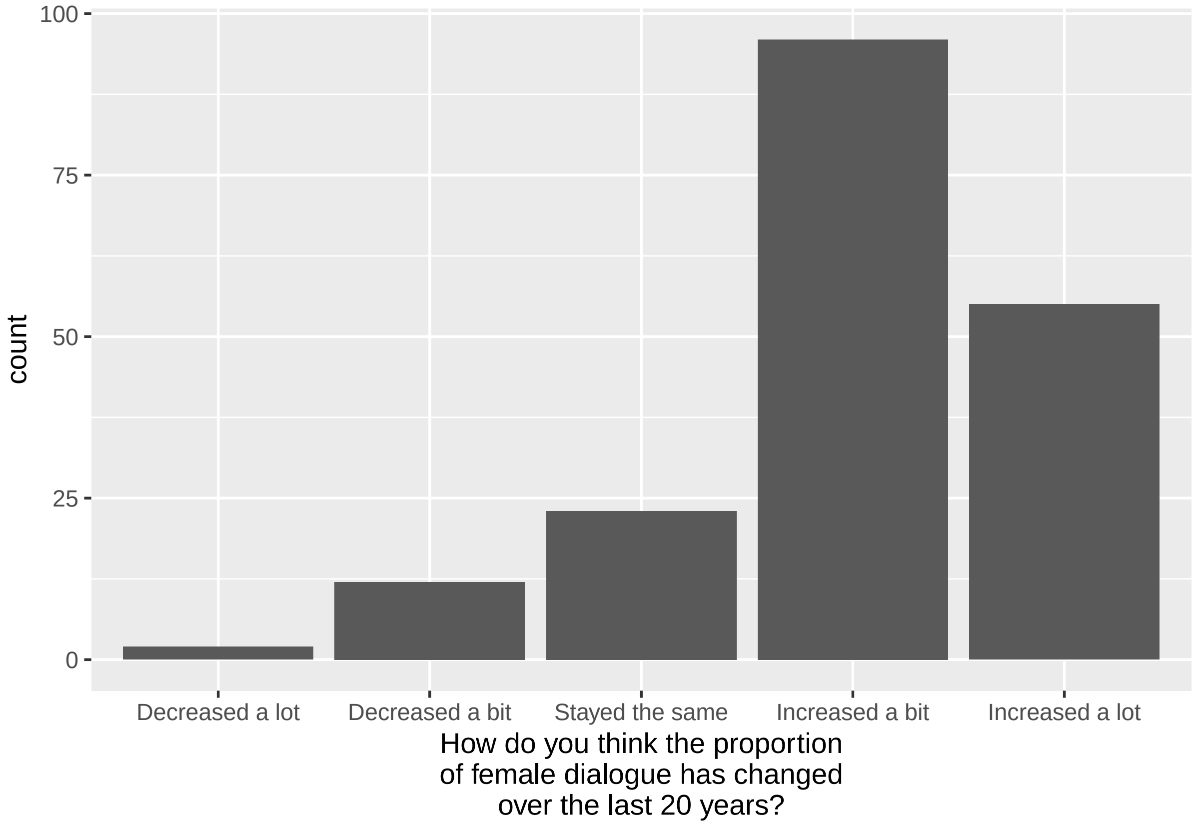
```
treeB = ctree(percentGamesMoreFemale ~
          age2 + factor(framing) + factor(gender) + rpgExperience,
          data=d)
plot(treeB)
```


1
factor(framing)
p = 0.01
male
female
Node 2 (n = 90)
Node 3 (n = 98)
100
80
60
40
20
0
100
80
60
40
20
0

Change over time

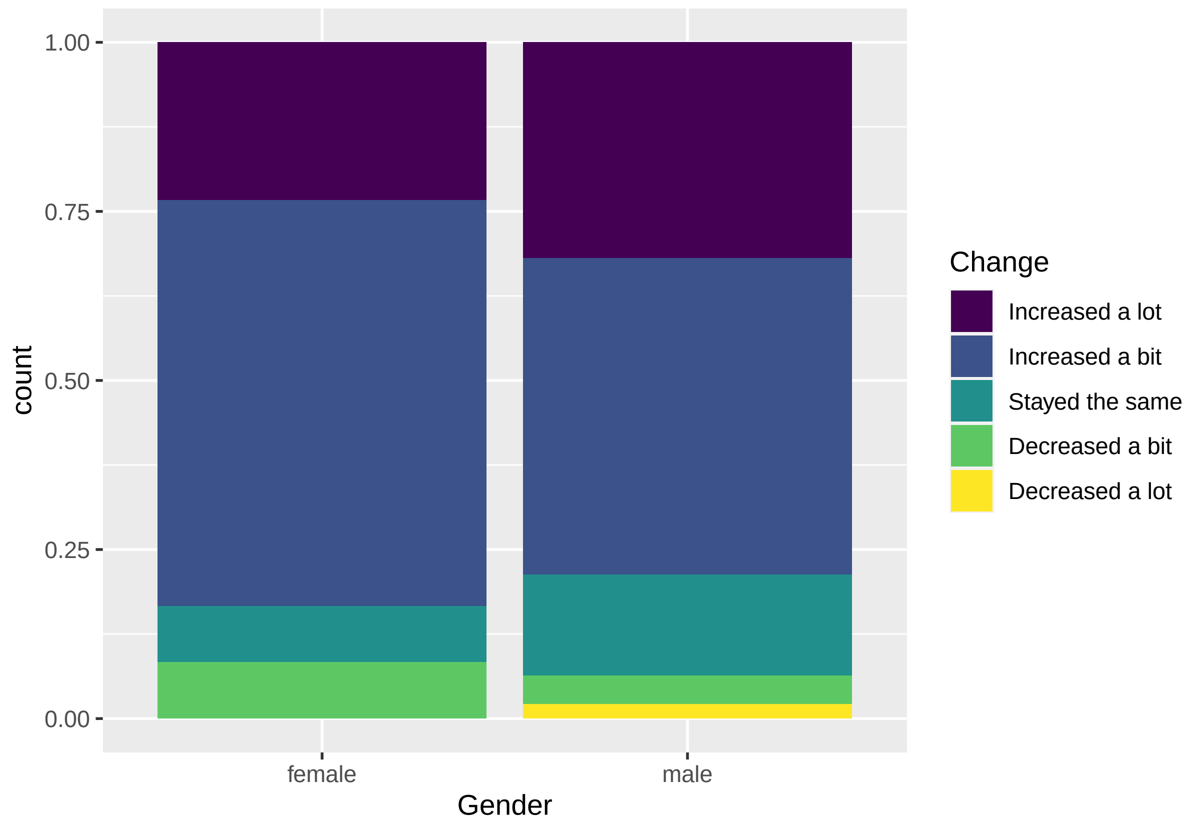
Summary of the responses to the question of “How do you think the proportion of female dialogue has changed over the last 20 years?”. Most respondents say it has increased a bit.

```
ggplot(d,aes(x=FemaleChange20y)) +
  geom_bar() +
  xlab("How do you think the proportion\nof female dialogue has changed\nover the last 20 years?")
```



Variation by gender:

```
d$Change = factor(d$FemaleChange20y,
  levels = c("Increased a lot",
                   "Increased a bit",
                   "Stayed the same",
                   "Decreased a bit",
                   "Decreased a lot"))

ggplot(d[d$gender2 %in% c("female","male"),],
       aes(x=gender2,fill=Change)) +
  geom_bar(position="fill") +
  xlab("Gender")+
  scale_fill_viridis_d()
```



Test whether the responses differ by gender:

```
changeTable = table(d[d$gender2 %in% c("female", "male"),]$FemaleChange20y,
                   d[d$gender2 %in% c("female", "male"),]$gender2)
changeTable
```

```
##
##                   female male
##   Decreased a lot      0    2
##   Decreased a bit      5    4
##   Stayed the same      5   14
##   Increased a bit     36   44
##   Increased a lot     14   30
```

```
fisher.test(changeTable)
```

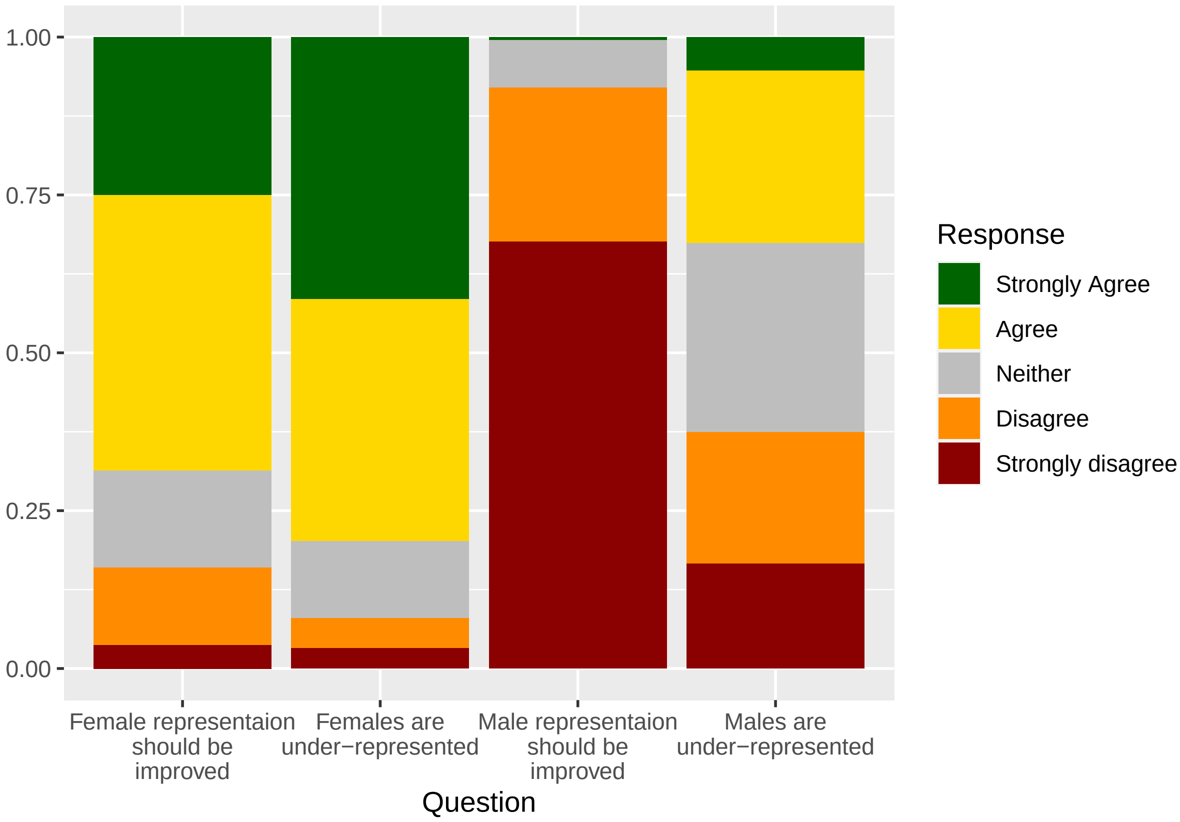
```
##
##  Fisher's Exact Test for Count Data
##
## data:  changeTable
## p-value = 0.2326
## alternative hypothesis: two.sided
```

No significant variation by gender.

Representation

Responses for the 5-point agreement scales.

```
OpLevels = rev(c("Strongly disagree","Disagree",
           "Neither agree nor disagree","Agree",
           "Strongly Agree"))
OpLables = rev(c("Strongly disagree","Disagree",
           "Neither","Agree",
           "Strongly Agree"))
d$OpFemaleRepImprove = factor(d$OpFemaleRepImprove,levels = OpLevels,labels=OpLables,ordered=T)
d$OpMaleRepImprove = factor(d$OpMaleRepImprove,levels = OpLevels,labels=OpLables,ordered=T)
d$OpFemaleUnderRepresented = factor(d$OpFemaleUnderRepresented,levels = OpLevels,labels=OpLables,ordere
d$OpMaleUnderRepresented = factor(d$OpMaleUnderRepresented,levels = OpLevels,labels=OpLables,ordered=T)

dx = data.frame(
  Response = c(
    d$OpFemaleRepImprove,
    d$OpMaleRepImprove,
    d$OpFemaleUnderRepresented,
    d$OpMaleUnderRepresented),
  Question = c(
    rep("Females are\nunder-represented",nrow(d)),
    rep("Males are\nunder-represented",nrow(d)),
    rep("Female representaion\nshould be\nimproved",nrow(d)),
    rep("Male representaion\nshould be\nimproved",nrow(d))
  )
)

ggplot(dx[!is.na(dx$Response),],
       aes(x=Question,fill=Response)) +
  geom_bar(position="fill") +
  scale_fill_manual(values=c("darkgreen","gold","gray","darkorange","darkred")) +
  theme(axis.title.y = element_blank())
```



The majority of respondents agreed that females are under-represented and represention should be improved. The majority disagreed that male representation should be improved. And there was a mixed response to whether males are under-represented. This suggests that there may be different factors and key issues for improving female representation and male representation.

Distribution

In the main analysis of the corpus, the proportion of female dialogue as a whole was not particularly surprising to the authors. The authors were more surprised by the distribution, with very few games havng more than 50% female dialogue. Question 6 of the survey included detailed questions that attempted to identify what players thought the distribution looked like.

6. This question takes a bit more thought. Imagine we picked 10 popular RPG games. How many games do you think would have …

- … between 0% and 10% [female/male] dialogue?
- … between 10% and 30% [female/male] dialogue?
- … between 10% and 30% [female/male] dialogue?
- … between 50% and 70% [female/male] dialogue?
- … between 70% and 90% [female/male] dialogue?
- … between 90% and 100% [female/male] dialogue?

Participants were free to specify any number between 0 and 10 in each of their responses, so their totals could add up to more than the 10 we asked for. We need to control for this so that participants specifying the distribution for 100 games doesn't bias the estimate.

Instead of working with raw numbers, we convert each particpant's responses into proportions across the scale:

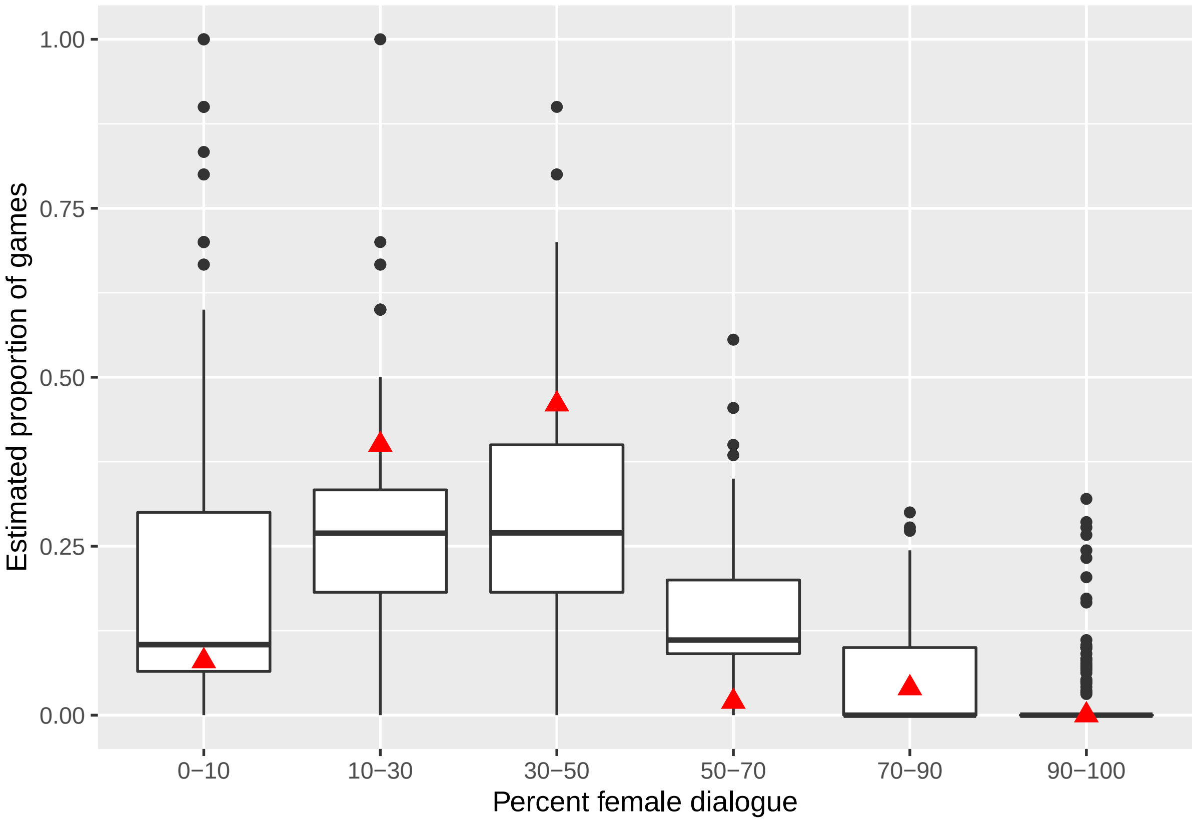
```
f = d[,grepl("f[1-9]0",names(d))]
rs = rowSums(f,na.rm = T)
fNormed = f/rowSums(f,na.rm=T)
fNormed[is.na(fNormed)] = 0
```

Collect the data:

```
rx = data.frame(
  n = unlist(f),
  p = unlist(fNormed),
  category.boundary = rep(c(10,30,50,70,90,100),each=nrow(f)),
  category.middle = rep(c(5,20,40,60,80,95),each=nrow(f)),
  category.names = rep(c("0-10","10-30","30-50","50-70","70-90","90-100"),each=nrow(f)),
  rowSum = rep(rs,6)
)

brks = c(10,30,50,70,90,101)
tn = table(cut(trueFemaleDialogueDist*100,c(0,brks)))
tp = tn/sum(tn)
trueDist = data.frame(
  p = as.vector(tp),
  category.middle = c(5,20,40,60,80,95),
  category.names = c("0-10","10-30","30-50","50-70","70-90","90-100"))
```

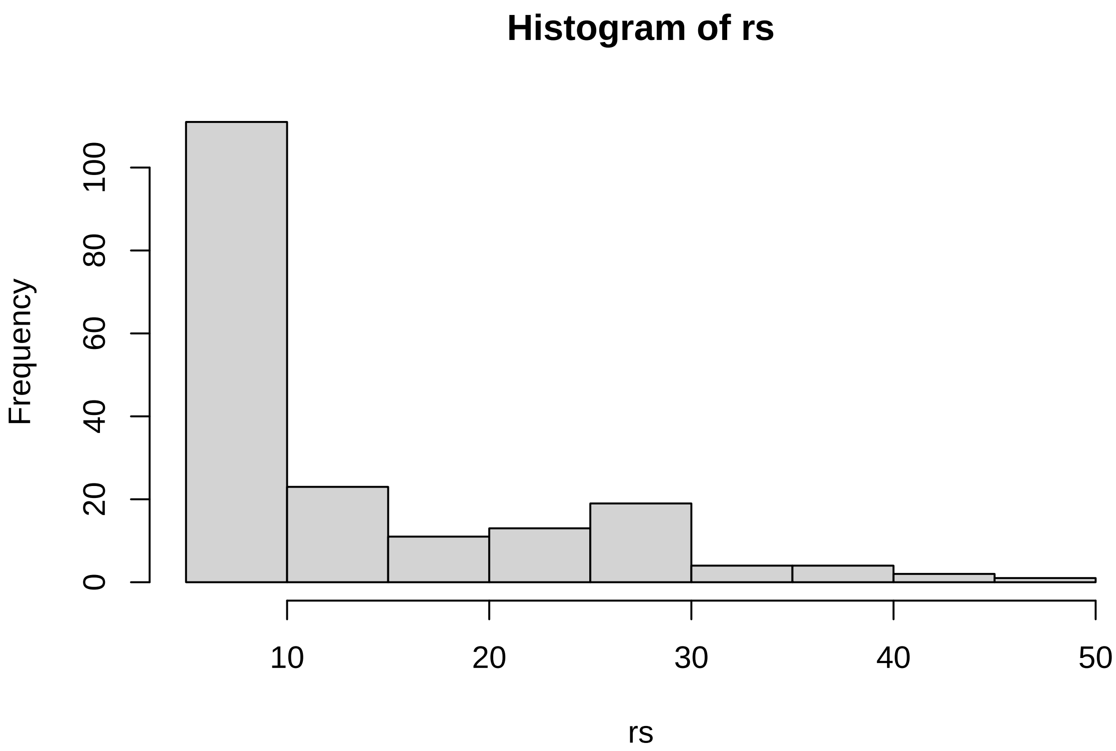
Plot the mean guesses as boxplots, with the true distribution indicated as red triangles

```
ggplot(rx, aes(x=factor(category.names),y=p))+
  geom_boxplot() +
  geom_point(data=trueDist,aes(y=p),stat="identity",
             colour="red",size=3,shape="triangle") +
  xlab("Percent female dialogue") +
  ylab("Estimated proportion of games")
```



Did people understand the question? What totals did they actual use?

```
hist(rs)
```

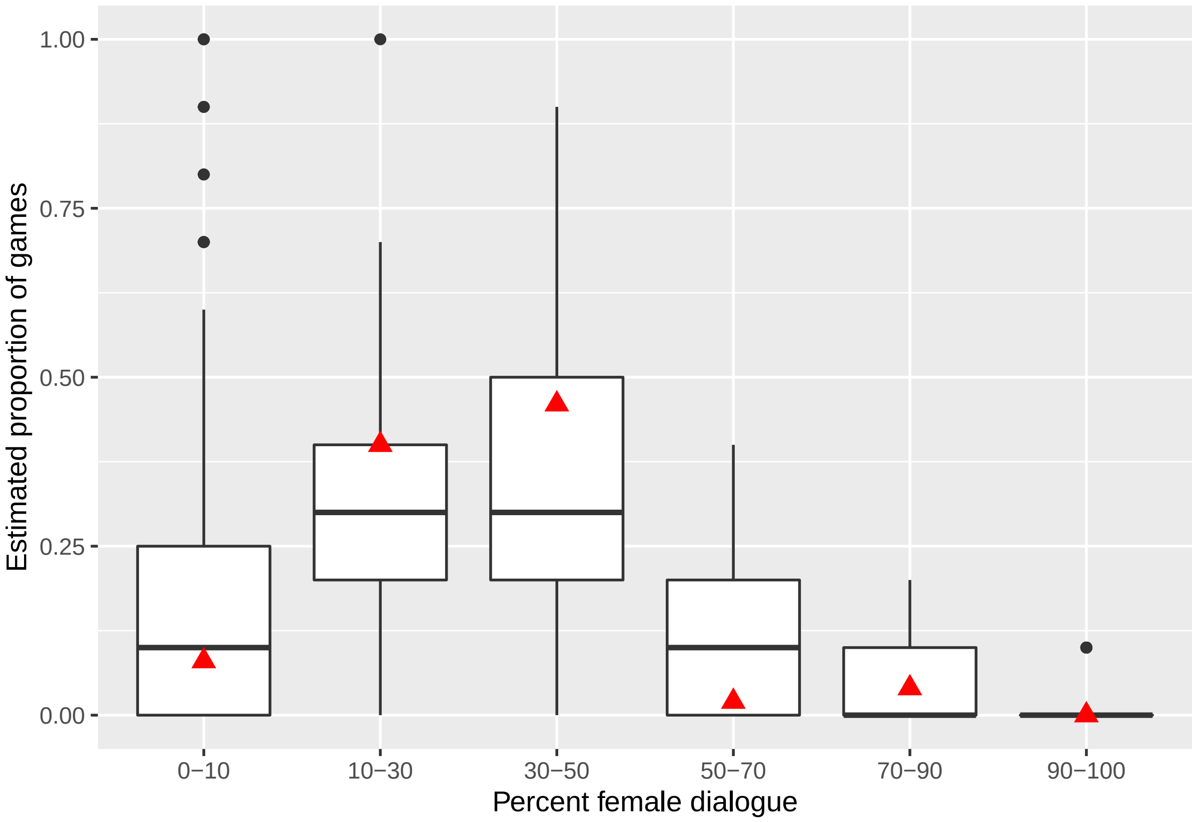



```r
# Proportion of participants responding with 10 games
sum(rs==10)/length(rs)
```

```
## [1] 0.5478723
```

Only about half of respondents used 10 items. So let's see the distribution for only participants that specified a total of 10:

```r
ggplot(rx[rx$rowSum==10,], aes(x=factor(category.names),y=p))+
  geom_boxplot() +
  geom_point(data=trueDist,aes(y=p),stat="identity",
             colour="red",size=3,shape="triangle")+
  xlab("Percent female dialogue") +
  ylab("Estimated proportion of games")
```



People seem to over-estimate the number of games that have more than 50% dialogue.

For each percentage category except 70-90, the mean guesses are significantly different from the true values:

```
distTest = data.frame()
for(i in 1:nrow(trueDist)){
  mx = rx$category.middle==trueDist[i,]$category.middle
  tt = t.test(rx[mx,]$p,
         mu=trueDist[i,]$p)
  distTest = rbind(distTest,data.frame(
    category = rx[mx,]$category.names[1],
    participantMean = tt$estimate, trueMean = trueDist[i,]$p,
    t = tt$statistic,p = tt$p.value
  ))
}
# Adjust for multiple comparisons
distTest$p = distTest$p * 6
distTest
```

```
##             category participantMean trueMean          t            p
## mean of x       0-10      0.21343244     0.08   7.918398 1.245801e-12
## mean of x1     10-30      0.27086989     0.40 -11.447707 1.468153e-22
## mean of x2     30-50      0.30081765     0.46 -11.725708 2.221112e-23
## mean of x3     50-70      0.13679580     0.02  15.784302 1.884724e-35
## mean of x4     70-90      0.05439421     0.04   2.873463 2.717884e-02
## mean of x5    90-100      0.02369000     0.00   5.626078 3.987537e-07
```

Write out stats for the 50-70 category for the paper

```
tx = distTest[distTest$category=="50-70",]
cat(round(100*tx$participantMean),file="../../results/latexStats/Survey_dist50_70_partGuess.tex")
cat(round(100*tx$trueMean),file="../../results/latexStats/Survey_dist50_70_true.tex")
px = "p < 0.0001"
if(tx$p>0.0001){
  px = paste0("p = ",round(tx$p,4))
}
cat(paste0("t = ",round(tx$t,2),", ",px)
  ,file="../../results/latexStats/Survey_dist50_70_stat.tex")
```

Conclusion

The mean guess for the average percentage of female dialogue was 30.88% (sd = 12.02). This is not signficiantly different from the empirical estimate. Male respondents estimated a higher percentage of female dialogue than female respondents.

The mean guess for the percentage of games with more than 50% female dialogue was 24.08% (sd = 23.68). This is significantly higher than the empirical estimate. The main predictor of this measure was the whether the survey was framed to ask about female or male dialogue.

7 Gender differences in emotion cues in Oblivion

Gender differences in emotion cues in *Oblivion*

Abstract

The emotional content of lines of dialogue from *Elder Scrolls IV: Oblivion* were analysed, testing various hypotheses about how their distribution might be biased by the gender of the speaking NPC. Support was found for the hypothesis that female NPCs are "backgrounded" (given a smaller range of emotional dialogue) and that emotions which are seen as more mascunline were more likely to be given to male NPCs, and conversely emotions seen as more feminine were more likely to be given to female NPCs.

Introduction

The data source for dialogue from *Elder Scrolls IV: Oblivion* (Bethesda Softworks, 2006) includes emotion and intensity cues for about 62,000 lines of dialogue. These were used by the game engine to dynamically animate the faces of the NPCs speaking the lines. They provide a window into the content of the dialogue and an opportunity to test how this content differs between genders.

The emotions are: Anger, Disgust, Fear, Happiness, Sadness, Surprise and 'Neutral'. The emotion intensity scores range from 0 to 100, though there are only 20 unique values. *Oblivion* contains dialogue for named NPCs (with a backstory, quests, etc.) and generic NPCs (e.g. guards). The main analysis only considers unique characters' lines of dialogue, since generic characters tended to have duplicated lines of dialogue for male and female NPCs.

We propose three hypotheses for how the distribution of emotions might be influenced by gender biases.

The first hypothesis derives from the observation that female characters are often "backgrounded" in video games. Carillo Masso (2011) studied imagery and dialogue from *Diablo* and *World of Warcraft*. It was found that masculine pronouns were used three times as frequently as feminine pronouns and there are no significant female characters in the main storyline of the game as portrayed in the game cinematics. Carillo Masso explains this using Fairclough (2003)'s notion of "backgrounding": the presence of female characters needs to be assumed and implied by the game player. Similarly, Miller and Summers (2007, p. 739), find that female characters were more often "supplemental" in video games.

For this study, gender-biased "backgrounding" would predict that female NPCs would play a smaller range of roles and less important roles. This suggests that their dialogue would be more likely to be "neutral" than male dialogue and less likely to have any of the other emotions than male dialogue. By having a higher proportion of neutral utterances compared to other emotions, female characters would be more frequently excluded from the main 'drama' and emotional narrative of the game, thus instead fading into the background.

The second hypothesis is that the game will reflect the tendency (or the belief) in the real world for females to publicly express more emotions than males (Kring & Gordon, 1998; Jansz, 2000; Timmers et al., 2003). This would predict the opposite of the first hypothesis: that male NPCs would be more likely to express neutral emotions and less likely to express other emotions compared to females (except perhaps anger, see e.g. Fischer & Evers, 2011).

The third hypothesis is based on the idea of gendered emotions: the distribution will reflect stereotypes and cultural beliefs about which emotions are more 'feminine' or 'masculine'. Stereotypically feminine emotions would include happiness, surprise, sadness, fear, and disgust, while sterotypically masculine emotions would include anger (see Brody & Hall, 2008). This would predict that female NPCs would be more likely to express all emotions except neutral and anger, compared to male NPCs. We note that this third hypothesis is independent of the first two: female dialogue might show less emotional diversity than male dialogue while at the same time males are more likely to display anger.

For the intensity scores, the “backgrounding” hypothesis would predict that female NPCs more “neutral” or “default” intensity. Alternatively, various studies (e.g. Robinson & Johnson, 1997) show that women are believed to express more intense emotions than men, which would predict that intensity scores for female NPCs would be higher than for male NPCs.

Examples of facial emotion, from https://www.youtube.com/watch?v=chalWNAUjp8

```
knitr::include_graphics("Emotions.png")
```

Anger

Disgust

Fear

Happy

Sad

Surprise

Load libraries

```
library(rjson)
library(ggplot2)
library(lmtest)
library(sjPlot)
library(pander)
library(knitr)
library(quanteda)
library(quanteda.textstats)
library(entropy)
library(brms)
```

Load data

Load data and select only lines of dialogue:

```
d = fromJSON(file="../../data/ElderScrolls/Oblivion/data.json")
d = d[[1]]

m = fromJSON(file="../../data/ElderScrolls/Oblivion/meta.json")
charGroups = m$characterGroups

emotions = data.frame(
  name = sapply(d,function(X){names(X)[1]}),
  cue = sapply(d,function(X){X[["_Emotion"]]}),
  gender = "Female",
  charType = "Unique",
  stringsAsFactors = F
)
dialogue = sapply(d,function(X){as.character(X[[1]])})
```

Assign gender, generic/unique and dialogue text to each observation from metadata:

```
emotions[emotions$name %in% charGroups$male,]$gender = "Male"
emotions[emotions$name %in%
           c(charGroups$GenericMale,
             charGroups$GenericFemale),]$charType = "Generic"
emotions[grepl("Generic",emotions$name),]$charType = "Generic"
```

`emotions` is now a data frame where each row is a line of dialogue and CharType is whether the character is a generic character or a named, unique character.

Parse emotion and intensity into seperate columns:

```
emotions$emotion = sapply(strsplit(emotions$cue," "),head,n=1)
emotions$intensity = sapply(strsplit(emotions$cue," "),tail,n=1)
emotions$intensity = as.numeric(emotions$intensity)
```

Select only unique characters (remove generic characters):

```
dialogue.generic = dialogue[emotions$charType=="Generic"]
emotions.generic = emotions[emotions$charType=="Generic",]

dialogue = dialogue[emotions$charType=="Unique"]
emotions = emotions[emotions$charType=="Unique",]
```

Some example sentences:

```
set.seed(106)
med = tapply(which(emotions$intensity==50),
```

```
              emotions[emotions$intensity==50,]$emotion,
              sample, size=1)

high = tapply(which(emotions$intensity==100),
              emotions[emotions$intensity==100,]$emotion,
              sample, size=1)
examp = data.frame(
  Emotion = rep(sort(unique(emotions$emotion)),each=2),
  Intensity = rep(c(50,100),times=length(sort(unique(emotions$emotion)))),
  dialogue = c(rbind(dialogue[med],dialogue[high])))
pander(examp)
```

Emotion	Intensity	dialogue
Anger	50	Order has always proved the stronger.
Anger	100	Ugh! Darkness, give me strength!
Disgust	50	I'm Krognak gro-Brok, landsman, and I don't much care who you are.
Disgust	100	Excuse me, the Empire doesn't run itself, you know. Submit a complaint to the usual department and I'm sure someone will take care of it.
Fear	50	Get to the Sigil Keep!
Fear	100	Bringer of Light, bless these wretches that they may see the path to your glory.
Happy	50	Don't worry, pal, you'll get used to losing after a while.
Happy	100	Capital! This may be the last piece of the puzzle. I need to spend more time with Savilla's Stone first.
Neutral	50	He's an odd one. I'm sure someone will take care of him soon.
Neutral	100	Now go, and may Sithis guide you in this new stage of your life's dark journey.
Sad	50	I can't make that deal. Even for you.
Sad	100	Armies can only mean one thing… more sorrow and more death.
Surprise	50	Yes, honored Listener? Do you have orders for me?
Surprise	100	They're using Hist Sap? And they claim to have brought a tree into Cyrodiil? Amazing. I can't imagine what the sap might do to non-Argonians.

Analysis

Emotion categories

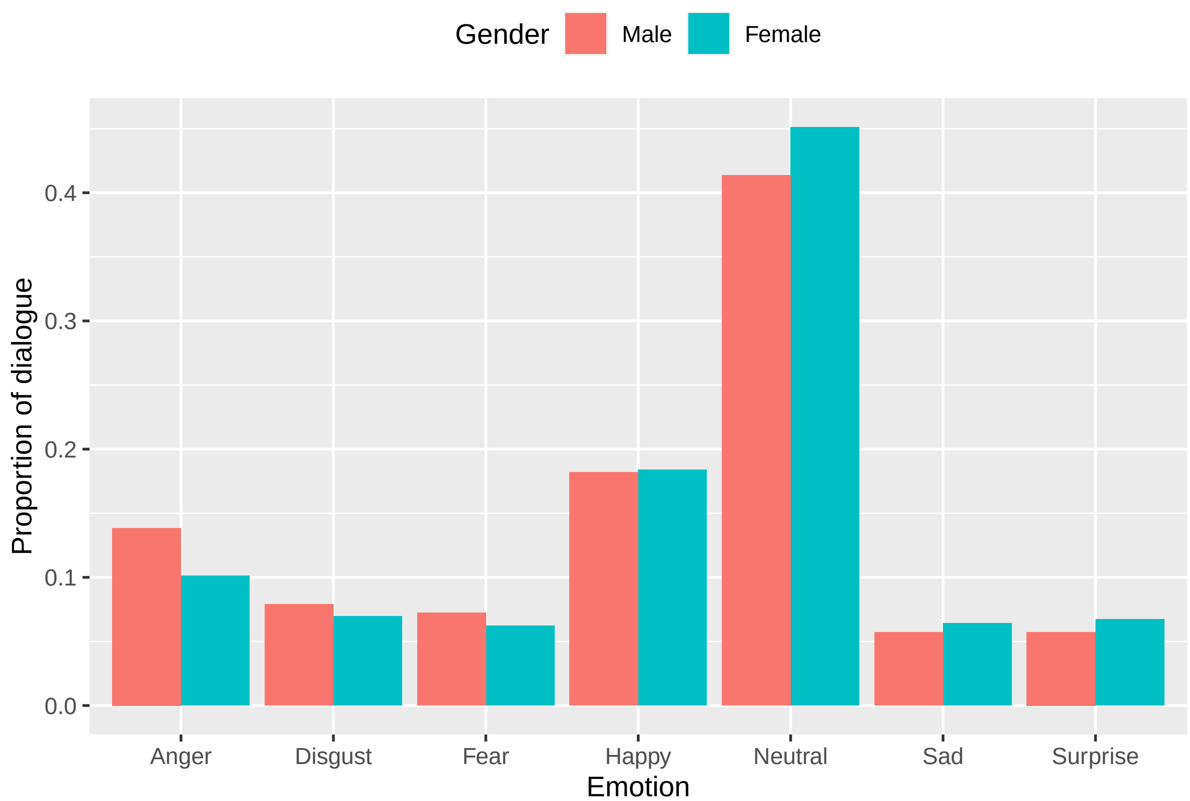
The tables below show the emotion cues assigned to male and female characters (raw numbers and proportionally by gender):

```
tab = table(emotions$emotion,emotions$gender)
colnames(tab) = paste(colnames(tab)," (N)")
propTab = prop.table(tab,margin=2)
colnames(propTab) = gsub("\\(N\\)","\\(%\\)",colnames(propTab))
kable(tab,"latex")
```

	Female (N)	Male (N)
Anger	978	2234
Disgust	674	1276
Fear	602	1166
Happy	1779	2936
Neutral	4362	6674
Sad	622	921
Surprise	650	925

```
kable(round(propTab*100,1),"latex")
```

	Female (%)	Male (%)
Anger	10.1	13.8
Disgust	7.0	7.9
Fear	6.2	7.2
Happy	18.4	18.2
Neutral	45.1	41.4
Sad	6.4	5.7
Surprise	6.7	5.7



Basic chi square test of proportion of emotion types by gender:

```
chiSqEmotion = chisq.test(tab)
chiSqEmotion
```

```
##
##  Pearson's Chi-squared test
##
## data:  tab
## X-squared = 118.5, df = 6, p-value < 2.2e-16
```

The distribution of emotions by gender is significantly different from what would be expected in a random distribution ($\chi^2(6) = 118.5$, p = 0.0001). Male NPCs have a higher proportion of anger, disgust, and fear dialogue compared to female NPCs. Female NPCs have a higher proportion of neutral, sad and surprise dialogue compared to male NPCs. The proportion of happy dialogue is roughly the same.

We can formally test the "backgrounding" hypothesis by comparing neutral and non-neutral dialogue:

```
emotions$neutral = factor(emotions$emotion=="Neutral",
                          levels=c(T,F),labels = c("Neutral","Non-neutral"))
tabN = table(emotions$neutral,emotions$gender)
kable(tabN,"latex")
```

	Female	Male
Neutral	4362	6674
Non-neutral	5305	9458

```
pt = round(prop.table(tabN,2)*100,1)
colnames(pt) = paste(colnames(pt),"(%)")
kable(pt,"latex")
```

	Female (%)	Male (%)
Neutral	45.1	41.4
Non-neutral	54.9	58.6

```
chiSqN = chisq.test(table(emotions$neutral,emotions$gender))
chiSqN
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(emotions$neutral, emotions$gender)
## X-squared = 34.599, df = 1, p-value = 4.051e-09
```

Female NPCs have a significantly more likely to have neutral emotion dialogue than male NPCs ($\chi^2(1) = 34.6$, $p < 0.0001$).

We can also formally test the gendered emotions hypothesis by creating a variable 'emotionGender' that is either masculine (for Anger) or feminine (for other non-neutral emotions). (this test is run on Non-neutral emotions only):

```
emotions$emotionGender = "feminine"
emotions$emotionGender[emotions$emotion=="Anger"] = "masculine"

tabEG = table(emotions[emotions$emotion!="Neutral",]$emotionGender,
          emotions[emotions$emotion!="Neutral",]$gender)
kable(tabEG,"latex")
```

	Female	Male
feminine	4327	7224
masculine	978	2234

```
ptEG = prop.table(tabEG,2)
colnames(ptEG) = paste(colnames(ptEG),"(%)")
kable(round(ptEG*100,1),"latex")
```

	Female (%)	Male (%)
feminine	81.6	76.4
masculine	18.4	23.6

```
chisqEG = chisq.test(tabEG)
chisqEG
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tabEG
## X-squared = 53.365, df = 1, p-value = 2.77e-13
```

There is a significant association between the gender of the NPC and masculinity/femininty of the emotion ($\chi^2(1) = 53.36$, $p < 0.0001$). Male NPCs are more likely to use masculine emotions (anger) than female NPCs, and female NPCs are more likely to use feminine emotions (happiness, sadness, surprise, fear and disgust).

Testing the difference between the "backgrounding" and gendered emotions hypotheses is hard. The "neutral" emotions aren't masculine or feminie, and the analysis above removes them from the data. But this is the key distinction for the backgrounding hypothesis. So the two variables can't be put together in a regression, and the two analyses above were based on different data, so can't be compared directly.

To address this, we create two regression models. The first model predicts the NPC's gender only based on neutral vs. non-neutral emotions. The second model uses a three-category variable that distinguishes "neutral" from "feminine" and "masculine". We then compare the fit of these models against each other. This effectively tests whether the prediction from the gendered emotions hypothesis helps predict the NPCs gender *over and above* the backgrounding hypothesis.

```
emotions$emotionGender[emotions$emotion=="Neutral"] = "neutral"
mNeutral = glm(gender=="Female" ~ neutral,
                data = emotions,family="binomial")
mEmGen = glm(gender=="Female" ~ emotionGender,
                data = emotions,family="binomial")
lrtest(mNeutral,mEmGen)
```

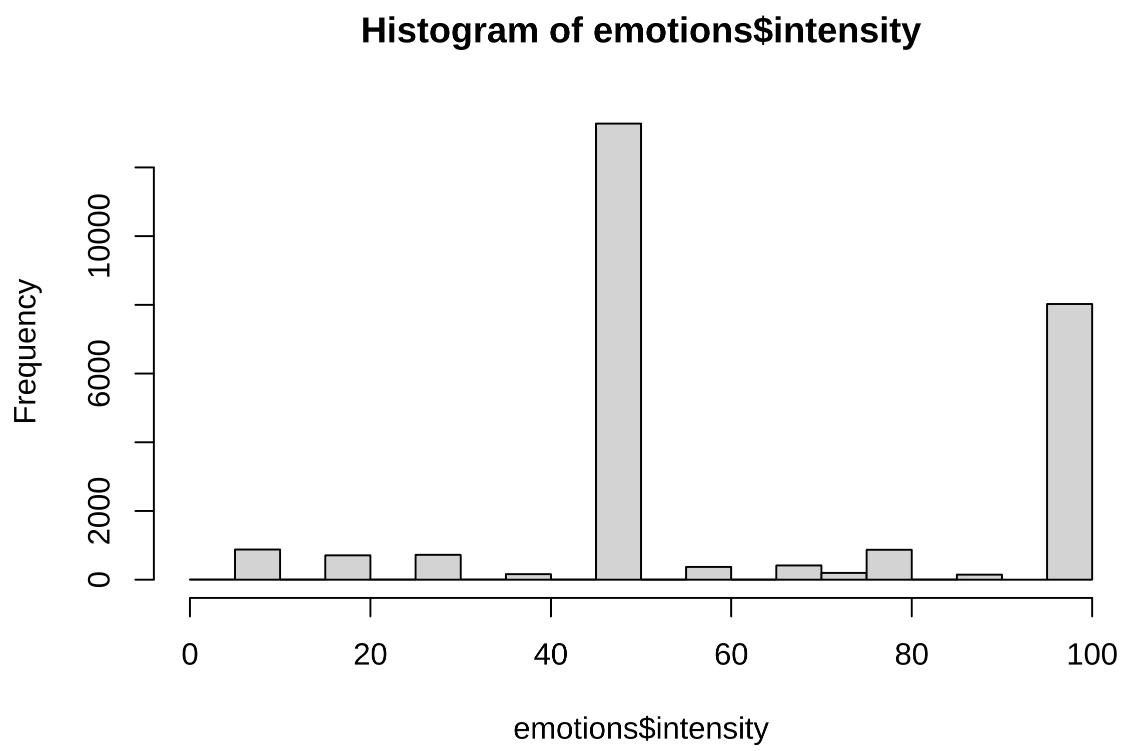
```
## Likelihood ratio test
##
## Model 1: gender == "Female" ~ neutral
## Model 2: gender == "Female" ~ emotionGender
##   #Df LogLik Df Chisq Pr(>Chisq)
## 1   2 -17046
## 2   3 -17019  1 54.67  1.425e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The improvement in the model fit for the second model is significant. This suggests that we should prefer the second model: the distribution of emotions is best explained by considering both neutral and gendered emotions (hypothesis 1 and 3).

Emotion intensity

The intensity of emotion cues in the Oblivion script is represented by a score from 0 – 100. However, there are only 20 unique values, and only 7 are used more than 1% of the time (see the histogram below).

```
emotions$intensityCategory =
  cut(emotions$intensity,
     breaks=c(-1,49,51,99,101),
     labels=c("Low","Medium","High","Max"))
hist(emotions$intensity,breaks=seq(0,100,by=5))
```



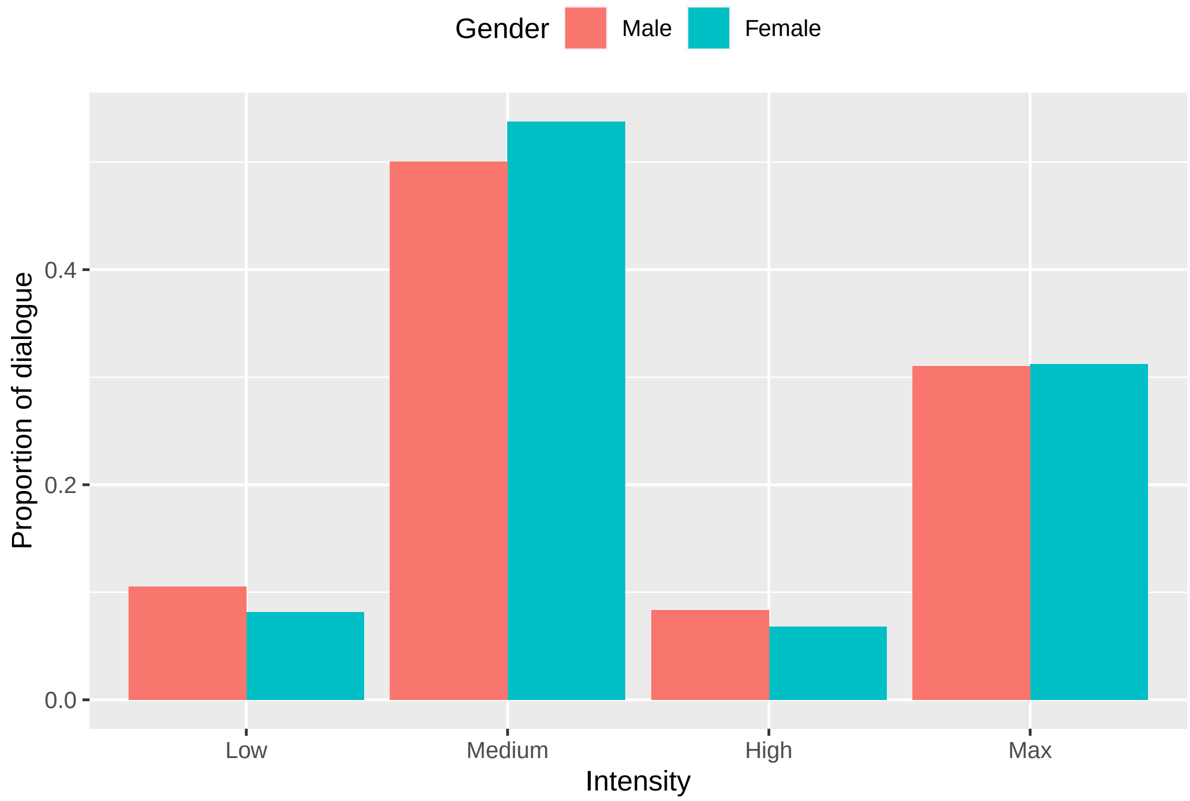
These scores were split into categories in order to facilitate analysis. Scores from 0 to 49 were classed as 'low', from 49 to 51 as 'medium', from 51-99 as 'high' and 100 as 'max'. The tables below show how lines of dialogue are distributed across intensity categories and genders.

```
tabI = table(emotions$intensityCategory,emotions$gender)
tabIprop = prop.table(tabI,2)
colnames(tabIprop) = paste(colnames(tabIprop),"(%)")
kable(tabI,"latex")
```

	Female	Male
Low	791	1701
Medium	5200	8076
High	658	1348
Max	3018	5007

```
kable(round(100*tabIprop,1),"latex")
```

	Female (%)	Male (%)
Low	8.2	10.5
Medium	53.8	50.1
High	6.8	8.4
Max	31.2	31.0




```
chiSqIntensity = chisq.test(tabI)
chiSqIntensity
```

```
##
##  Pearson's Chi-squared test
##
## data:  tabI
## X-squared = 69.971, df = 3, p-value = 4.331e-15
```

The distribution of intensity categories by gender is significantly different from what would be expected in a random distribution ($\chi^2(6) = 118.5$, p = 0.0001). Female NPCs are more likely to be given "Medium" intensity lines than male NPCs, and less likely to be given Low or High intensities. The proportion of "Max" intensities is roughly equal.

Treating the intensity scores as continuous, we find that the average intensity is about half a point lower for male NPCs than female NPCs. Below we test whether this is statistically significant using a permutation test (a t-test is not appropriate, since the distribution is not normal).

```
trueDiff = diff(tapply(emotions$intensity,emotions$gender,mean))
trueDiff
```

```
##       Male
## -0.4994472
```

```
permuteI = function(){
  diff(tapply(emotions$intensity,sample(emotions$gender),mean))
}
n = 10000
permutedDiff = replicate(n,permuteI())
p.value = sum(permutedDiff < trueDiff) / n
if(p.value==0){
  p.value = 1/n
}
z.score = (trueDiff- mean(permutedDiff)) / sd(permutedDiff)
```

```r
paste("Permutation z =",round(z.score,2), ", p = ",p.value)
```

```
## [1] "Permutation z = -1.43 , p =  0.0758"
```

The result is not significant and the effect size for the continuous data is very small. The patterns may be better captured by the categorical analysis above.

Exploratory analyses

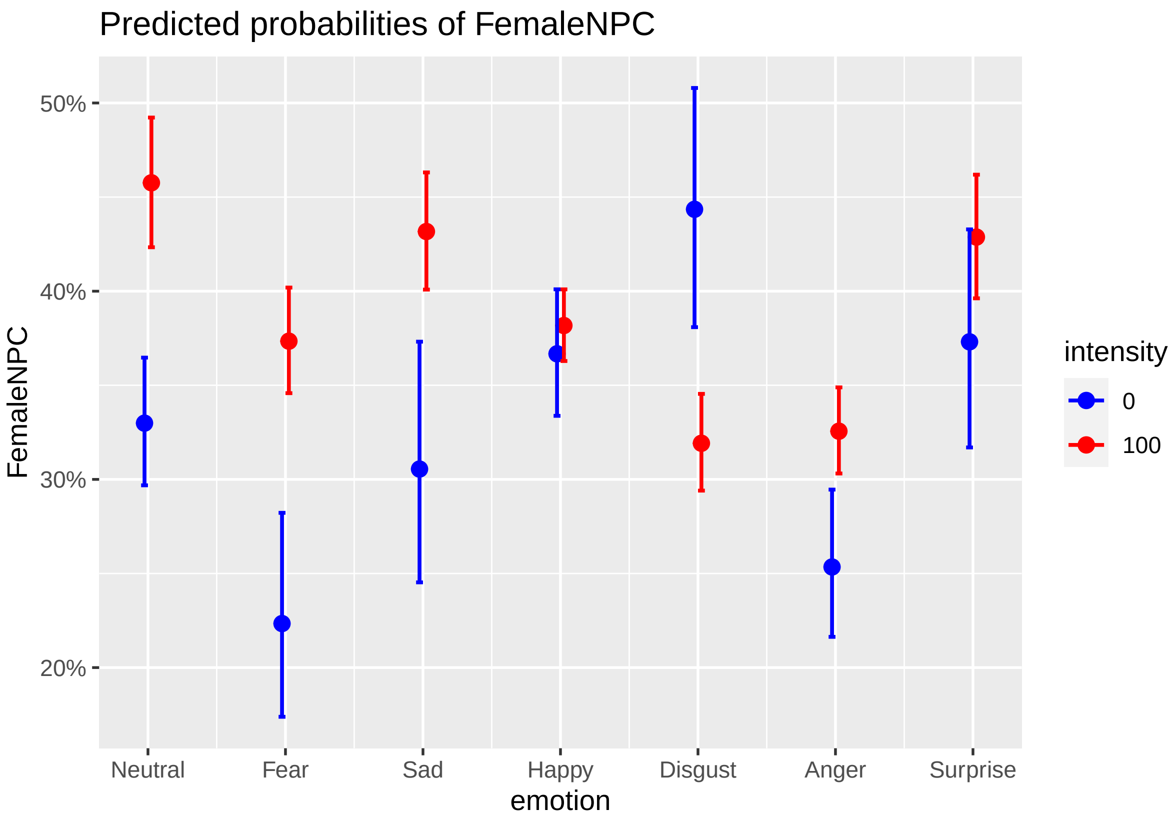
Interactions between emotion and intensity

We explored the interaction between emotion type and intensity. For simplicity, we use the continuous measure of intensity in a binomial regression predicting the gender of the NPC:

```
table(emotions$emotion,emotions$intensityCategory,emotions$gender)
```

```
## , ,  = Female
##
##
##             Low Medium High  Max
##   Anger     135    245  121  477
##   Disgust    96    145   68  365
##   Fear       45    104   66  387
##   Happy     296    442  250  791
##   Neutral    47   4006   29  280
##   Sad        65    105   69  383
##   Surprise  107    153   55  335
##
## , ,  = Male
##
##
##             Low Medium High  Max
##   Anger     298    566  480  890
##   Disgust   123    265  113  775
##   Fear      158    222  123  663
##   Happy     585    627  366 1358
##   Neutral   263   5941   74  396
##   Sad       117    198  123  483
##   Surprise  157    257   69  442
```

```
emotions$FemaleNPC = factor(emotions$gender=="Female")
mCombo = glm(FemaleNPC ~ emotion* intensity,
            data = emotions,family="binomial")
plot_model(mCombo,"int",colors = c("blue","red"))
```



For anger, fear and sadness, increasing intensity is associated with a higher likelihood of being from a female NPC. For disgust, the opposite was true: higher intensity was associated with a higher likelihood of being from a male NPC.

Emotions per character

We looked at the range of emotions per character. It's possible that, while females may be more likely to exhibit neutral emotions, an individual female may be more likely to exhibit a greater range of emotions than an individual man. Or put it another way, a female may exhibit a range of emotions while a man may specialise in one emotion.

```
LinesPerChar = as.vector(table(emotions$name))
emotionEntropy = tapply(factor(emotions$emotion), emotions$name, function(X){
  entropy(table(X))
})
emotionNum = tapply(factor(emotions$emotion),emotions$name,function(X){length(unique(X))})

ent = data.frame(name = names(emotionEntropy),
                 entropy= emotionEntropy,
                 emotionNum = emotionNum,
                 lines=LinesPerChar)
ent$lines.log = log(ent$lines)
ent$gender = emotions[match(ent$name,emotions$name),]$gender
ent$gender = factor(ent$gender,levels=c("Male","Female"))
contrasts(ent$gender) = contr.sum(2)/2
```

The table below shows the proportion of characters displaying only 1 emotion, 2 emotions, 3 emotions etc.

```
table(ent$emotionNum,ent$gender)
```

```
##
##     Male Female
##   1  147     61
##   2  143     71
##   3   93     63
##   4   90     49
##   5   72     54
##   6   83     50
##   7   52     38
```

```
prop.table(table(ent$emotionNum,ent$gender),2)
```

```
##
##           Male     Female
##   1 0.21617647 0.15803109
##   2 0.21029412 0.18393782
##   3 0.13676471 0.16321244
##   4 0.13235294 0.12694301
##   5 0.10588235 0.13989637
##   6 0.12205882 0.12953368
##   7 0.07647059 0.09844560
```

```
barplot(t(prop.table(table(ent$emotionNum,ent$gender),2)),
        beside = T,
        xlab="Number of emotions",
        ylab = "Proportion of characters (within gender)",
        legend.text = c("Male","Female"),col=c("blue","pink"))
```

It looks like the distribution for males is more biased towards low ranges of emotions, while the distribution for females is slighly higher.

However, this effect might be confounded by the amount of dialouge. There is no significant effect of gender on the number of emotions displayed when the (log) number of lines is considered:

```
mx = glm(emotionNum ~ lines.log+gender, data = ent, family="poisson")
summary(mx)
```

```
##
## Call:
## glm(formula = emotionNum ~ lines.log + gender, family = "poisson",
##     data = ent)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.13764  -0.46977  -0.05695   0.38011   1.60908
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.38192    0.03842   9.942   <2e-16 ***
## lines.log    0.33809    0.01183  28.579   <2e-16 ***
## gender1     -0.04042    0.03376  -1.197    0.231
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1212.99  on 1065  degrees of freedom
## Residual deviance:  394.72  on 1063  degrees of freedom
## AIC: 3564.4
##
```

```
## Number of Fisher Scoring iterations: 4
```

Similarly, looking at the entropy of emotions displayed per-character shows that females have higher entropy than males:

```
t.test(ent$entropy~ent$gender)
```

```
##
##  Welch Two Sample t-test
##
## data:  ent$entropy by ent$gender
## t = -2.7609, df = 828.08, p-value = 0.005892
## alternative hypothesis: true difference in means between group Male and group Female is not equal
## 95 percent confidence interval:
##  -0.16129027 -0.02724875
## sample estimates:
##   mean in group Male mean in group Female
##            0.8223120            0.9165815
```

However, this is also confounded by the number of lines of dialogue for each character. Below we show that there is no significant effect of gender when taking into account the (log) number of lines each character speaks.

```
# The entropy measure has a normal distribution,
# but also lots of characters who only display one emotion.
#  (entropy of zero)
# So we use a zero-inflated beta distribution.

# First, scale entropy to be between zero and 1
ent$entropy2 = ent$entropy / (max(ent$entropy)+0.00001)
bmx = brm(entropy2 ~ lines.log + gender,data=ent, family = "zero_inflated_beta")
```

```
## Running /Library/Frameworks/R.framework/Resources/bin/R CMD SHLIB foo.c
## clang -arch arm64 -I"/Library/Frameworks/R.framework/Resources/include" -DNDEBUG   -I"/Library/Fr
## In file included from <built-in>:1:
## In file included from /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/StanHe
## In file included from /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/RcppEi
## In file included from /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/RcppEi
## /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/RcppEigen/include/Eigen/src/
## namespace Eigen {
## ^
## /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/RcppEigen/include/Eigen/src/
## namespace Eigen {
##                ^
##                ;
## In file included from <built-in>:1:
## In file included from /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/StanHe
## In file included from /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/RcppEi
## /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library/RcppEigen/include/Eigen/Core
## #include <complex>
##          ^~~~~~~~~
## 3 errors generated.
## make: *** [foo.o] Error 1
##
## SAMPLING FOR MODEL 'dea086090905fb2839955020334863e6' NOW (CHAIN 1).
## Chain 1:
## Chain 1: Gradient evaluation took 0.000592 seconds
## Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 5.92 seconds.
## Chain 1: Adjust your expectations accordingly!
## Chain 1:
## Chain 1:
```

```
## Chain 1: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 1: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 1: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 1: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 1: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 1: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 1: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 1: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 1: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 1: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 1: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 1: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 1:
## Chain 1:  Elapsed Time: 2.52281 seconds (Warm-up)
## Chain 1:                2.54467 seconds (Sampling)
## Chain 1:                5.06748 seconds (Total)
## Chain 1:
##
## SAMPLING FOR MODEL 'dea086090905fb2839955020334863e6' NOW (CHAIN 2).
## Chain 2:
## Chain 2: Gradient evaluation took 0.000393 seconds
## Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 3.93 seconds.
## Chain 2: Adjust your expectations accordingly!
## Chain 2:
## Chain 2:
## Chain 2: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 2: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 2: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 2: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 2: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 2: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 2: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 2: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 2: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 2: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 2: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 2: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 2:
## Chain 2:  Elapsed Time: 2.50907 seconds (Warm-up)
## Chain 2:                2.48588 seconds (Sampling)
## Chain 2:                4.99494 seconds (Total)
## Chain 2:
##
## SAMPLING FOR MODEL 'dea086090905fb2839955020334863e6' NOW (CHAIN 3).
## Chain 3:
## Chain 3: Gradient evaluation took 0.000396 seconds
## Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 3.96 seconds.
## Chain 3: Adjust your expectations accordingly!
## Chain 3:
## Chain 3:
## Chain 3: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 3: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 3: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 3: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 3: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 3: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 3: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 3: Iteration: 1200 / 2000 [ 60%]  (Sampling)
```

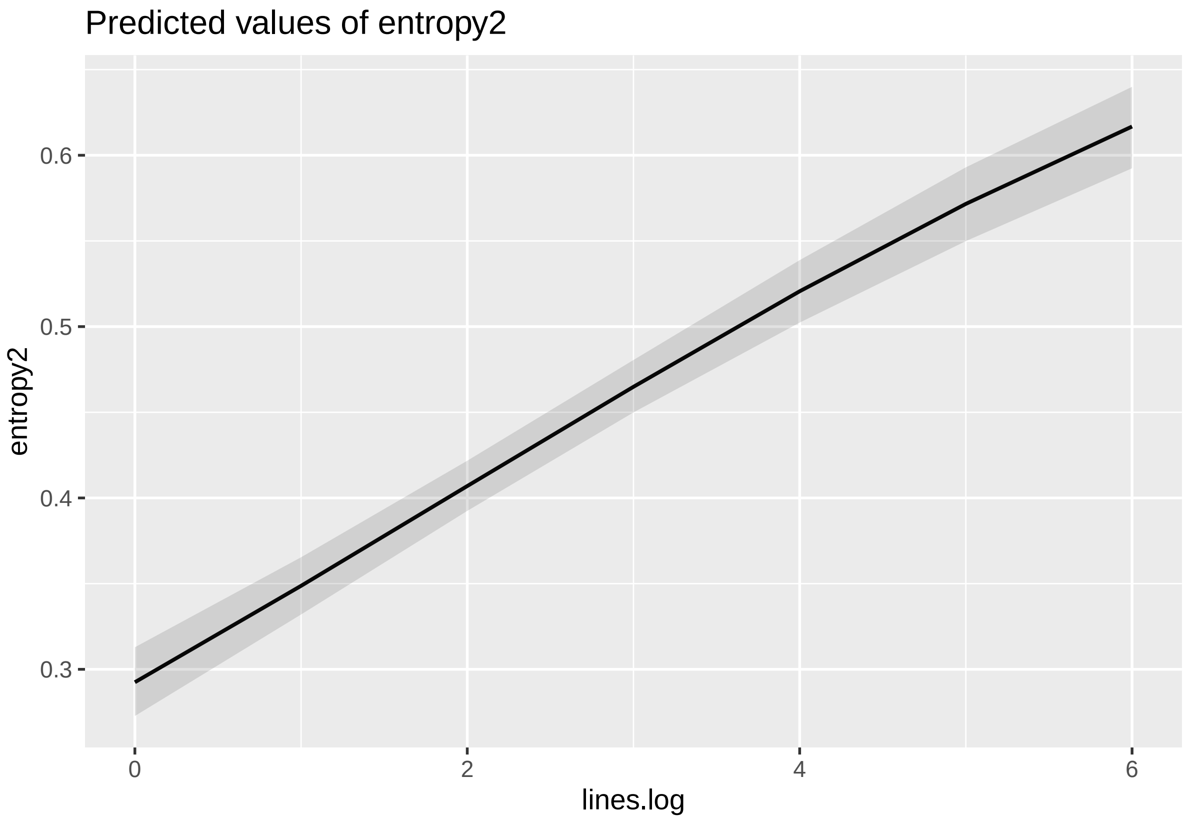
```
## Chain 3: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 3: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 3: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 3: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 3:
## Chain 3:  Elapsed Time: 2.59876 seconds (Warm-up)
## Chain 3:                2.31574 seconds (Sampling)
## Chain 3:                4.91451 seconds (Total)
## Chain 3:
##
## SAMPLING FOR MODEL 'dea086090905fb2839955020334863e6' NOW (CHAIN 4).
## Chain 4:
## Chain 4: Gradient evaluation took 0.000393 seconds
## Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 3.93 seconds.
## Chain 4: Adjust your expectations accordingly!
## Chain 4:
## Chain 4:
## Chain 4: Iteration:    1 / 2000 [  0%]  (Warmup)
## Chain 4: Iteration:  200 / 2000 [ 10%]  (Warmup)
## Chain 4: Iteration:  400 / 2000 [ 20%]  (Warmup)
## Chain 4: Iteration:  600 / 2000 [ 30%]  (Warmup)
## Chain 4: Iteration:  800 / 2000 [ 40%]  (Warmup)
## Chain 4: Iteration: 1000 / 2000 [ 50%]  (Warmup)
## Chain 4: Iteration: 1001 / 2000 [ 50%]  (Sampling)
## Chain 4: Iteration: 1200 / 2000 [ 60%]  (Sampling)
## Chain 4: Iteration: 1400 / 2000 [ 70%]  (Sampling)
## Chain 4: Iteration: 1600 / 2000 [ 80%]  (Sampling)
## Chain 4: Iteration: 1800 / 2000 [ 90%]  (Sampling)
## Chain 4: Iteration: 2000 / 2000 [100%]  (Sampling)
## Chain 4:
## Chain 4:  Elapsed Time: 2.52473 seconds (Warm-up)
## Chain 4:                2.72139 seconds (Sampling)
## Chain 4:                5.24611 seconds (Total)
## Chain 4:
```

```
summary(bmx)
```

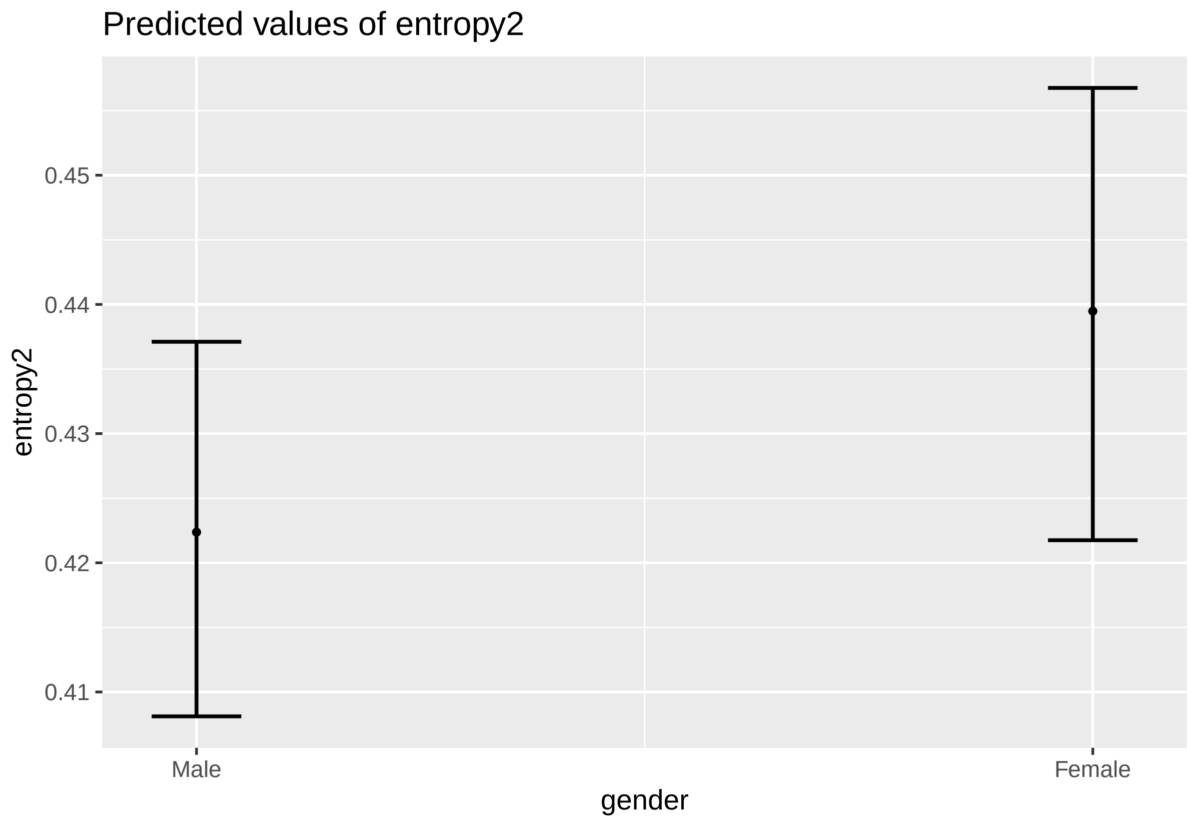
```
##  Family: zero_inflated_beta
##   Links: mu = logit; phi = identity; zi = identity
## Formula: entropy2 ~ lines.log + gender
##    Data: ent (Number of observations: 1066)
##   Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
##          total post-warmup draws = 4000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept    -0.52      0.06    -0.64    -0.40 1.00     4228     3044
## lines.log     0.29      0.02     0.25     0.33 1.00     4819     3345
## gender1      -0.09      0.05    -0.19     0.02 1.00     4499     2883
##
## Family Specific Parameters:
##     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## phi     6.21      0.29     5.67     6.78 1.00     4891     3269
## zi      0.20      0.01     0.17     0.22 1.00     4231     2347
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```

```r
plot_model(bmx,"pred")
```

```
## $lines.log
```




```
##
## $gender
```



Predicted values of entropy2
entropy2
0.45
0.44
0.43
0.42
0.41
Male
Female
gender

Cues in angry dialogue

There was a difference between genders in the emotion and intensity. One of the biggest differences was for lines marked with the "anger" emotion. However, is there a difference between genders in the linguistic expression of these emotions? That is, do male and female characters express anger differently?

Below, we extract frequencies of profanities in male and female dialogue for lines tagged with the "anger" emotion.

```
dialogueAnger= dialogue[emotions$emotion=="Anger"]
anger = corpus(dialogueAnger,
               docvars = data.frame(
                 gender = emotions[emotions$emotion=="Anger",]$gender))
toks_nopunct <- tokens(anger, remove_punct = TRUE)
totalWords = table(toks_nopunct$gender)

dfmat <- dfm(toks_nopunct)
tstat_freq <- textstat_frequency(dfmat, n = 40)

swears = read.csv("https://gist.githubusercontent.com/tjrobinson/2366772/raw/97329ead3d5ab06160c3c7a
                  stringsAsFactors = F,header = F)
swears = swears[,1]
swears = swears[!swears %in% c("snatch")]
nx = c("hell","dago","ass")
swears[!swears %in% nx]= paste0(swears[!swears %in% nx],"*")
swears = c(swears,"vermin","scum")
swears = dictionary(list(swears=swears))
toks_swears = tokens_keep(toks_nopunct,swears)
dfmat_swears <- dfm(toks_swears)
tstat_freq_swears <- textstat_frequency(dfmat_swears, groups = gender)

tots = sapply(unique(tstat_freq_swears$feature),function(X){
  z = tstat_freq_swears[tstat_freq_swears$feature==X,]
  f = z[z$group=="Female",]$frequency
  m = z[z$group=="Male",]$frequency
  c(Female = ifelse(length(f)==0,0,f),
    Male = ifelse(length(m)==0,1,m))
  })
tots = t(tots)
swearSums = colSums(tots)
tots = rbind(tots,TOTAL = swearSums)
tots = rbind(tots,TOTAL.PerThousand = 1000 * (swearSums/totalWords))
tots
```

```
##                     Female    Male
## damned             7.00000 12.0000
## damn               5.00000 28.0000
## bloody             3.00000  1.0000
## bastard            3.00000  6.0000
## hell               2.00000  6.0000
## bastards           1.00000 15.0000
## arse               1.00000  1.0000
## scum               0.00000 10.0000
## ass                0.00000  5.0000
## pisses             0.00000  4.0000
## vermin             0.00000  3.0000
## damnable           0.00000  1.0000
## asses              0.00000  1.0000
## crap               0.00000  1.0000
## goddam             0.00000  1.0000
```

```
## cocky               0.00000  1.0000
## goddamn             0.00000  1.0000
## mongrel             0.00000  1.0000
## TOTAL              22.00000 98.0000
## TOTAL.PerThousand 22.49489 43.8675
```

```
chisqSwear = chisq.test(rbind(swearSums,totalWords - swearSums))
chiSqString = paste0("female frequency per thousand words = ",
                     round(tots[nrow(tots),1],1),
                     ", male frequency per thousand words = ",
                     round(tots[nrow(tots),2],1),", ",
        "$\\chi^2$(",chisqSwear$parameter,
        ") = ",round(chisqSwear$statistic,2),
        ", p = ",signif(chisqSwear$p.value,2))
cat(chiSqString,file="../../results/latexStats/OblivionSwearingChiSq.tex")
```

Male characters use profanities twice as often as female characters in angry dialogue (female frequency per thousand words = 22.5, male frequency per thousand words = 43.9, $\chi^2(1) = 8.06$, p = 0.0045).

Dialogue by generic NPCs

Oblivion also includes spoken dialogue by "generic" NPCs, for example guards. These lines are usually duplicated for each gender. For example, the phrase "It's time for you to go" is recorded for both generic Imperial male NPC and a generic Imperial female NPC (also for all other races in the game).

However, not all lines are duplicated. There are some unique lines of dialogue for generic characters.

```
# Find unique dialogue
tab = table(dialogue.generic)
uniqueGenericDialogue = which(dialogue.generic %in% names(tab[tab==1]))
```

There are 942 unique lines of dialogue for generic characters. We can identify how these are distributed by gender:

```
uniqueGenericDialogueByGender = rbind(
  Unique = table(emotions.generic[uniqueGenericDialogue,]$gender),
  Duplicated = table(emotions.generic[-uniqueGenericDialogue,]$gender))
kable(uniqueGenericDialogueByGender,"latex")
```

	Female	Male
Unique	140	802
Duplicated	16573	19174

```
ugpt= prop.table(uniqueGenericDialogueByGender,2)
colnames(ugpt) = paste(colnames(ugpt),"(%)")
kable(round(ugpt*100,1),"latex")
```

	Female (%)	Male (%)
Unique	0.8	4
Duplicated	99.2	96

```
chiSqUniqueGenericDialogue = chisq.test(uniqueGenericDialogueByGender)
chiSqUniqueGenericDialogue
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  uniqueGenericDialogueByGender
## X-squared = 365.91, df = 1, p-value < 2.2e-16
```

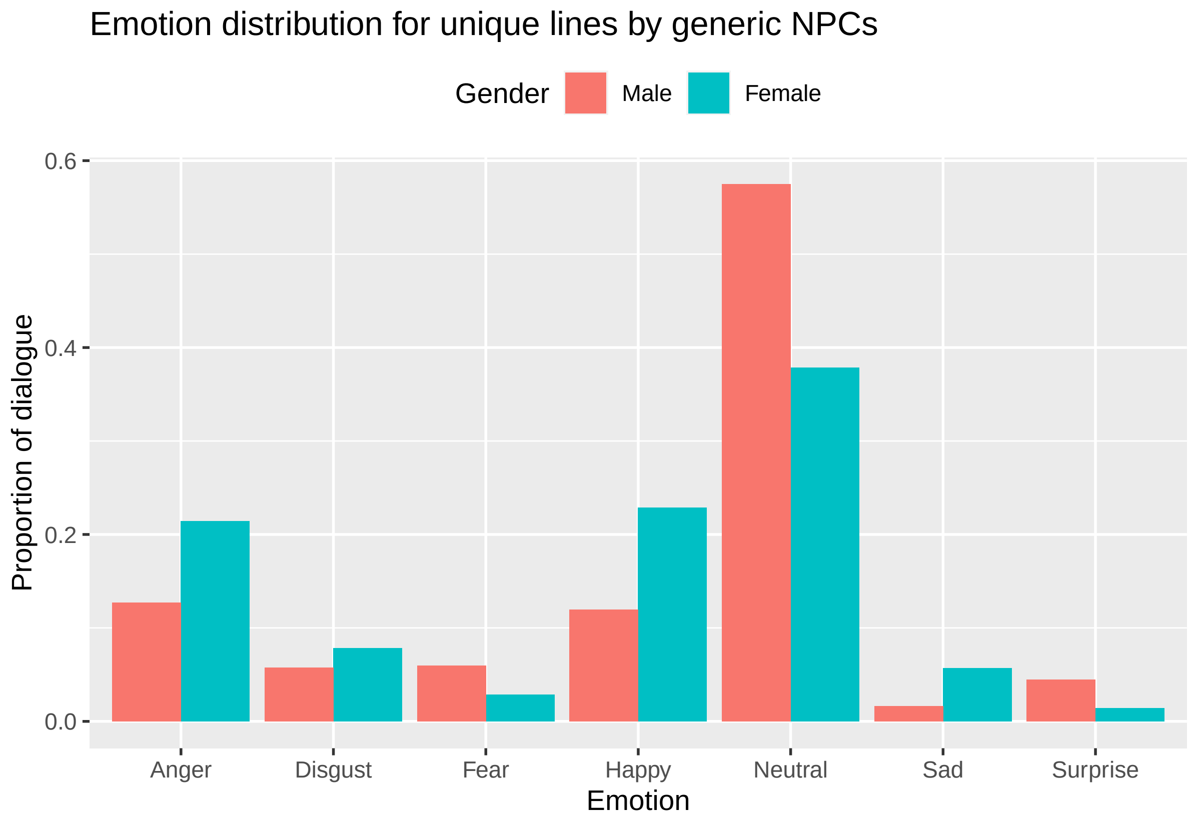
There is a significant imbalance: generic male NPCs are more than four times more likely to be given unique lines of dialogue than generic female NPCs.

```
p.val = round(chiSqUniqueGenericDialogue$p.value,3)
if(p.val < 0.0001){ p.val = "0.0001"}else{p.val = paste("=",p.val)}
chiSqString = paste0("$\\chi^2$(",chiSqUniqueGenericDialogue$parameter,
                     ") = ",round(chiSqUniqueGenericDialogue$statistic,2),
                      ", p = ",p.val)
cat(chiSqString, file="../../results/latexStats/OblivionUniqueGenericDialogueChiSq.tex")
```

We can also look at how the emotions are distributed:

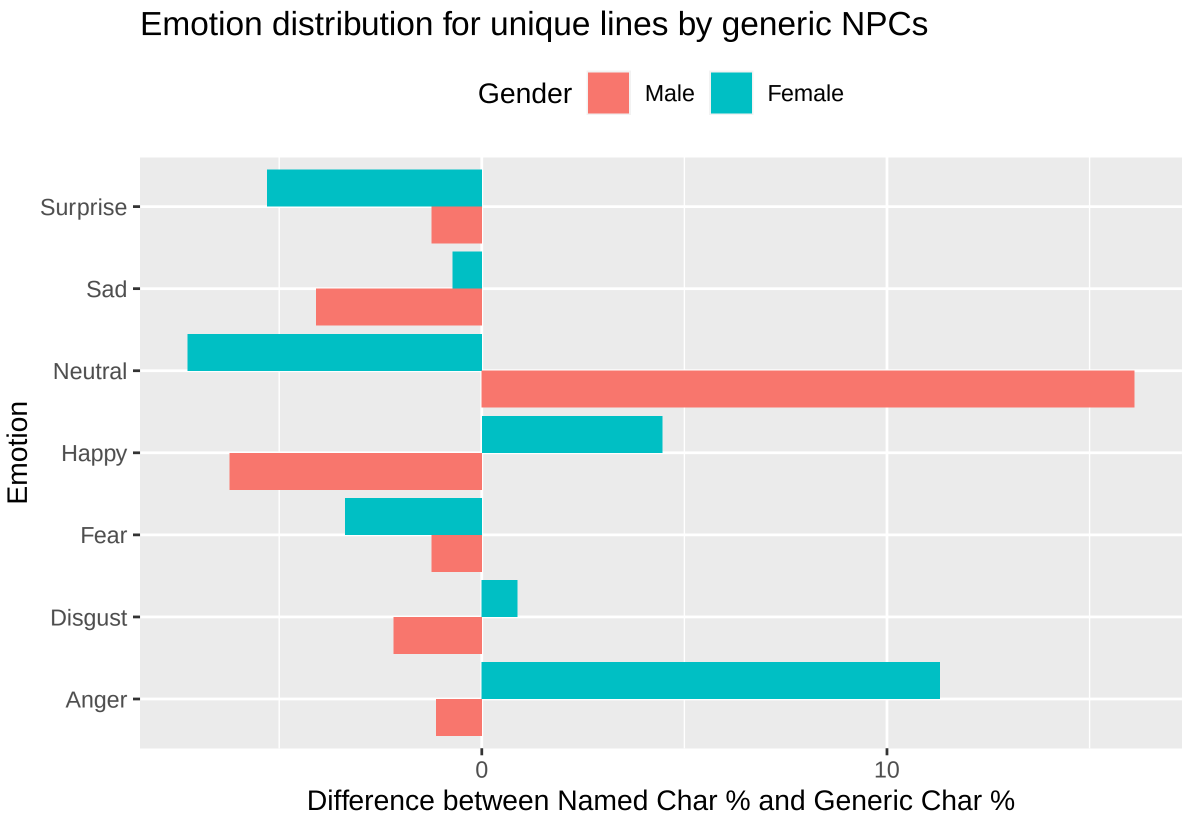
```
ugdByGenderAndEmotion = table(
  emotions.generic[uniqueGenericDialogue,]$emotion,
  emotions.generic[uniqueGenericDialogue,]$gender)

ugdByGenderAndEmotionProp = prop.table(ugdByGenderAndEmotion,2)

dxG = data.frame(
  n = as.vector(ugdByGenderAndEmotionProp),
  Gender = rep(c("Female","Male"),each=nrow(ugdByGenderAndEmotionProp)),
  Emotion = rep(rownames(ugdByGenderAndEmotionProp),2)
)
```

```r
dxG$Gender = factor(dxG$Gender,levels = c("Male","Female"))
ggplot(dxG, aes(fill=Gender,y=n, x=Emotion)) +
    geom_bar(position="dodge", stat="identity") +
  ylab("Proportion of dialogue") +
  ggtitle("Emotion distribution for unique lines by generic NPCs") +
  theme(legend.position = "top")
```



The biggest differences between the named and generic characters are for females: they are less likely to have neutral emotions, are more likely to have happy and angry emotions. The differences between the two types of character are visualised below:

```r
dx$diff = 100*(dxG$n - dx$n)
ggplot(dx,aes(y=diff,fill=Gender,x=Emotion)) +
  geom_bar(position="dodge", stat="identity") +
  theme(legend.position = "top") +
  coord_flip() +
  ggtitle("Emotion distribution for unique lines by generic NPCs") +
  ylab("Difference between Named Char % and Generic Char %")
```


Emotion distribution for unique lines by generic NPCs
Gender
Male
Female
Surprise
Sad
Neutral
Happy
Fear
Disgust
Anger
Emotion
0
10
Difference between Named Char % and Generic Char %

Conclusion

Compared to Male NPCs, Female NPCs are statistically more likely to be given ‘neutral’ emotion cues with ‘medium’ intensity. This supports the hypothesis that female characters are “backgrounded”. This may be because they have a narrower range of roles or are less relevant at the most ‘emotional’ or dramatic moments in the game. However, this hypothesis did not predict the finding that female NPCs would express more sadness and surprise.

We also found support for the gendered emotions hypothesis: female NPCs were more likely to express feminine emotions and male NPCs were more likely to express masculine emotions. However, this is mainly a test of anger versus the other (non-neutral) emotions. This hypothesis also does not predict the difference in neutral emotion lines.

We found no support for the hypothesis based on the belief that women express more emotion than men. However, there are some hints that emotion is expressed differently. For example, male NPCs were more likely to use curse words when expressing anger than female NPCs. Although it is often believed that, in the real world, men use curse words more frequently than women, some studies show no difference between genders (e.g. McEnery 2006, p. 29).

The patterns in the generic character dialogue require more analysis. Tellingly, there is less unique dialogue for female generic NPCs, fitting with the “backgrounding” hypothesis.

References

Bethesda Softworks (2006) The Elder Scrolls IV: Oblivion [Video Game]. Cheng.

Brody, L. R.; Hall, J. A. (2008). “Gender and emotion in context”. Handbook of Emotions. 3: 395–408.

Carillo Masso, I. 2011. The grips of fantasy: The construction of female characters in and beyond virtual game worlds. In: Ensslin, A.M., E ed. Creating Second Lives: Community, Identity and Spatiality as Constructions of the Virtual. New York: Routledge, pp. 113-142.

McEnery, T. 2006. Swearing in English bad language, purity and power from 1586 to the present. London, New York: Routledge.

Fairclough, N. 2003. Analysing discourse : textual analysis for social research. New York: Routledge.

Fairclough, N. et al. 2011. Critical Discourse Analysis. In: Van Dijk, T.A. ed. Discourse Studies: A Multidisciplinary Introduction. 2 ed. London: SAGE Publications Ltd.

Fischer, A. H., & Evers, C. (2011). The social costs and benefits of anger as a function of gender and relationship context. Sex roles, 65(1), 23-34.

Jansz, J (2000). “Masculine identity and restrictive emotionality”. Gender and Emotion. Gender and Emotion: Social Psychological Perspectives. pp. 166–186.

Kring, A. M.; Gordon, A. H. (1998). “Sex differences in emotion: expression, experience, and physiology”. Journal of Personality and Social Psychology. 74 (3): 686–703.

Miller, M. K. and Summers, A. 2007. Gender Differences in Video Game Characters’ Roles, Appearances, and Attire as Portrayed in Video Game Magazines. Sex roles 57(9), pp. 733-742.

Timmers, M., Fischer, A. H., & Manstead, A. S. R. (2003). Ability versus vulnerability: Beliefs about men’s and women’s emotional behavior. Cognition and Emotion, 17, 41–63.

8 Player choices

Player Choices

Introduction

Load libraries

```
library(rjson)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
source("https://raw.githubusercontent.com/datavizpyr/data/master/half_flat_violinplot.R")
```

Dialogue choices

We simulated an omnicient player who tries to maximise dialogue from one gender while minimising the dialogue from another (without repeat encounters).

Example: FFVII

The figure below shows a dialogue tree from Final Fantasy VII. Aerith (blue) initiates a conversation with the player character (red). The total number of words spoken by each character is listed at the bottom of each branch of the tree. Depending on the player's decisions, they might experience 70% female dialogue (left), or 60% (middle) or 58% (right).

There are various ways of measuring the variation created by possible choices. One way is to simulate a player that makes random decisions: for each dialogue tree, make a random choice of responses and record the proportions you observe. Alternatively, an omniscient player who had full knowledge of the dialogue tree could make choices to either maximise or minimise the proportion of female dialogue.

There are therefore several possible measures for the proportion of female dialogue:

- The proportion of female dialogue written by the game authors (the main measure in the rest of the paper)
- The range of proportions from a player making random decisions.
- The proportion of female dialogue experienced if an omniscient player tried to maximise female dialogue.
- The proportion of female dialogue experienced if an omniscient player tried to maximise male dialogue.

These measures were estimated using the script `processing/getChoiceVariation.py`. Proportions from a random player were estimated by creating 100 scripts from random decisions for each game without re-visiting

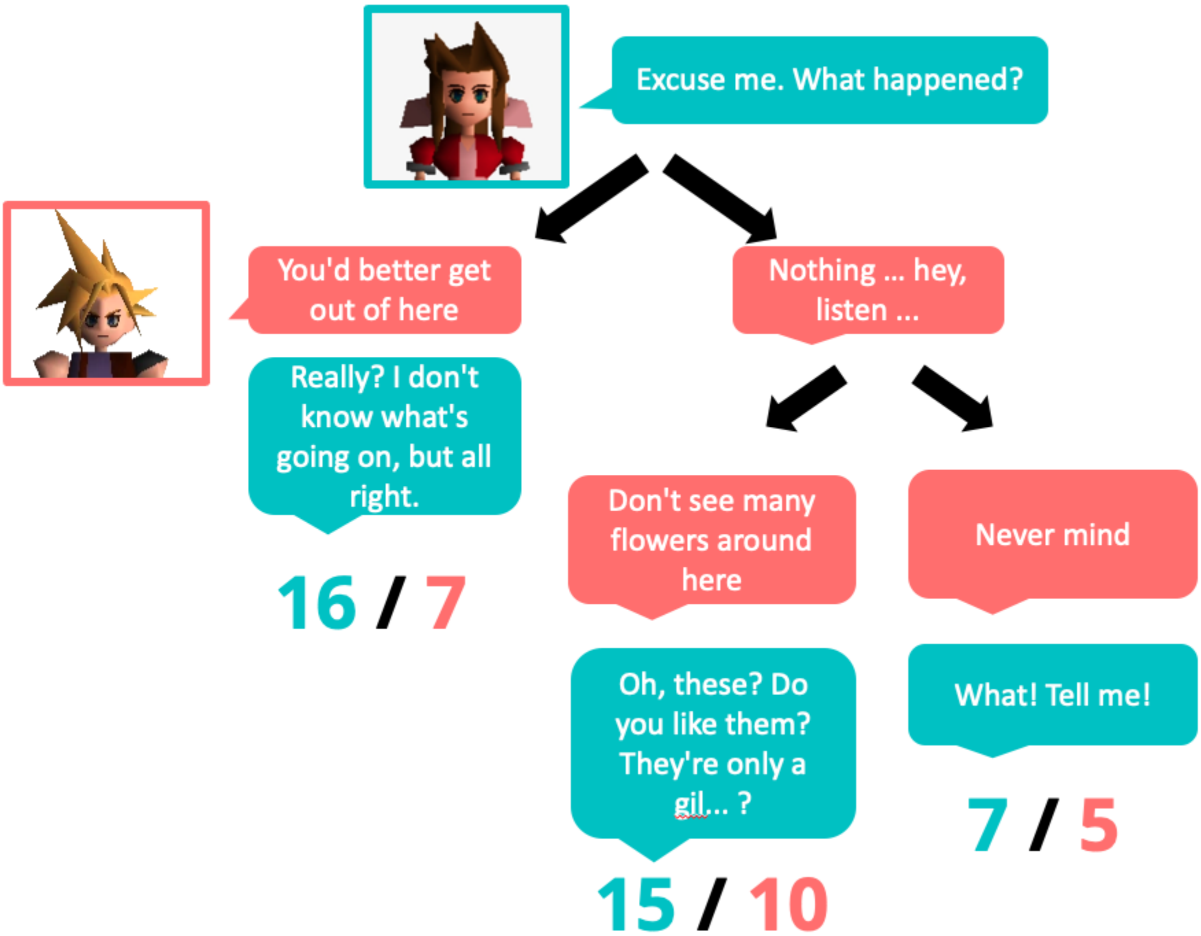


Figure 1: Example dialogue tree

any lines. For each of these random scripts, the proportion of female dialogue is calculated. Then we measure the mean of these proportions as well as the range of values (using a summary statistic like the 95% quantile is possible, but in practice the range of random options is very small).

Proportions from an omniscient player were obtained by exhaustive search through each independent dialogue tree. This assumes that choice blocks in the main script level are independent, which is not always true (e.g. in Mass Effect, certain dialogue decisions at one point in the game affect which options are available later). However, obtaining this high-level data is difficult.

For the omnisicent player results, let's walk through an example from Final Fantasy VII. First, we load the data generated by the python script:

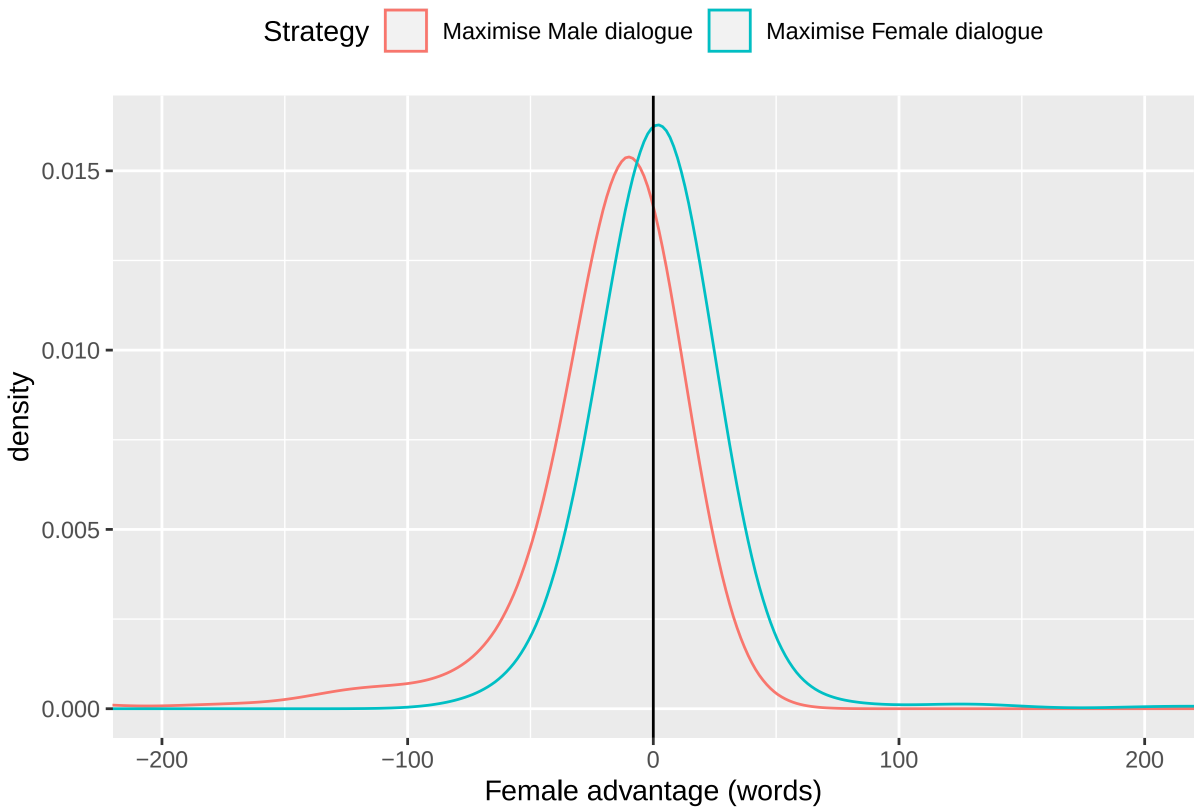
```
folder = "../data/FinalFantasy/FFVII/"
d = read.csv(paste0(folder,"/choiceVariation.csv"),stringsAsFactors = F)
d = d[!is.na(d$maxM.femaleWords),]
```

We can now work out the "female advantage": the difference between the number of female words and the number of male words (higher than zero = more female words than male words).

```
d$maxM.femaleAdvantage = d$maxM.femaleWords - d$maxM.maleWords
d$maxF.femaleAdvantage = d$maxF.femaleWords - d$maxF.maleWords
```

Collate data: for each dialogue tree, plot the distribution of female advantage for two strategies: maximising female dialogue and maximising male dialogue:

```
dx = data.frame(
  FemaleAdvantage = c(d$maxF.femaleAdvantage,d$maxM.femaleAdvantage),
  Strategy = rep(c("Maximise Female dialogue","Maximise Male dialogue"),
                 times = c(nrow(d),nrow(d)))
  )
dx$Strategy = relevel(factor(dx$Strategy),"Maximise Male dialogue")

ggplot(dx, aes(x=FemaleAdvantage,color=Strategy)) +
  geom_density(bw=20) +
  theme(legend.position = "top") +
  xlab("Female advantage (words)") +
  geom_vline(xintercept = 0) +
  coord_cartesian(xlim = c(-200,200))
```



Note that the female distribution is centered around zero, while the male distribution is biased to the left: there are more opportunities to maximise male dialogue.

Work out some stats:

```
# Test if distribution is centered around zero.
maxFTest = t.test(d$maxF.femaleAdvantage, mu = 0)
# prop trees with more female dialogue than male
maxF.moreF = prop.table(table(d$maxF.femaleWords > d$maxF.maleWords))[2]
# mean female advantage
maxF.meanFAdv = mean(d$maxF.femaleAdvantage)

# Same for max M strategy
maxMTest = t.test(d$maxM.femaleAdvantage, mu = 0)
maxM.moreM = prop.table(table(d$maxM.maleWords > d$maxM.femaleWords))[2]
maxM.meanMAdv = -mean(d$maxM.femaleAdvantage)
```

In Final Fantasy VII, a player trying to maximise female dialogue over male would succeed in observing more female dialogue than male dialogue in 55% of dialogue trees, on average seeing 2.5 more words spoken by females than males in each dialogue tree (not significantly different from zero, $p = 0.03$).

When maximising male dialogue over female, a player would succeed in observing more male dialogue than female in 86.1%, on average seeing 24.4‘ more words from males than females (highly significant, $p < 0.0001$).

Analysis of dialogue choices for all games

We now apply the method above to all games. First, we load the data:

```
folders = list.dirs("../data", recursive = T)
folders = folders[sapply(folders,function(X){
  "choiceVariation.csv" %in% list.files(X)
})]

allGames= NULL
allData = NULL
transitionStrings = NULL
randChoices = NULL
for(folder in folders){
  print(folder)
  js = fromJSON(file = paste0(folder,"/meta.json"))
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  if(!alternativeMeasure){
    cvFile = paste0(folder,"/choiceVariation.csv")
    randFile = paste0(folder,"/stats_randomChoices.csv")
    statsFile = paste0(folder,"/stats.csv")
    if(file.exists(cvFile) & file.exists(statsFile) & file.exists(randFile)){

      stats = read.csv(statsFile,stringsAsFactors = F)
      d = read.csv(cvFile,stringsAsFactors = F)

      totalNonChoiceFemale = d[!is.na(d$totalNonChoice.femaleWords),]$totalNonChoice.femaleWords[1]
      totalNonChoiceMale = d[!is.na(d$totalNonChoice.maleWords),]$totalNonChoice.maleWords[1]
      d = d[is.na(d$totalNonChoice.maleWords),]
      allData = rbind(allData,d[,c("folder","maxF.maleWords",
                                   "maxF.femaleWords","maxM.maleWords",
                                   "maxM.femaleWords")])
      minFemaleProportion = (totalNonChoiceFemale+ sum(d$maxM.femaleWords)) /
          (totalNonChoiceFemale + totalNonChoiceMale +
             sum(d$maxM.femaleWords) + sum(d$maxM.maleWords))
      maxFemaleProportion = (totalNonChoiceFemale+ sum(d$maxF.femaleWords)) /
          (totalNonChoiceFemale + totalNonChoiceMale +
            sum(d$maxF.femaleWords) + sum(d$maxF.maleWords))

      mainFemaleProp = stats[stats$group=="female",]$words/
          (stats[stats$group=="female",]$words + stats[stats$group=="male",]$words)

      diffMainToFemaleMax = maxFemaleProportion - mainFemaleProp
      diffMainToFemaleMin = minFemaleProportion - mainFemaleProp

      allGames = rbind(allGames,data.frame(
        folder = folder,
        game = js$game,
        series = js$series,
        minFemaleProportion = minFemaleProportion,
        maxFemaleProportion = maxFemaleProportion,
        mainFemaleProp = mainFemaleProp,
```

```
        diffMainToFemaleMax = diffMainToFemaleMax,
        diffMainToFemaleMin = diffMainToFemaleMin,
        stringsAsFactors = F
      ))

      # Distributions from random choices
      rand = read.csv(randFile, stringsAsFactors = F)
      rand$folder = folder
      rand$maleWords = rand$maleWords + totalNonChoiceMale
      rand$femaleWords = rand$femaleWords + totalNonChoiceFemale
      rand$femaleProp = rand$femaleWords / (rand$maleWords + rand$femaleWords)
      randChoices = rbind(randChoices,rand)

    }
  }
}
```

```
## [1] "../data/ChronoTrigger/ChronoTrigger"
## [1] "../data/DragonAge/DragonAge2"
## [1] "../data/DragonAge/DragonAgeOrigins"
## [1] "../data/DragonAge/DragonAgeOrigins_B"
## [1] "../data/ElderScrolls/Daggerfall"
## [1] "../data/ElderScrolls/Morrowind"
## [1] "../data/ElderScrolls/Skyrim"
## [1] "../data/FinalFantasy/FFII"
## [1] "../data/FinalFantasy/FFIX"
## [1] "../data/FinalFantasy/FFIX_B"
## [1] "../data/FinalFantasy/FFVI_B"
## [1] "../data/FinalFantasy/FFVII"
## [1] "../data/FinalFantasy/FFVII_Remake"
## [1] "../data/FinalFantasy/FFVIII"
## [1] "../data/FinalFantasy/FFX"
## [1] "../data/FinalFantasy/FFX_B"
## [1] "../data/FinalFantasy/FFX2"
## [1] "../data/FinalFantasy/FFXII_B"
## [1] "../data/FinalFantasy/FFXIII-2"
## [1] "../data/FinalFantasy/FFXIV"
## [1] "../data/FinalFantasy/FFXV"
## [1] "../data/MassEffect/MassEffect1"
## [1] "../data/MassEffect/MassEffect1B"
## [1] "../data/MassEffect/MassEffect2"
## [1] "../data/MassEffect/MassEffect3B"
## [1] "../data/MassEffect/MassEffect3C"
## [1] "../data/MassEffect/MassEffectAndromeda"
## [1] "../data/MonkeyIsland/TheCurseOfMonkeyIsland"
## [1] "../data/Persona/Persona5B"
## [1] "../data/StardewValley/StardewValley"
## [1] "../data/StarWarsKOTOR/StarWarsKOTOR"
## [1] "../data/SuperMarioRPG/SuperMarioRPG"
## [1] "../data/Test/Test"
```

```
allGames = allGames[order(allGames$minFemaleProportion),]
allGames$num = 1:nrow(allGames)
allGames$game = as.character(allGames$game)
```

```r
allGames$game[grepl("Mario",allGames$game)] = "SuperMario RPG"
allGames$game = factor(allGames$game,levels = allGames$game[order(allGames$maxFemaleProportion)])

randChoices$game = allGames[match(randChoices$folder,allGames$folder),]$game
```

We remove Persona 5, because it only records one path through a dialogue tree.

```r
allGames = allGames[allGames$game!="Persona 5",]
randChoices = randChoices[randChoices$game!="Persona 5",]
```

Collate data, including working out the mean proportion from the random player samples.

```r
meanRand = tapply(randChoices$femaleProp,randChoices$folder,mean)
quantileRand = tapply(randChoices$femaleProp,
                      randChoices$folder,
                      #quantile,probs=c(0.025,0.975))
                      range)
randStats = data.frame(
  folder = names(meanRand),
  mean = meanRand,
  low = unlist(lapply(quantileRand,head,n=1)),
  high = unlist(lapply(quantileRand,tail,n=1))
)
allGames$randomMean = randStats[match(allGames$folder,randStats$folder),]$mean
allGames$randomLow = randStats[match(allGames$folder,randStats$folder),]$low
allGames$randomHigh = randStats[match(allGames$folder,randStats$folder),]$high
```

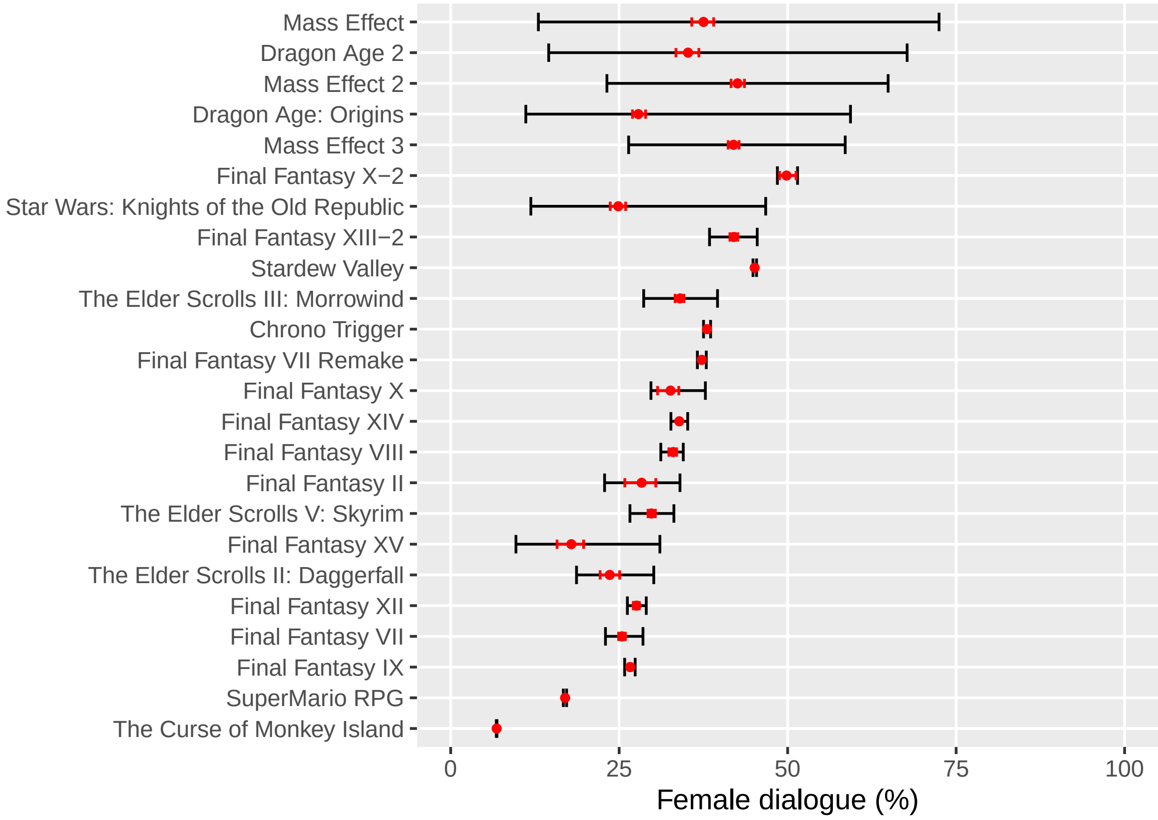
Stats for paper

```r
cat(nrow(allGames),file="../results/latexStats/choice_NumGames.tex")
```

Plot the variation in female dialogue proportions based on possible player choices. The graph shows:

- Red dot: The proportion of female dialogue written by the game authors (the main measure in the rest of the paper)
- Red whiskers: The range of proportions from a player making random decisions.
- Black whiskers: The theoretical range. That is, the range from the proportion of female dialogue experienced if an omniscient player tried to maximise male dialogue, to if they are trying to maximise female dialogue.

```r
choiceVarGraph = ggplot(data = allGames[!is.na(allGames$minFemaleProportion),],
       mapping = aes(y = minFemaleProportion*100,x=game)) +
  geom_errorbar(mapping = aes(ymin = minFemaleProportion*100,
                              ymax = maxFemaleProportion*100),
                width=0.6) +
  geom_errorbar(mapping = aes(ymin = randomLow*100,
                              ymax = randomHigh*100),
                width=0.3, colour="red") +
  geom_point(mapping=aes(y=randomMean*100,x=game),
                shape=16, color="red") +
  theme(panel.grid.minor.x = element_blank()) +
  ylab("Female dialogue (%)")+
  xlab("") +
  coord_flip(ylim=c(0,100))+
  scale_y_continuous(breaks=c(0,25,50,75,100))
choiceVarGraph
```




```
pdf("../results/graphs/Choices_MinMax.pdf", height=3,width=6)
choiceVarGraph
dev.off()
```

```
## pdf
##   2
```

Stats

```
allData$maxM.femaleAdvantage = allData$maxM.femaleWords - allData$maxM.maleWords
allData$maxF.femaleAdvantage = allData$maxF.femaleWords - allData$maxF.maleWords

# Test if distribution is centered around zero.
AmaxFTest = t.test(allData$maxF.femaleAdvantage, mu = 0)
AmaxFTest.p = round(AmaxFTest$p.value,2)
if(AmaxFTest.p==0){
  AmaxFTest.p = "< 0.001"
} else{
  AmaxFTest.p = paste("=",AmaxFTest.p)
}
cat(AmaxFTest.p, file="../results/latexStats/choice_maxF_p.tex")

# prop trees with more female dialogue than male
AmaxF.moreF = prop.table(table(allData$maxF.femaleWords > allData$maxF.maleWords))[2]
cat(round(AmaxF.moreF*100,1),file="../results/latexStats/choice_maxF_propMoreF.tex")
# mean female advantage
AmaxF.meanFAdv = mean(allData$maxF.femaleAdvantage)
```

```r
if(AmaxF.meanFAdv<0){
  AmaxF.meanFAdv = paste(-round(AmaxF.meanFAdv,1),"fewer")
} else{
  AmaxF.meanFAdv = paste(round(AmaxF.meanFAdv,1),"more")
}
cat(AmaxF.meanFAdv,file="../results/latexStats/choice_maxF_meanAdv.tex")

# Same for max M strategy
AmaxMTest = t.test(allData$maxM.femaleAdvantage, mu = 0)
cat(round(AmaxMTest$p.value,2), file="../results/latexStats/choice_maxM_p.tex")
AmaxM.moreM = prop.table(table(allData$maxM.maleWords > allData$maxM.femaleWords))[2]
cat(round(AmaxM.moreM*100,1),file="../results/latexStats/choice_maxM_propMoreM.tex")
AmaxM.meanMAdv = -mean(allData$maxM.femaleAdvantage)
cat(round(AmaxM.meanMAdv,1),file="../results/latexStats/choice_maxM_meanAdv.tex")

# test whether there is a difference between strategies
maxM.Vs.MaxF = t.test(allData$maxF.femaleAdvantage,
                      allData$maxM.femaleAdvantage,paired = T)
px = round(maxM.Vs.MaxF$p.value,3)
if(px==0){
  px = "< 0.001"
} else{
  px = paste("=",px)
}
maxM.Vs.MaxFStat = paste0("t = ",round(maxM.Vs.MaxF$statistic,2),
                          ", p ",px)
cat(maxM.Vs.MaxFStat, file="../results/latexStats/choice_maxF_vs_maxM.tex")
```

Only 6 games out of 24 exhibit more than 50% female dialogue when attepmting to maximise female dialogue.

Across all games, a player trying to maximise female dialogue over male would succeed in observing more female dialogue than male dialogue in 35.7% of dialogue trees, on average seeing 10.2 more words spoken by females than males in each dialogue tree (not significantly different from zero, $p = 0$).

When maximising male dialogue over female, a player would succeed in observing more male dialogue than female in 64.6%, on average seeing 33.4 more words from males than females (highly significant, $p < 0.0001$).

Choice of gender for main player character

Some games allow players to choose the gender of their character. To what extent can this shift the proportion of dialogue for each gender?

Load all games with a "playerChoice" group, and work out the proportions if the player character is male or female:

```
folders = list.dirs("../data", recursive = T)
folders = folders[sapply(folders,function(X){
  "meta.json" %in% list.files(X)
})]

allG= NULL
for(folder in folders){
  js = fromJSON(file = paste0(folder,"/meta.json"))
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  statsFile = paste0(folder,"/stats_by_character.csv")
  if(file.exists(statsFile) & !alternativeMeasure){
    stats = read.csv(statsFile,stringsAsFactors = F)

    if("playerChoice" %in% stats$group){
      fWords = sum(stats[stats$group=="female",]$words,na.rm=T)
      mWords = sum(stats[stats$group=="male",]$words,na.rm=T)
      playerChoice = sum(stats[stats$group=="playerChoice",]$words,na.rm=T)

      allG = rbind(allG,
        data.frame(
          folder = stats$folder[1],
          game = stats$game[1],
          series = stats$series[1],
          propFemaleDialogueWhenPCIsMale = fWords / (fWords + mWords + playerChoice),
          propFemaleDialogueWhenPCIsFemale = (fWords + playerChoice) / (fWords + mWords + playerChoice)
        ))
      }
    }
}
allG = allG[order(allG$series),]
```

Summary of proportions when Player Character is male/female:

```
dxx = allG[,
      c("game","propFemaleDialogueWhenPCIsMale",
        "propFemaleDialogueWhenPCIsFemale")]
dxx[dxx$game == "Star Wars: Knights of the Old Republic",]$game =
  "Star Wars: KOTOR"
dxx$game = gsub("The Elder Scrolls .+?: ","",dxx$game)
knitr::kable(dxx)
```

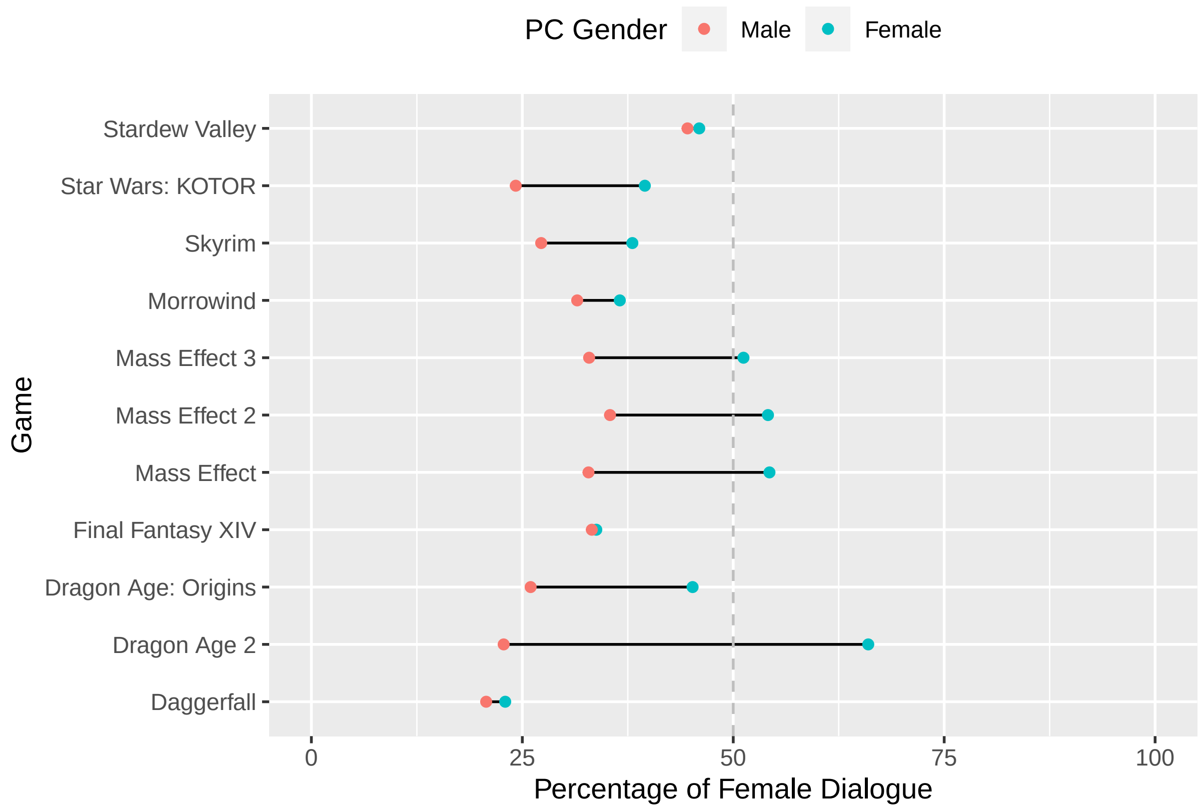
	game	propFemaleDialogueWhenPCIsMale	propFemaleDialogueWhenPCIsFemale
1	Dragon Age 2	0.2278847	0.6600277
2	Dragon Age: Origins	0.2598299	0.4518054
6	Final Fantasy XIV	0.3323781	0.3376912
7	Mass Effect	0.3284165	0.5429446
8	Mass Effect 2	0.3539331	0.5411705
9	Mass Effect 3	0.3291586	0.5121435
11	Star Wars: KOTOR	0.2422015	0.3952390
10	Stardew Valley	0.4457957	0.4596919
3	Daggerfall	0.2071458	0.2297331
4	Morrowind	0.3150428	0.3655770
5	Skyrim	0.2723502	0.3804433

Stats for paper:

```
numMoreThan50 = sum(allG$propFemaleDialogueWhenPCIsFemale>=0.5)
numMoreLess50 = sum(allG$propFemaleDialogueWhenPCIsFemale<0.5)
cat(nrow(allG),file="../results/latexStats/choice_PC_numGames.tex")
cat(nrow(numMoreThan50),file="../results/latexStats/choice_PC_numMoreThan50.tex")
cat(nrow(numMoreLess50),file="../results/latexStats/choice_PC_numMoreLess50.tex")

highestGame = allG[order(allG$propFemaleDialogueWhenPCIsFemale,decreasing = T),]$game[1]
highestPer = round(100*allG[
  order(allG$propFemaleDialogueWhenPCIsFemale,
        decreasing = T),]$propFemaleDialogueWhenPCIsFemale[1],1)
lowestPer =  round(100*allG[
  order(allG$propFemaleDialogueWhenPCIsFemale,
        decreasing = T),]$propFemaleDialogueWhenPCIsMale[1],1)
cat(highestGame,file="../results/latexStats/choice_PC_bestGame.tex")
cat(highestPer,file="../results/latexStats/choice_PC_bestGamePercentFemaleDialogue.tex")
cat(lowestPer,file="../results/latexStats/choice_PC_bestGamePercentFemaleDialogue_ifMale.tex")
```

```
gd = data.frame(Game = rep(dxx$game,2))
gd$`Percentage of Female Dialogue` =
      100* c(dxx$propFemaleDialogueWhenPCIsFemale,
           dxx$propFemaleDialogueWhenPCIsMale)
gd$`PC Gender` = rep(c("Female","Male"),each=nrow(dxx))
gd$`PC Gender` = factor(gd$`PC Gender`,levels=c("Male","Female"))
ggplot(gd,aes(y=`Percentage of Female Dialogue`,
              x=Game, colour=`PC Gender`)) +
  geom_line(aes(group=Game),colour='black')+
  geom_point() +
  scale_y_continuous(limits = c(0,100),
                     breaks=c(0,25,50,75,100))+
```

```
geom_hline(yintercept = 50,linetype="dashed",col="gray")+
coord_flip() +
theme(legend.position = "top")
```



Summary

11 games in our sample allow the player to choose the gender of their character. Of these, only 4 games have more than 50% female dialogue when the player character was female, the highest being for Dragon Age 2 with 66% versus 22.8% female dialogue if the player character is male.

9 Transitions between characters

Gender biases in dialogue transitions

Introduction

As well as the amount of dialogue given to female characters, studies of film and television have shown biases in who they speak to. The “Bechdel test” or “Bechdel-Wallace test” is a popular illustration of a bias against females talking to other females. The test originated as an idea in a comic strip (Bechdel, 1986), and has become more widely known. A film passes the Bechdel test if it depicts:

1. At least two named female characters …
2. … who talk to each other
3. … about something other than a man.

For example, we are reasonably sure that *The Secret of Monkey Island* does not pass the Bechdel test, simply because there are so few female characters and there are no scenes with two female characters in. According to our corpus, there apparently is only one case of a transition between two females in any of the three Monkey Island games, though this turns out to be the following from Monkey Island 2:

> “Woman 1 watching spit contest”: “(claps)”
>
> “Woman 2 watching spit contest”: “(claps)”

There is no linguistic content, and these women are not named. So all three games apparently fail the Bechdel test.

The Bechdel test has been influential and revealing, but there are issues with applying it as a measure of gender bias in video game dialogue. The first issue is that the Bechdel test is very strict. The data in our corpus is often a sample, and sometimes only one possible play-through. Other games have procedural elements. For any given game, it would be very hard to prove that there is *no possibility* of dialogue between two women.

The second issue is that video games often have more dialogue than the average TV show or film. So the probability of there being *no* female-female dialogue is very low. So most games would pass the test. However, this doesn’t mean that there are no problems, and it doesn’t show which games have more problems than others.

The third issue is that it’s possible for a game to potentially pass the Bechdel test, while not actually passing it for a specific player’s experience. The test is well suited to short scripts with canonical forms (like TV, Film, books), but conceptually more difficult to apply to interactive texts like video games.

There are alternative measures, such as the “Mako Mori test” which test whether a female character has a narrative ark. However, this also has the issues above.

Therefore, for this study it makes more sense to apply a probabilistic test that shows a bias away from an expected value. This would help quantify the relative degree of bias between games. So, in the following sections, we estimate the frequencies of dialogue transitions between characters of different genders and assess these in relation to baselines calculated using a permutation procedure.

Measuring transitions

Data for most games in the corpus captures the local order of dialogue. That is, while players might experience scenes in a different order (e.g. by choosing to speak to characters in a different order), the order within a scene is consistent. For these games, we want to count the number of dialogue transitions between characters of different genders. This can be done by iterating through the game script and counting transitions from one character to the next. This is similar to previous “consecutive” approaches to automating the Betchdel test that build an interaction network between characters based on a consecutive lines of dialogue (Agarwal et al., 2015).

However, neither games nor films have a single continuous conversation. Transitions should only be counted for characters who are actually talking *to each other*. Automated approaches applied to film assume that all characters within a scene are effectively talking to each other, so count only the transitions that happen within scenes (Weng et al., 2009; Agarwal et al., 2015). Some games have scene or location transitions, encoded in "LOCATION" lines. During parsing, games where conversation boundaries were detectable but not labelled were indicated with an ACTION line with the value "—". However, scenes within films are often much shorter than in games (there may be long periods in a game without a narrative "cut"). Furthermore, dialogue is often in short bursts, separated by longer game playing sections. Therefore, in addition to scenes and locations, we treat action descriptions as scene boundaries.

A python script (`processing/DialogueTransitions.py`) stepped through the script, identifying dialogue transitions between two non-identical characters. It looked up these characters' relative gender categories, and incremented a count for transitions from the first to the second category. In dialogue trees with choices, the algorithm counted all possible dialogue transitions that a player might experience (considering only branching options without re-visiting earlier parts in the tree). The script output the frequency of transitions between each gender category (stored as `transitions.csv`) and the sequence of individual speaker categories and scene breaks (stored as `transitions_all.csv`) for use in the premutation test (see below).

The automated method here will obviously not always be accurate. Agarwal et al., report that their method of scene boundary detection has accuracies of around 65% for identifying whether women talk to each other, so it is likely that there will be many errors. However, the aim of the current study is to identify large-scale gender biases in lingusitic behaviour. So we proceed with the assumption that the errors will be unbiased by gender. That is, the scene boundary cues we use are not distributed differently depending on the gender of the surrounding characters.

Calculating expected transitions

For each game, we now have the empirical number of transitions between each gender category. For example, Final Fantasy VII has the following transition frequencies, where the rows indicate the gender of the character before the transition and the columns indicate the gender of the character after the transition:

```
##        male female
## male   2666   1150
## female 1199    225
```

The frequency of female-to-female dialogue is low, but is it significantly lower than we would expect by chance? This is more difficult to answer than it might appear. A game that has no biases in transitions between female characters won't necessarily have an equal number of each type of transition (female-to-female, female-to-male, etc.). For example, imagine a game with many more male characters than female characters. It is likely that this would have more male-to-male transitions than female-to-female transitions, even if the creators were designing many opportunities for female characters to talk to each other. So while a low proportion of female-to-female transitions might be evidence for a general bias against female dialogue, it is not necessarily evidence for a bias against *transitions between* female characters. This requires comparing the transitions we observe to a "baseline" of values we would expect by chance if there was no bias in who talks to who.

The calculation of this baseline is difficult to compute analytically, since it depends on the number of male and female lines, and the number of scene boundaries. Instead, we use permutation: the full sequence of gender categories and scene boundaries for a game is randomly re-shuffled. This creates a hypothetical script with the same numbers of lines for each gender category and the same number of scene boundaries. This step is repeated many times to produce many scripts (10,000 times for the analysis of all games together, and 1,000 times for each individual game analysis). On average, we would expect these hypothetical scripts to break any implicit gender biases in dialogue transitions (they should have an 'unbiased' frequency of transitions between each gender category). The frequency of transitions are be re-calculated for each hypothetical script to get a distribution of permuted transition frequencies. These can be compared to the "real" frequencies. From this we can calculate the following for a given transition between gender categories:

- The mean frequency of permuted transitions, which serve as the expected values if there was no gender bias.

- A z-score which represents a normalised distance between the “real” value and the expected values.
- A p-value: The p-value represents the proportion of hypothetical scripts that exhibited a more extreme frequency than the empirical frequency. This serves as an indicator of the likelihood that the empirical frequency would be produced by chance.

There are four possible transition types between two genders, but the frequencies for these are obviously not independent of each other. Therefore, we focus on two measures: the proportion of female-to-female transitions (to reflect the Bechdel test), and the proportion of male-to-male transitions (as the orthogonal measure). In principle, these measures are relatively independent, so a game could have a high or low value for either measure.

Load libraries

```
library(rjson)
library(Gmisc)
library(dplyr)
library(RColorBrewer)
library(grid)
library(ggplot2)
library(ggrepel)
```

Set some parameters for the permutation tests:

```
numberOfPermutationsForAllGames = 10000
numberOfPermutationsForEachGame = 1000
```

Load data

```
folders = list.dirs("../data", recursive = T)
folders = folders[sapply(folders,function(X){
  "transitions.csv" %in% list.files(X)
})]


allGames= NULL
transitionStrings = NULL
for(folder in folders){
  print(folder)
  shortName = tail(strsplit(folder,"/")[[1]],1)
  js = fromJSON(file = paste0(folder,"/meta.json"))
  alternativeMeasure = FALSE
  if(!is.null(js$alternativeMeasure)){
    alternativeMeasure = js$alternativeMeasure
  }
  suitableForTransitions = TRUE
  if(!is.null(js$sourceFeatures$dialogueOrder)){
      suitableForTransitions = js$sourceFeatures$dialogueOrder
  }

  if(suitableForTransitions & (!alternativeMeasure)){
    sbcFile = paste0(folder,"/transitions.csv")
    if(file.exists(sbcFile)){
      d = read.csv(sbcFile,stringsAsFactors = F)
      if(nrow(d)>0){
        d$shortName = shortName
      }
      allGames = rbind(allGames,d)
    }
    tsFile = paste0(folder,"/transitions_all.txt")
    if(file.exists(tsFile)){
      ts = suppressWarnings(readLines(tsFile)[1])
      transitionStrings = rbind(transitionStrings,
                    data.frame(folder=folder,series=d$series[1],
                               game=d$game[1],
                               ts = ts,
                               stringsAsFactors = F))
    }
  }
```

```
}
```

```
## [1] "../data/ChronoTrigger/ChronoTrigger"
## [1] "../data/DragonAge/DragonAge2"
## [1] "../data/DragonAge/DragonAgeInquisition"
## [1] "../data/DragonAge/DragonAgeOrigins_B"
## [1] "../data/ElderScrolls/Daggerfall"
## [1] "../data/ElderScrolls/Morrowind"
## [1] "../data/ElderScrolls/Oblivion"
## [1] "../data/ElderScrolls/Skyrim"
## [1] "../data/FinalFantasy/FFI"
## [1] "../data/FinalFantasy/FFII"
## [1] "../data/FinalFantasy/FFIII"
## [1] "../data/FinalFantasy/FFIV_DS"
## [1] "../data/FinalFantasy/FFIX"
## [1] "../data/FinalFantasy/FFIX_B"
## [1] "../data/FinalFantasy/FFV"
## [1] "../data/FinalFantasy/FFVI"
## [1] "../data/FinalFantasy/FFVII"
## [1] "../data/FinalFantasy/FFVII_Remake"
## [1] "../data/FinalFantasy/FFVIII"
## [1] "../data/FinalFantasy/FFX"
## [1] "../data/FinalFantasy/FFX_B"
## [1] "../data/FinalFantasy/FFX2"
## [1] "../data/FinalFantasy/FFXII"
## [1] "../data/FinalFantasy/FFXII_B"
## [1] "../data/FinalFantasy/FFXIII"
## [1] "../data/FinalFantasy/FFXIII-2"
## [1] "../data/FinalFantasy/FFXIII-LR"
## [1] "../data/FinalFantasy/FFXIV"
## [1] "../data/FinalFantasy/FFXV"
## [1] "../data/Horizon/HorizonZeroDawn"
## [1] "../data/KingdomHearts/KingdomHearts"
## [1] "../data/KingdomHearts/KingdomHearts_B"
## [1] "../data/KingdomHearts/KingdomHearts2"
## [1] "../data/KingdomHearts/KingdomHearts3"
## [1] "../data/KingdomHearts/KingdomHearts3D"
## [1] "../data/KingsQuest/KingsQuest1"
## [1] "../data/KingsQuest/KingsQuest2"
## [1] "../data/KingsQuest/KingsQuest3"
## [1] "../data/KingsQuest/KingsQuest4"
## [1] "../data/KingsQuest/KingsQuest5"
## [1] "../data/KingsQuest/KingsQuest6"
## [1] "../data/KingsQuest/KingsQuest7"
## [1] "../data/KingsQuest/KingsQuest8"
## [1] "../data/KingsQuest/KingsQuestChapters"
## [1] "../data/MassEffect/MassEffect1"
## [1] "../data/MassEffect/MassEffect1B"
## [1] "../data/MassEffect/MassEffect2"
## [1] "../data/MassEffect/MassEffect3"
## [1] "../data/MassEffect/MassEffect3C"
## [1] "../data/MassEffect/MassEffectAndromeda"
## [1] "../data/MonkeyIsland/MonkeyIsland2"
## [1] "../data/MonkeyIsland/TheCurseOfMonkeyIsland"
## [1] "../data/MonkeyIsland/TheSecretOfMonkeyIsland"
## [1] "../data/Persona/Persona3"
## [1] "../data/Persona/Persona4"
## [1] "../data/Persona/Persona5"
```

```
## [1] "../data/Persona/Persona5B"
## [1] "../data/StardewValley/StardewValley"
## [1] "../data/StarWarsKOTOR/StarWarsKOTOR"
## [1] "../data/SuperMarioRPG/SuperMarioRPG"
```

```
# Trim final stroke to make consistent
transitionStrings$folder = gsub("/$","",transitionStrings$folder)
allGames$folder = gsub("/$","",allGames$folder)
```

Analogue of the Betchdel test: try to find games with no transitions between two female characters:

```
hasFemaleToFemale = sapply(unique(allGames$folder), function(fld){
  any(allGames[allGames$folder==fld,]$from=="female" & allGames[allGames$folder==fld,]$to=="female",
})

t(t(names(hasFemaleToFemale[!hasFemaleToFemale])))
```

```
##      [,1]
## [1,] "../data/FinalFantasy/FFI"
## [2,] "../data/MassEffect/MassEffect1B"
## [3,] "../data/MassEffect/MassEffect2"
## [4,] "../data/MassEffect/MassEffect3C"
## [5,] "../data/MonkeyIsland/TheCurseOfMonkeyIsland"
## [6,] "../data/MonkeyIsland/TheSecretOfMonkeyIsland"
```

Permutation tests

Function to count different types of transition and run the permutation tests:

```
getTransitions = function(allGames, raw=F){

  maleToMale = sum(allGames[allGames$from=="male" & allGames$to=="male",]$frequency)
  maleToFemale = sum(allGames[allGames$from=="male" & allGames$to=="female",]$frequency)
  femaleToMale = sum(allGames[allGames$from=="female" & allGames$to=="male",]$frequency)
  femaleToFemale = sum(allGames[allGames$from=="female" & allGames$to=="female",]$frequency)

  if(raw){
    maleToMaleP = maleToMale
    maleToFemaleP = maleToFemale
    femaleToMaleP = femaleToMale
    femaleToFemaleP = femaleToFemale
  }
  else{
    total = sum(maleToMale,maleToFemale,femaleToMale,femaleToFemale)
    maleToMaleP = maleToMale/(maleToMale+maleToFemale)
    maleToFemaleP = maleToFemale/(maleToMale+maleToFemale)
    femaleToMaleP = femaleToMale/(femaleToMale+femaleToFemale)
    femaleToFemaleP = femaleToFemale/(femaleToMale+femaleToFemale)
  }

  transitionTable =matrix(c(maleToMaleP, femaleToMaleP, maleToFemaleP, femaleToFemaleP),nrow=2)
  rownames(transitionTable) = c("m","f")
  colnames(transitionTable) = c("m","f")

  return(transitionTable)

}
```

Functions to extract and print the stats:

```r
permuteTransitionString = function(X){
  # permute order
  X = sample(X)
  # calculate table of transitions using lag
  tab = table(X[1:(length(X)-1)], X[2:length(X)])
  tab = tab[c("m","f"),c("m","f")]
  tab[is.na(tab)] = 0
  transitionProbs = prop.table(tab,1)
}

getZP = function(tpPerm,trueProb){
  Z = (trueProb - mean(tpPerm))/sd(tpPerm)
  sx = sum(tpPerm >trueProb)
  P = 1/length(tpPerm)
  if(sx>0){
    P = sx/length(tpPerm)
  }
  if(Z<0){
    P = 1 - P
  }
  if(P==0){
    P = 1/length(tpPerm)
  }
  return(c(mean = mean(tpPerm), z = Z,p = P))
}

getPermutedStats = function(ts, trueTransitionProbs,numPerm){
  lines = strsplit(ts,"")[[1]]
  transProbsPerm = replicate(numPerm, permuteTransitionString(lines))
  m2m = getZP(transProbsPerm["m","m",],trueTransitionProbs["m","m"])
  m2f = getZP(transProbsPerm["m","f",],trueTransitionProbs["m","f"])
  f2f = getZP(transProbsPerm["f","f",],trueTransitionProbs["f","f"])
  f2m = getZP(transProbsPerm["f","m",],trueTransitionProbs["f","m"])

  return(c(m2m, f2m, m2f, f2f))
}
```

Run transition permutation tests for the corpus as a whole:

```
trueTransitions.AllGames = getTransitions(allGames)
trueTransitions.AllGames.Raw = getTransitions(allGames,raw=TRUE)

trueTransitionsString.AllGames = paste0(transitionStrings$ts)
permutedTransitionStats.AllGames =
  getPermutedStats(trueTransitionsString.AllGames,
                   trueTransitions.AllGames,
                   numberOfPermutationsForAllGames)
names(permutedTransitionStats.AllGames) = paste(
  rep(c("m2m", "f2m", "m2f", "f2f"),each=3),
  names(permutedTransitionStats.AllGames),sep=".")

t(t(permutedTransitionStats.AllGames))
```

```
##                  [,1]
## m2m.mean   0.5678064
## m2m.z     10.3953934
## m2m.p      0.0001000
## f2m.mean   0.5681016
## f2m.z      9.2559500
## f2m.p      0.0001000
## m2f.mean   0.4321936
## m2f.z    -10.3953934
## m2f.p      0.0001000
## f2f.mean   0.4318984
## f2f.z     -9.2559500
## f2f.p      0.0001000
```

Stats for each game:

```
set.seed(238)
print("Running stats ...")
```

```
## [1] "Running stats ..."
```

```
permutationResults = NULL
for(folder in unique(allGames$folder)){
  print(folder)
  dx = allGames[allGames$folder == folder,]
  trueTransitionProbs = getTransitions(dx)
  trueTransitionProbsFlat = as.vector(trueTransitionProbs)
  names(trueTransitionProbsFlat) = c("m2m", "f2m", "m2f", "f2f")

  trueTransitionRaw = getTransitions(dx,raw = T)
  trueTransitionRawFlat = as.vector(trueTransitionRaw)
  names(trueTransitionRawFlat) = c("m2m", "f2m", "m2f", "f2f")

  if(any(is.nan(trueTransitionProbs))){
    print(paste("No data for ",folder))
  } else{
    trueTransitionString = transitionStrings[transitionStrings$folder==folder,]$ts
    permutedTransitionStats = getPermutedStats(trueTransitionString,trueTransitionProbs,numberOfPerm
    names(permutedTransitionStats) = paste(rep(c("m2m", "f2m", "m2f", "f2f"),each=3),names(permutedT

    res = data.frame(folder=dx$folder[1],
                     series = dx$series[1],
                     game = dx$game[1],
                     shortName = dx$shortName[1])
```

```r
    tt = matrix(trueTransitionProbsFlat,nrow=1)
    colnames(tt) = names(trueTransitionProbsFlat)
    res = cbind(res,tt)

    ttr = matrix(trueTransitionRawFlat,nrow=1)
    colnames(ttr) = paste0(names(trueTransitionRawFlat),".raw")
    res = cbind(res,ttr)

    pt = matrix(permutedTransitionStats,nrow=1)
    colnames(pt) = names(permutedTransitionStats)
    res = cbind(res,pt)

    permutationResults = rbind(permutationResults,res)
  }

}
```

```
## [1] "../data/ChronoTrigger/ChronoTrigger"
## [1] "../data/DragonAge/DragonAge2"
## [1] "../data/DragonAge/DragonAgeOrigins_B"
## [1] "../data/ElderScrolls/Skyrim"
## [1] "../data/FinalFantasy/FFI"
## [1] "../data/FinalFantasy/FFII"
## [1] "../data/FinalFantasy/FFIV_DS"
## [1] "../data/FinalFantasy/FFIX_B"
## [1] "../data/FinalFantasy/FFV"
## [1] "../data/FinalFantasy/FFVI"
## [1] "../data/FinalFantasy/FFVII"
## [1] "../data/FinalFantasy/FFVII_Remake"
## [1] "../data/FinalFantasy/FFVIII"
## [1] "../data/FinalFantasy/FFX_B"
## [1] "../data/FinalFantasy/FFX2"
## [1] "../data/FinalFantasy/FFXII_B"
## [1] "../data/FinalFantasy/FFXIII"
## [1] "../data/FinalFantasy/FFXIII-2"
## [1] "../data/FinalFantasy/FFXIII-LR"
## [1] "../data/FinalFantasy/FFXIV"
## [1] "../data/FinalFantasy/FFXV"
## [1] "../data/Horizon/HorizonZeroDawn"
## [1] "../data/KingdomHearts/KingdomHearts_B"
## [1] "../data/KingdomHearts/KingdomHearts2"
## [1] "../data/KingdomHearts/KingdomHearts3"
## [1] "../data/KingdomHearts/KingdomHearts3D"
## [1] "../data/KingsQuest/KingsQuest6"
## [1] "../data/KingsQuest/KingsQuest7"
## [1] "../data/MassEffect/MassEffect1B"
## [1] "No data for  ../data/MassEffect/MassEffect1B"
## [1] "../data/MassEffect/MassEffect2"
## [1] "../data/MassEffect/MassEffect3C"
## [1] "No data for  ../data/MassEffect/MassEffect3C"
## [1] "../data/MonkeyIsland/MonkeyIsland2"
## [1] "../data/MonkeyIsland/TheCurseOfMonkeyIsland"
## [1] "../data/MonkeyIsland/TheSecretOfMonkeyIsland"
## [1] "../data/Persona/Persona3"
## [1] "../data/Persona/Persona4"
## [1] "../data/Persona/Persona5B"
## [1] "../data/StardewValley/StardewValley"
## [1] "../data/StarWarsKOTOR/StarWarsKOTOR"
```

```
## [1] "../data/SuperMarioRPG/SuperMarioRPG"
```

```
permutationResults$diffExpEmp.m2m = permutationResults$m2m - permutationResults$m2m.mean
permutationResults$diffExpEmp.f2f = permutationResults$f2f - permutationResults$f2f.mean
```

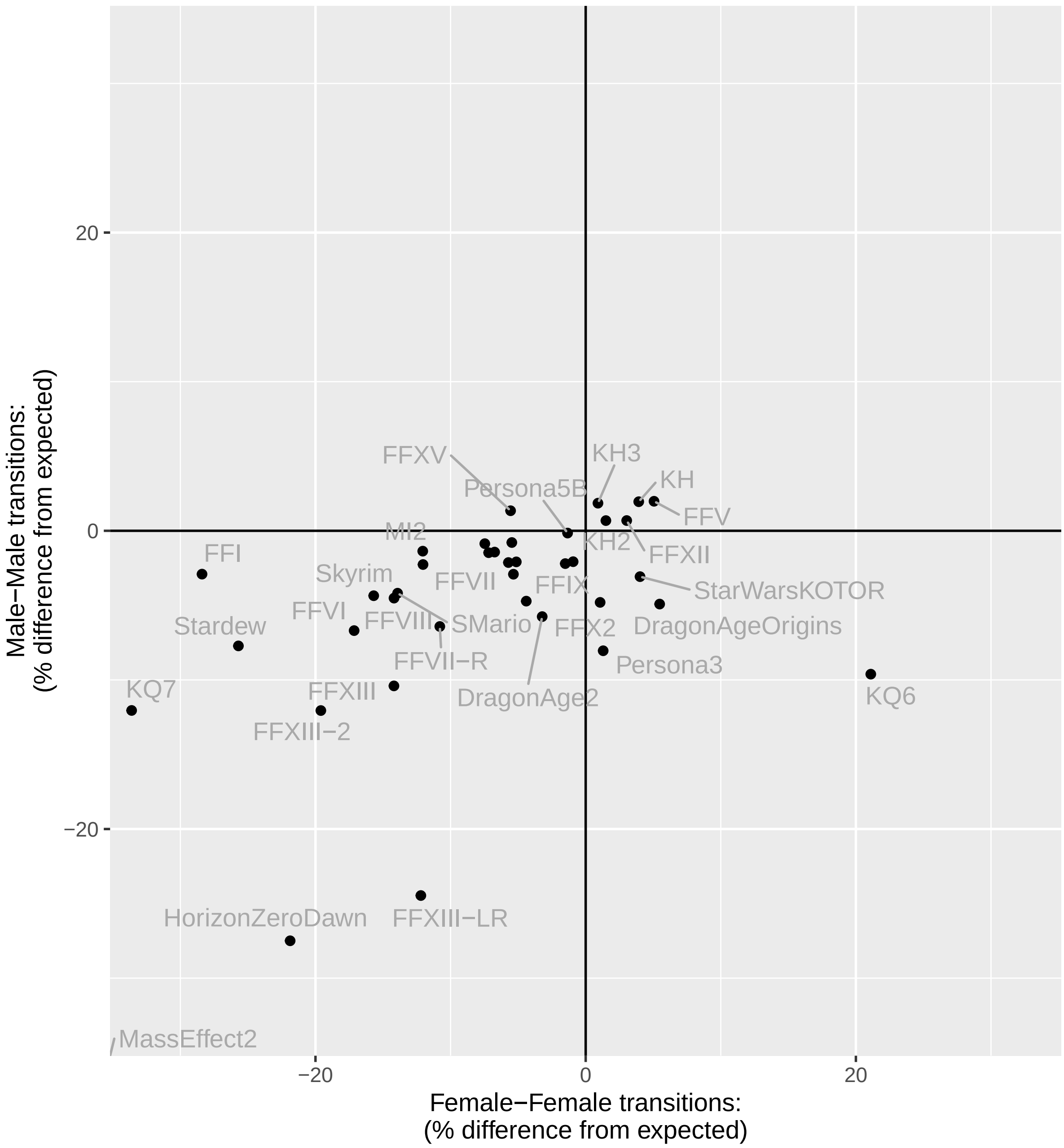
Write the data:

```
write.csv(permutationResults, "../results/transitionsPermutationTest.csv",row.names = F)
```

Plot difference from expected

```
permutationResults$shortName2 = permutationResults$shortName
permutationResults$shortName2 = gsub("KingdomHearts","KH",permutationResults$shortName2)
permutationResults$shortName2 = gsub("KingsQuest","KQ",permutationResults$shortName2)
permutationResults$shortName2 = gsub("_Remake","-R",permutationResults$shortName2)
permutationResults$shortName2 = gsub("_B","",permutationResults$shortName2)
permutationResults$shortName2 = gsub("_DS","",permutationResults$shortName2)
permutationResults$shortName2[permutationResults$shortName2=="TheSecretOfMonkeyIsland"] = "MI1"
permutationResults$shortName2[permutationResults$shortName2=="MonkeyIsland2"] = "MI2"
permutationResults$shortName2[permutationResults$shortName2=="TheCurseOfMonkeyIsland"] = "MI3"
permutationResults$shortName2[permutationResults$shortName2=="SuperMarioRPG"] = "SMario"
permutationResults$shortName2[permutationResults$shortName2=="StardewValley"] = "Stardew"
permutationResults$shortName2[permutationResults$shortName2=="FFX_B"] = "FFX"

permResPlot =   ggplot(permutationResults,
    aes(x=diffExpEmp.f2f*100,y=diffExpEmp.m2m*100)) +
    geom_point() +
    coord_cartesian(ylim=c(-32,32),xlim=c(-32,32))+
    ylab("Male-Male transitions:\n(% difference from expected)") +
    geom_hline(yintercept = 0) + geom_vline(xintercept = 0) +
    xlab("Female-Female transitions:\n(% difference from expected)") +
    geom_text_repel(aes(label=shortName2),color="dark gray",force = 10)
permResPlot
```

```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```




```r
pdf('../results/graphs/transitions/Transitions_DiffFromExpected.pdf', width=6,height=6)
permResPlot
```

```
## Warning: ggrepel: 12 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

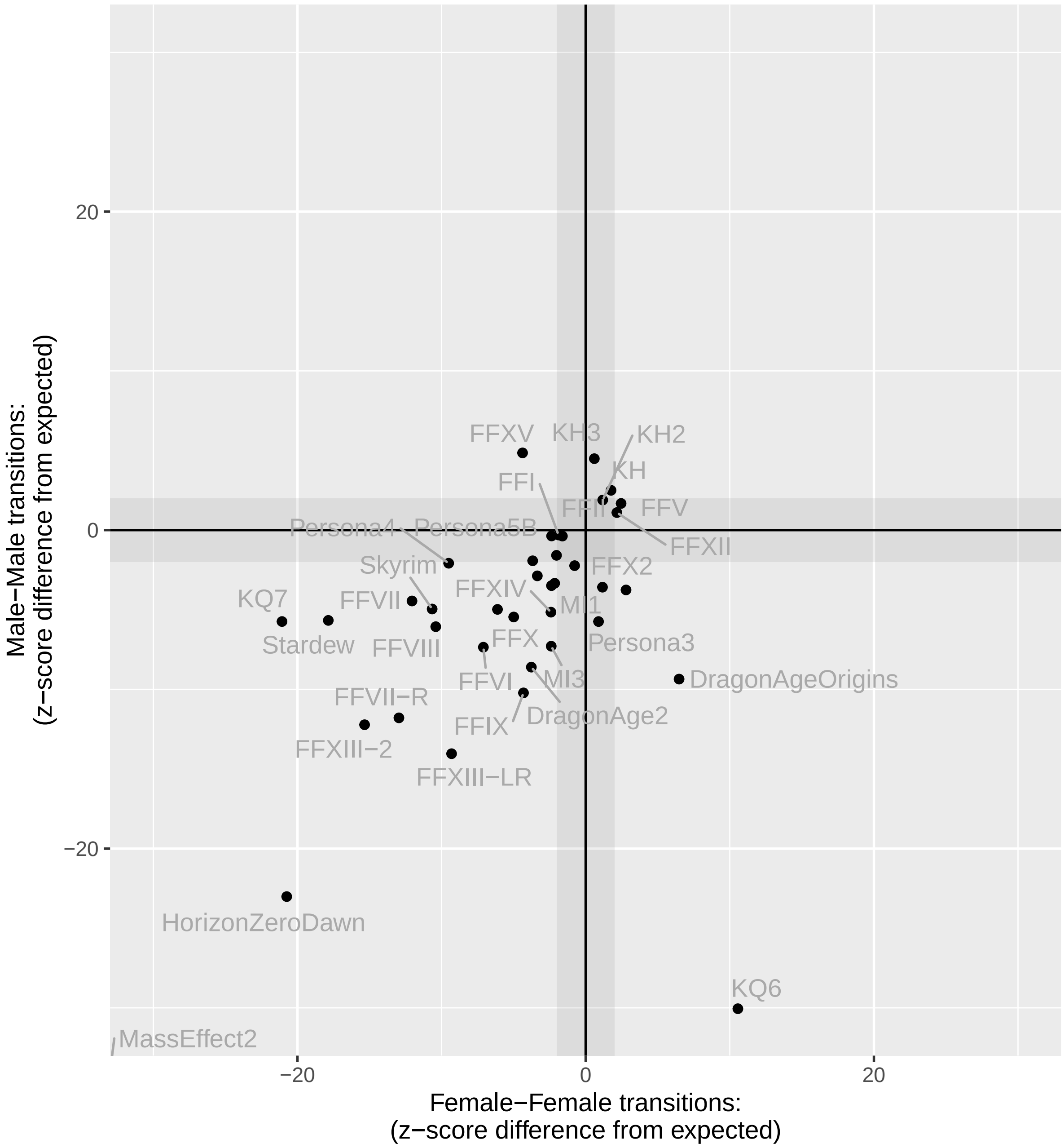
```r
dev.off()
```

```
## pdf
##   2
```

```r
  ggplot(permutationResults,aes(x=f2f.z,y=m2m.z)) +
    annotate("rect", xmin = -1000, xmax = 1000, ymin = -2, ymax = 2, alpha = .1) +
    annotate("rect", xmin = -2, xmax = 2, ymin = -1000, ymax = 1000, alpha = .1) +
    geom_point() +
    coord_cartesian(ylim=c(-30,30),xlim=c(-30,30))+
    ylab("Male-Male transitions:\n(z-score difference from expected)") +
    geom_hline(yintercept = 0) + geom_vline(xintercept = 0) +
```

```
    xlab("Female-Female transitions:\n(z-score difference from expected)") +
    geom_text_repel(aes(label=shortName2),color="dark gray",force = 10)
```

```
## Warning: ggrepel: 7 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```




```
xs = c(-20,-10,0,10,20)
ls = log(100+xs)
permResultsZScore = ggplot(permutationResults,aes(x=log(100+f2f.z),y=log(100+m2m.z))) +
    annotate("rect", xmin = -1000, xmax = 1000, ymin = log(100-2), ymax = log(102), alpha = .1) +
    annotate("rect", xmin = log(100-2), xmax = log(102), ymin = -1000, ymax = 1000, alpha = .1) +
    geom_point() +
    coord_cartesian(ylim=c(4.25,4.9),xlim=c(4.25,4.9))+
    ylab("Male-Male transitions:\n(z-score difference from expected)") +
 #   geom_hline(yintercept = 0) + geom_vline(xintercept = 0) +
    xlab("Female-Female transitions:\n(z-score difference from expected)") +
    geom_text_repel(aes(label=shortName2),color="dark gray",force = 10) +
```

```
    scale_x_continuous(breaks=ls,labels=xs,
       sec.axis = sec_axis(~.*1,
                           breaks = ls,
                           labels=c("Fewer female-female transitions\nthan expected","",
                                    "As expected","",
                                    "More female-female transitions\nthan expected")))+
  scale_y_continuous(breaks=ls,labels=xs) +
    theme(panel.grid.minor = element_blank())



pdf("../results/graphs/transitions/Transitions_DiffFromExpected_ZScores.pdf",width=6.5,height=6)
permResultsZScore
```

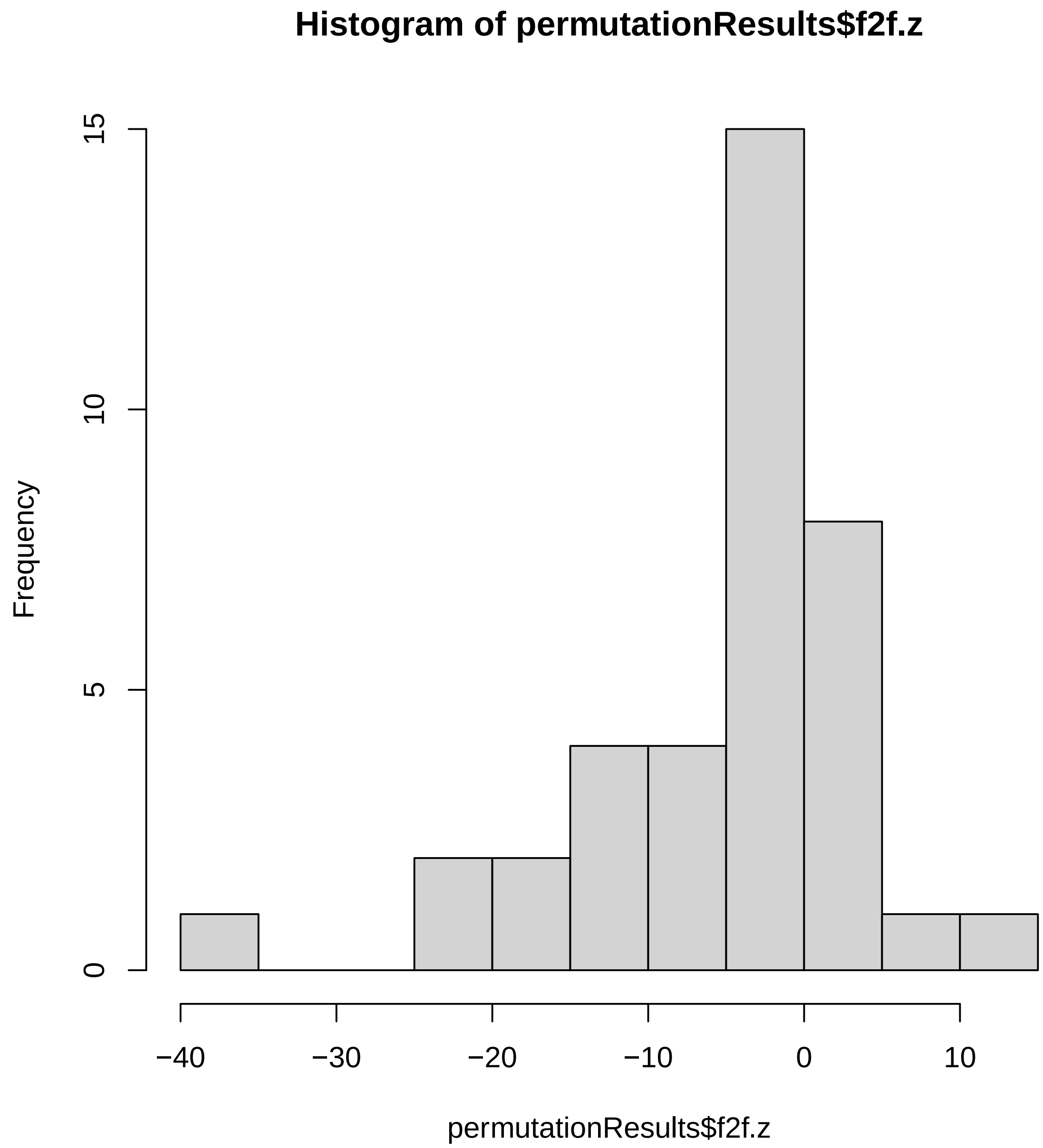
```
## Warning: ggrepel: 17 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

```
dev.off()
```

```
## pdf
##   2
```

In general, transitions within genders is lower than expected, while transitions between genders is higher than expected.

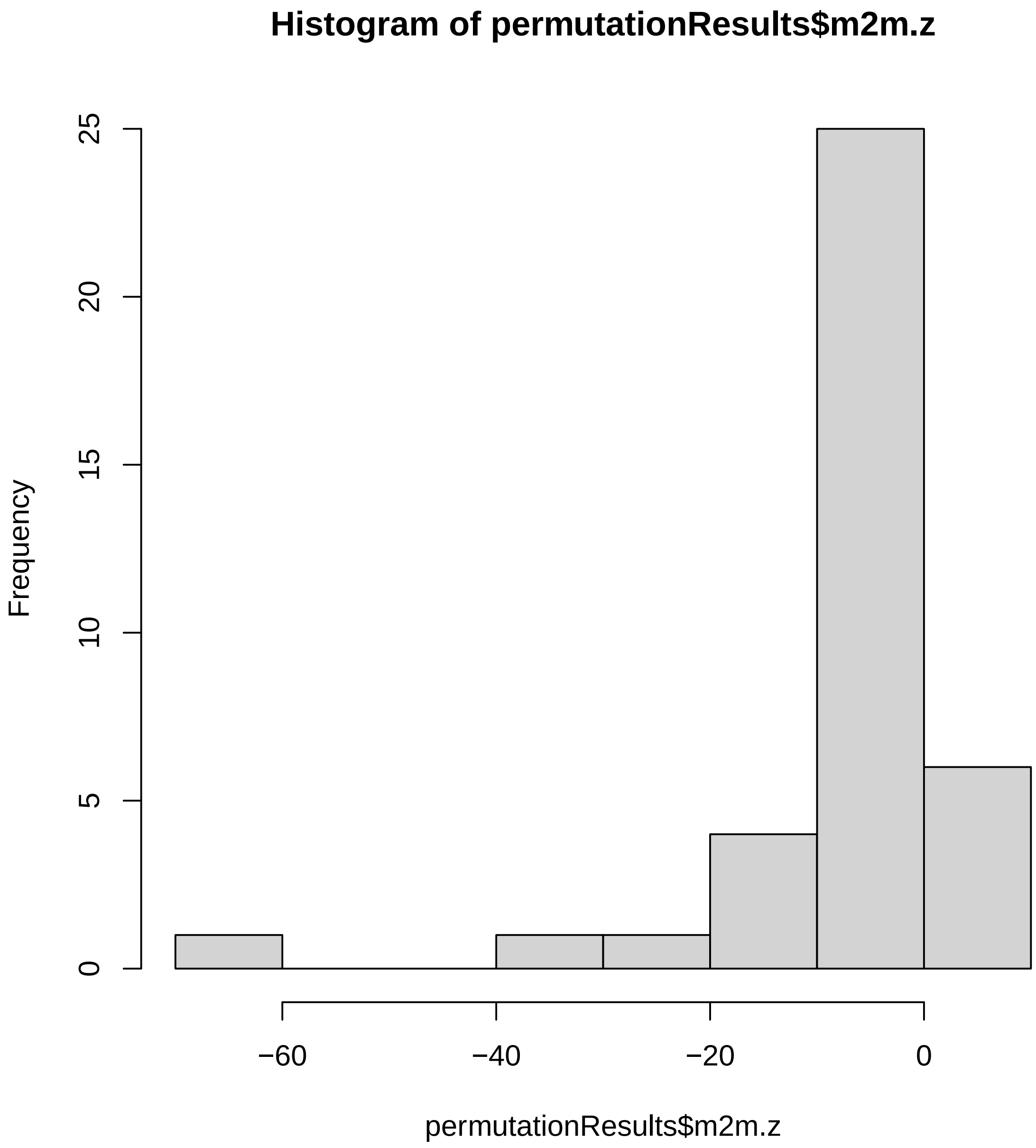
```
hist(permutationResults$f2f.z)
```




```
mean(permutationResults$f2f.z)
```

```
## [1] -5.348596
```

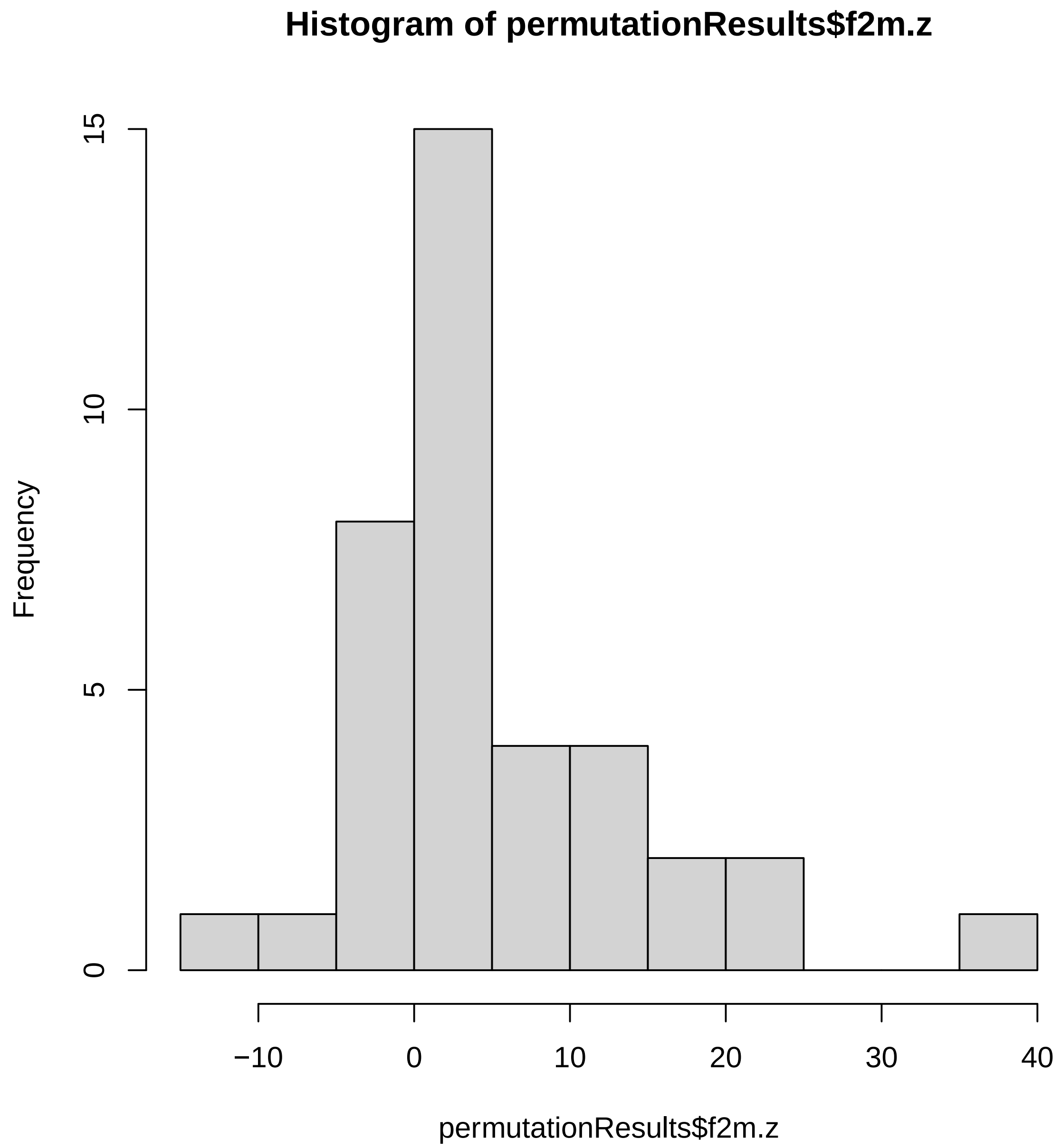
```
hist(permutationResults$m2m.z)
```




```
mean(permutationResults$m2m.z)
```

```
## [1] -6.826839
```

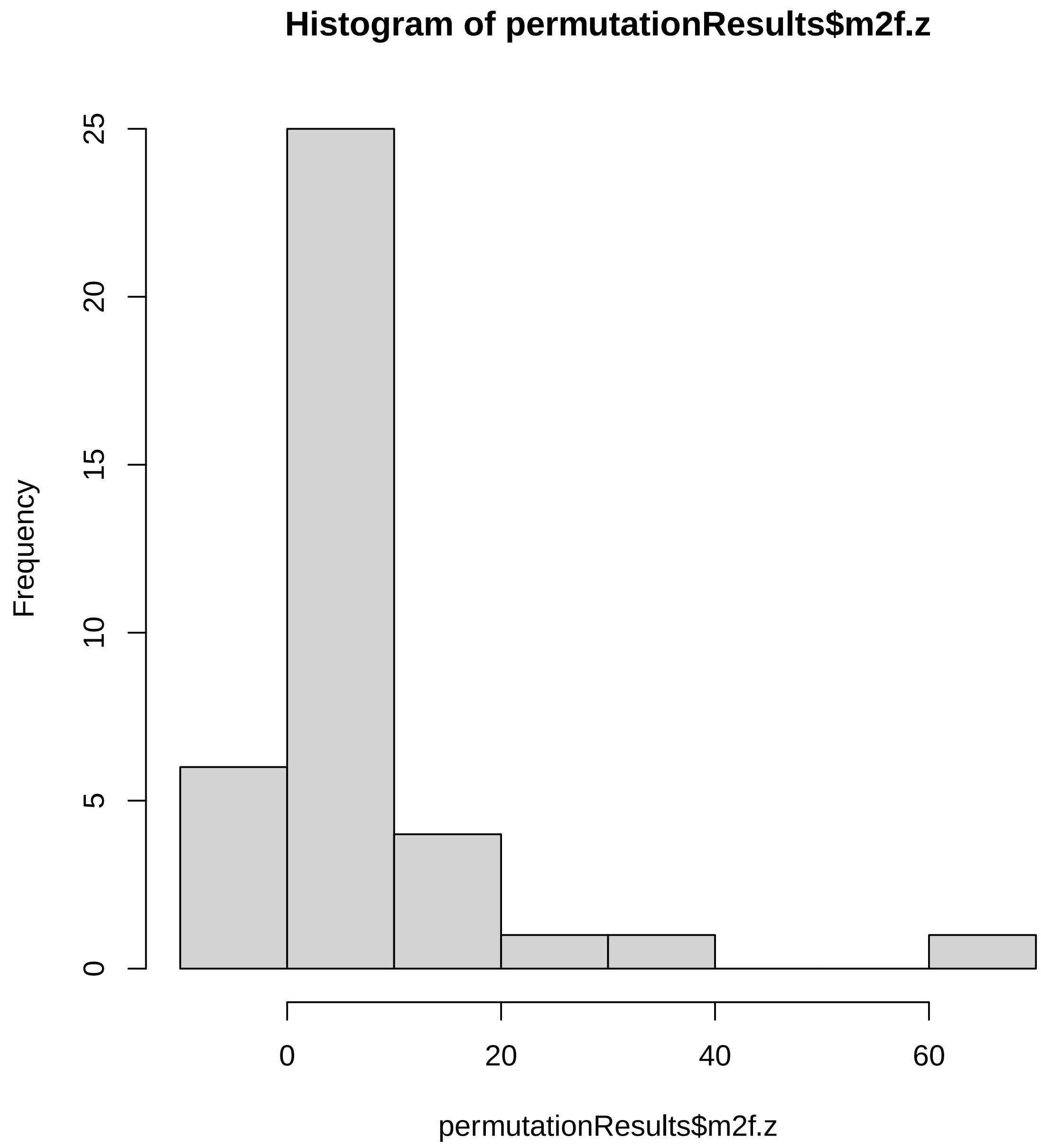
```
hist(permutationResults$f2m.z)
```




```
mean(permutationResults$f2m.z)
```

```
## [1] 5.348596
```

```
hist(permutationResults$m2f.z)
```




```
mean(permutationResults$m2f.z)
```

```
## [1] 6.826839
```

Number of games with within lower than expected female-to-female transitions:

```
lowF2F = table(permutationResults$f2f.z < 0 &
                 permutationResults$f2f.p<0.05)
lowM2M = table(permutationResults$m2m.z < 0 &
                 permutationResults$m2m.p<0.05)
lowF2F
```

```
##
## FALSE  TRUE
##    13    25
```

```
lowM2M
```

```
##
```

```
## FALSE  TRUE
##    10    28
```

25 games had significantly lower female-to-female transitions than expected by chance, and 25 games had significantly lower male-to-male transitions than expected by chance.

```
binom.test(table(permutationResults$f2f.z < permutationResults$m2m.z))
```

```
##
##  Exact binomial test
##
## data:  table(permutationResults$f2f.z < permutationResults$m2m.z)
## number of successes = 19, number of trials = 38, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.333789 0.666211
## sample estimates:
## probability of success
##                    0.5
```

```
# Some imbalances:
table(permutationResults$m2m.z < 0 & permutationResults$m2m.p<0.05,
      permutationResults$f2f.z < 0 & permutationResults$f2f.p<0.05)
```

```
##
##         FALSE TRUE
##   FALSE     7    3
##   TRUE      6   22
```

```
# Game with less m-m and more f-f than expected:
permutationResults[(permutationResults$m2m.z <0 &
                      permutationResults$m2m.p<0.05) &
                     (permutationResults$f2f.z >0 &
                        permutationResults$f2f.p<0.05),]
```

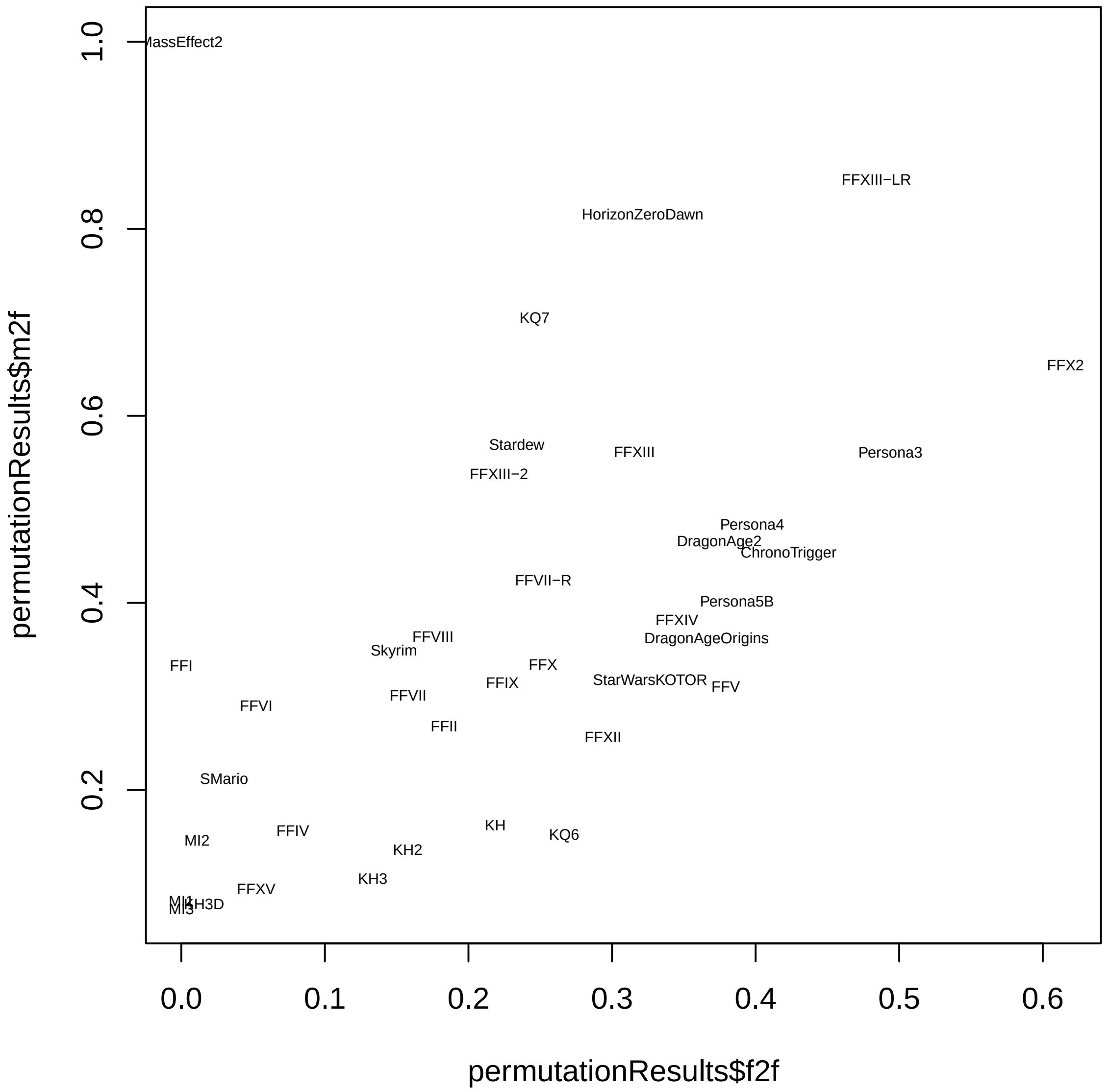
```
##                                  folder                                    series
## 3  ../data/DragonAge/DragonAgeOrigins_B                                Dragon Age
## 27       ../data/KingsQuest/KingsQuest6                              King's Quest
## 37  ../data/StarWarsKOTOR/StarWarsKOTOR Star Wars: Knights of the Old Republic
##                                     game          shortName       m2m
## 3                    Dragon Age: Origins DragonAgeOrigins_B 0.6392202
## 27                       King's Quest VI        KingsQuest6 0.8475798
## 37 Star Wars: Knights of the Old Republic      StarWarsKOTOR 0.6820782
##          f2m       m2f       f2f m2m.raw f2m.raw m2f.raw f2f.raw  m2m.mean
## 3  0.6341757 0.3607798 0.3658243    2787    1581    1573     912 0.6883674
## 27 0.7333333 0.1524202 0.2666667     823     110     148      40 0.9437190
## 37 0.6734463 0.3179218 0.3265537    1431     596     667     289 0.7128301
##         m2m.z m2m.p  f2m.mean      f2m.z f2m.p   m2f.mean     m2f.z m2f.p
## 3   -9.356370 0.001 0.6889318  -6.481471 0.001 0.31163261  9.356370 0.001
## 27 -30.054323 0.001 0.9443704 -10.561529 0.001 0.05628103 30.054323 0.001
## 37  -3.757067 0.001 0.7136756  -2.800403 0.004 0.28716987  3.757067 0.001
##      f2f.mean     f2f.z f2f.p diffExpEmp.m2m diffExpEmp.f2f       shortName2
## 3  0.31106823  6.481471 0.001    -0.04914720     0.05475608 DragonAgeOrigins
## 27 0.05562956 10.561529 0.001    -0.09613916     0.21103710              KQ6
## 37 0.28632442  2.800403 0.004    -0.03075196     0.04022926    StarWarsKOTOR
```

```
# Game with less f-f and more m-m than expected:
permutationResults[(permutationResults$m2m.z >0 &
                      permutationResults$m2m.p<0.05) &
                     (permutationResults$f2f.z <0 &
                        permutationResults$f2f.p<0.05),]
```

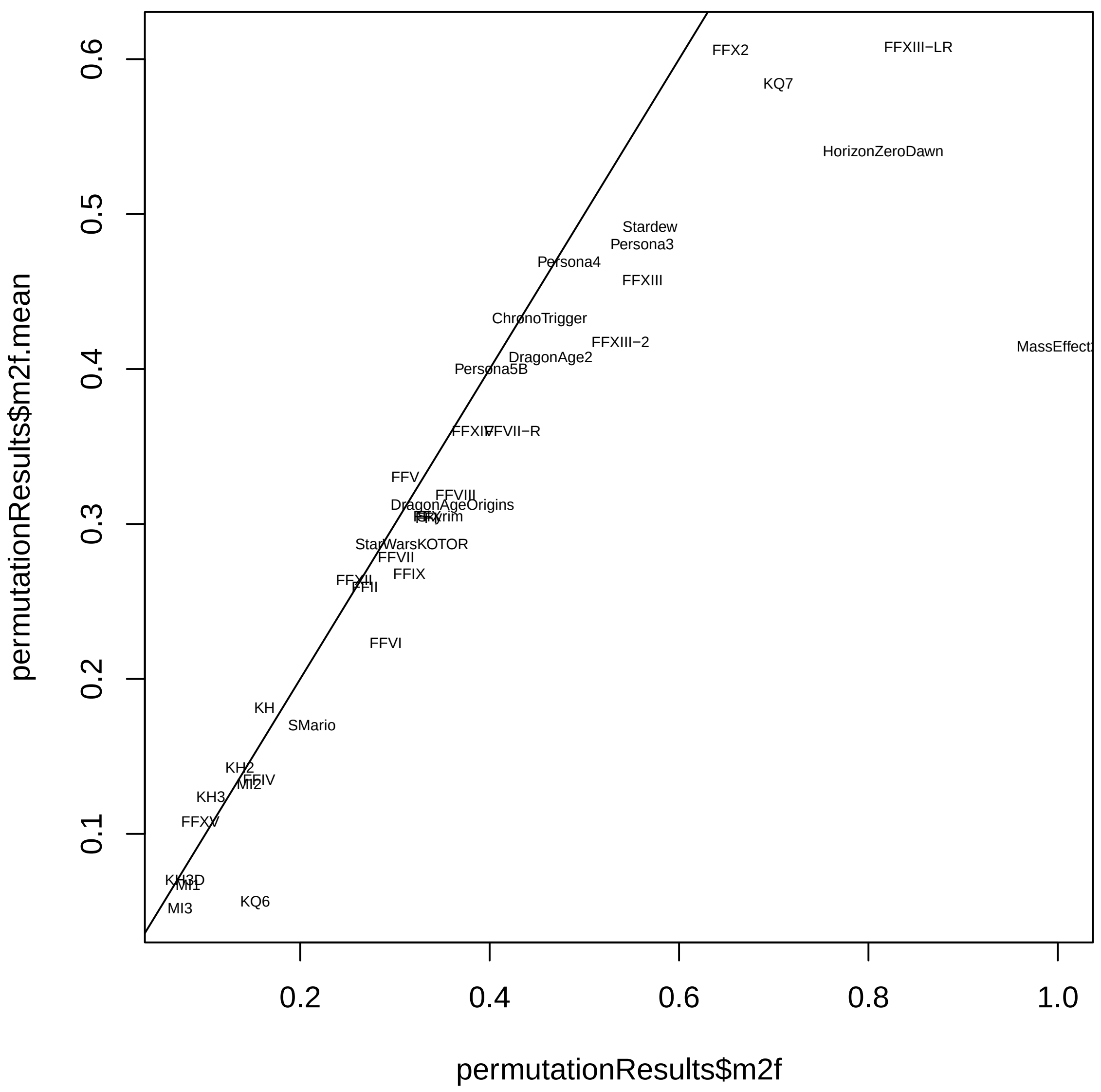
```
##                        folder        series              game shortName      m2m
## 21 ../data/FinalFantasy/FFXV Final Fantasy Final Fantasy XV      FFXV 0.905426
##          f2m        m2f        f2f m2m.raw f2m.raw m2f.raw f2f.raw  m2m.mean
## 21 0.9477612 0.09457398 0.05223881    4155     381     434      21 0.8919802
##       m2m.z m2m.p  f2m.mean    f2m.z f2m.p  m2f.mean     m2f.z m2f.p  f2f.mean
## 21 4.849541 0.001 0.8921863 4.37991 0.001 0.1080198 -4.849541 0.001 0.1078137
##       f2f.z f2f.p diffExpEmp.m2m diffExpEmp.f2f shortName2
## 21 -4.37991 0.001     0.01344582    -0.05557492       FFXV
```

The relationship between the different transition types is correlated, but there is some variation:

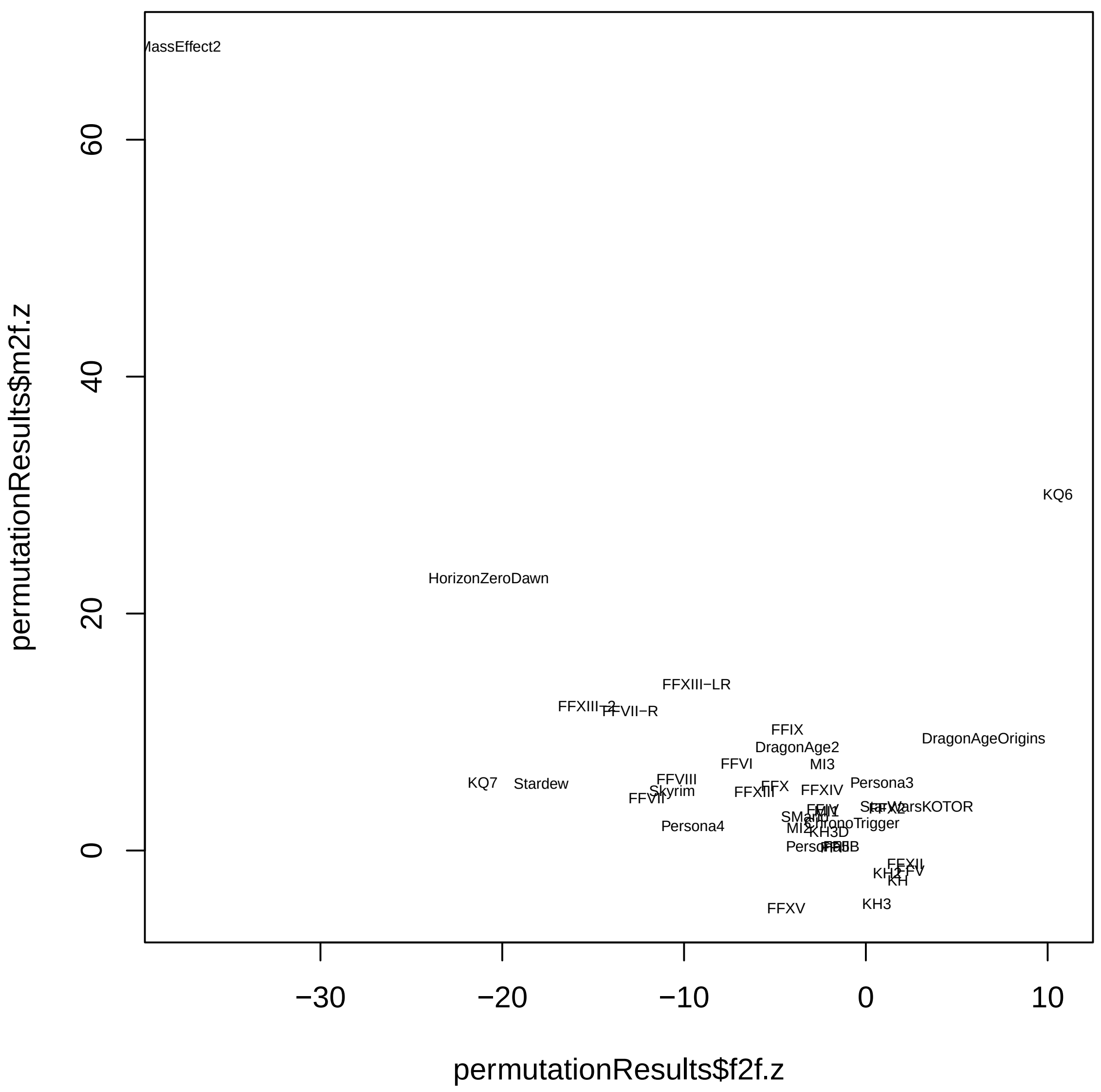
```
plot(permutationResults$f2f,permutationResults$m2f,col=NA)
text(permutationResults$f2f,permutationResults$m2f,permutationResults$shortName2,cex=0.5)
```



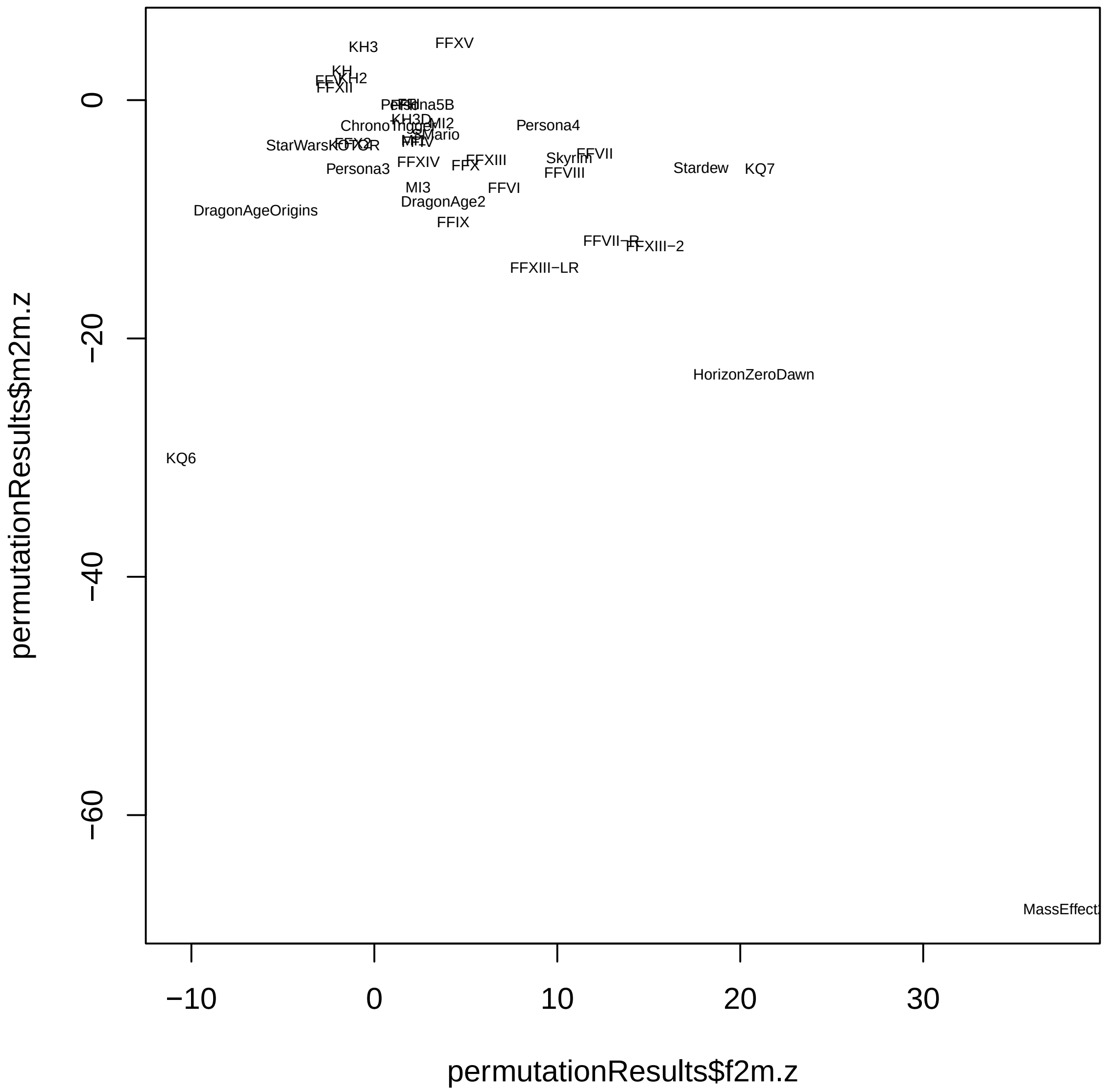

```
plot(permutationResults$m2f,permutationResults$m2f.mean,col=NA)
text(permutationResults$m2f,permutationResults$m2f.mean,permutationResults$shortName2,cex=0.5)
abline(0,1)
```



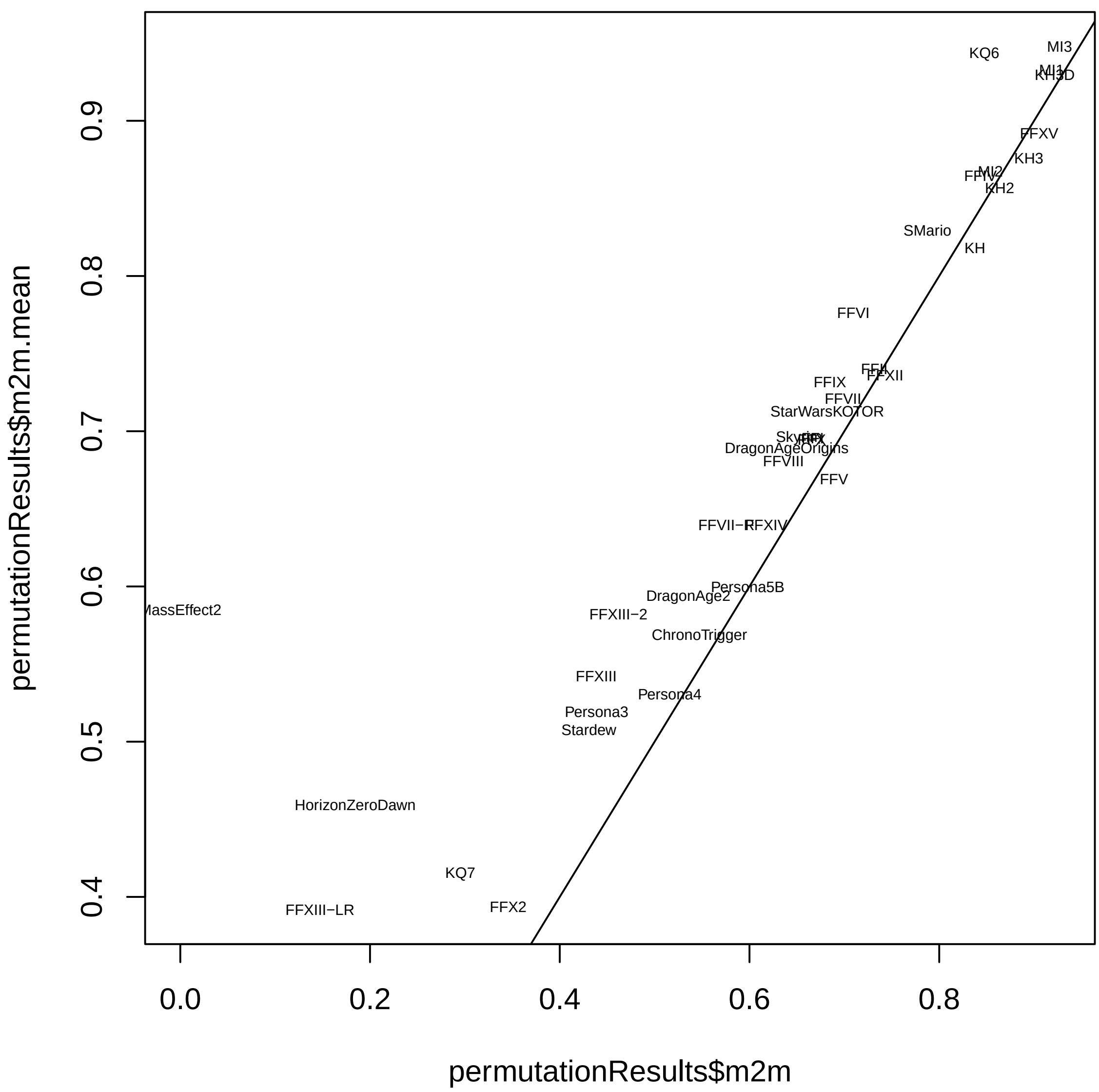

```
plot(permutationResults$f2f.z,permutationResults$m2f.z,col=NA)
text(permutationResults$f2f.z,permutationResults$m2f.z,permutationResults$shortName2,cex=0.5)
```




```
plot(permutationResults$f2m.z,permutationResults$m2m.z,col=NA)
text(permutationResults$f2m.z,permutationResults$m2m.z,permutationResults$shortName2,cex=0.5)
```




```
plot(permutationResults$m2m,permutationResults$m2m.mean,col=NA)
text(permutationResults$m2m,permutationResults$m2m.mean,permutationResults$shortName2,cex=0.5)
abline(0,1)
```



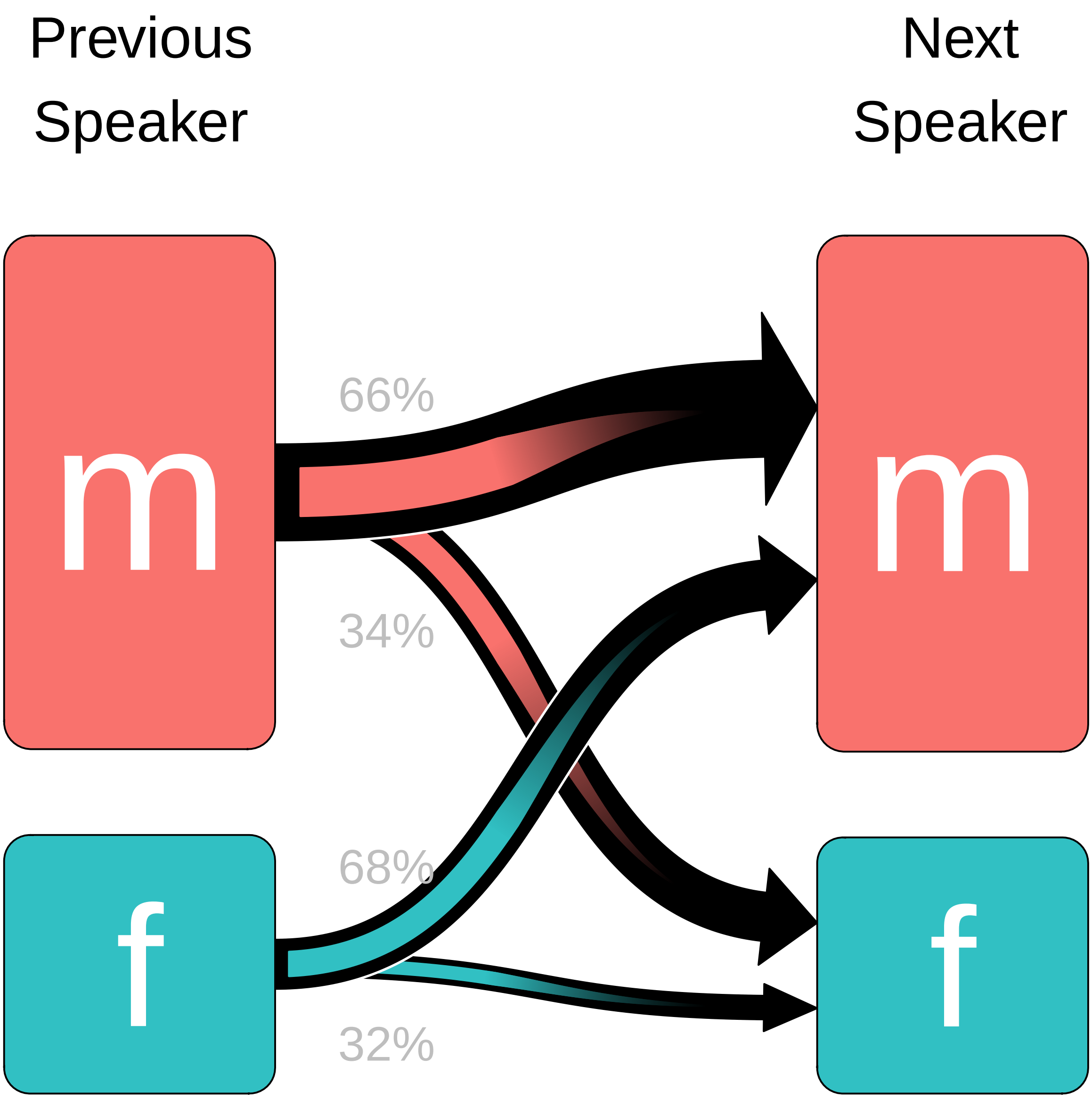
Fancy class for plotting transitions:

```
transitions <- trueTransitions.AllGames.Raw %>%
  getRefClass("Transition")$new(
      label=c("Previous\nSpeaker", "Next\nSpeaker"),
      skip_shadows = TRUE,
      min_lwd = unit(0.1, "mm"), max_lwd = unit(14, "mm"),
      box_label_cex = 2,
      fill_clr = list(c("#f9726d","#31c0c3"),c("#f9726d","#31c0c3")))
```

Plots for all games

Visualisation of overall transitions for all games:

```
transitions$render()
grid::grid.text(label=paste0(round(100*trueTransitions.AllGames),"%"),
                x=c(0.35,0.35,0.35,0.35),
                y=c(0.6,0.2, 0.4,0.05),
                gp=gpar(fontsize=20, col="grey"))
```

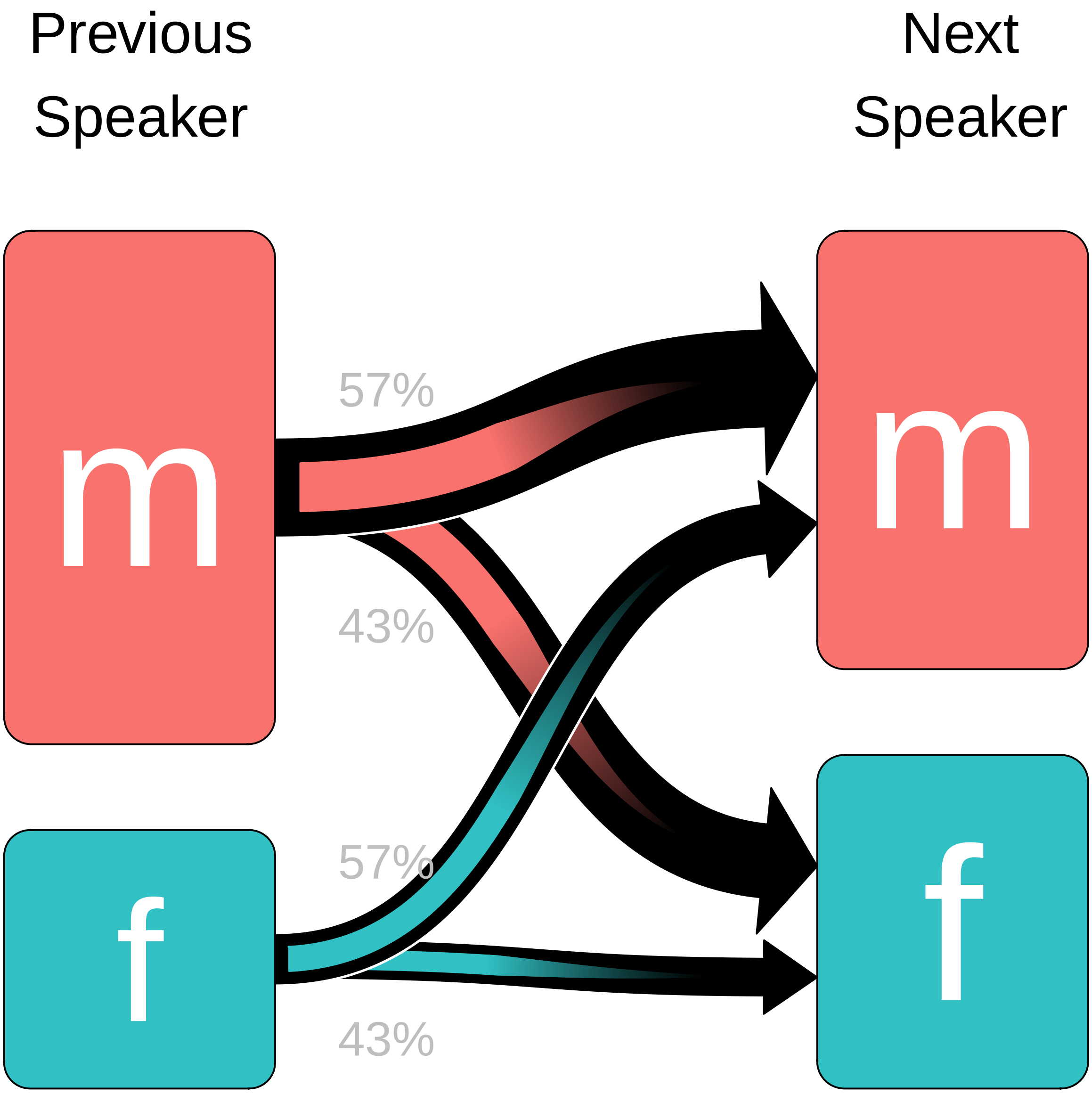



```
pdf("../results/graphs/transitions/Transitions.pdf")
transitions$render()
grid::grid.text(label=paste0(round(100*trueTransitions.AllGames),"%"),
                x=c(0.35,0.35,0.35,0.35),
                y=c(0.6,0.2, 0.4,0.05),
                gp=gpar(fontsize=20, col="grey"))
dev.off()
```

```
## pdf
##   2
```

Expected proportions under permutation for all games:

```
expectedProp = matrix(c(permutedTransitionStats.AllGames['m2m.mean'], permutedTransitionStats.AllGam
permutedTransitionStats.AllGames['f2m.mean'], permutedTransitionStats.AllGames['f2f.mean']),nrow=2,b
rownames(expectedProp) = c("m","f")
colnames(expectedProp) = c("m","f")

totalLinesPerGender = rowSums(trueTransitions.AllGames.Raw)
expectedRaw = round(expectedProp * totalLinesPerGender)

expectedTrans <- expectedRaw %>%
  getRefClass("Transition")$new(
    label=c("Previous\nSpeaker", "Next\nSpeaker"),
    skip_shadows = TRUE,
    min_lwd = unit(0.01, "mm"), max_lwd = unit(14, "mm"),
    box_label_cex = 2,
    fill_clr = list(c("#f9726d","#31c0c3"),c("#f9726d","#31c0c3")))

expectedTrans$render()
grid::grid.text(label=paste0(round(100*expectedProp),"%"),
                x=c(0.35,0.35,0.35,0.35),
                y=c(0.6,0.2, 0.4,0.05),
                gp=gpar(fontsize=20, col="grey"))
```




```
pdf("../results/graphs/transitions/Transitions_ExpectedAllGames.pdf")
expectedTrans$render()
grid::grid.text(label=paste0(round(100*expectedProp),"%"),
                x=c(0.35,0.35,0.35,0.35),
                y=c(0.6,0.2, 0.4,0.05),
                gp=gpar(fontsize=20, col="grey"))
dev.off()
```

```
## pdf
##   2
```

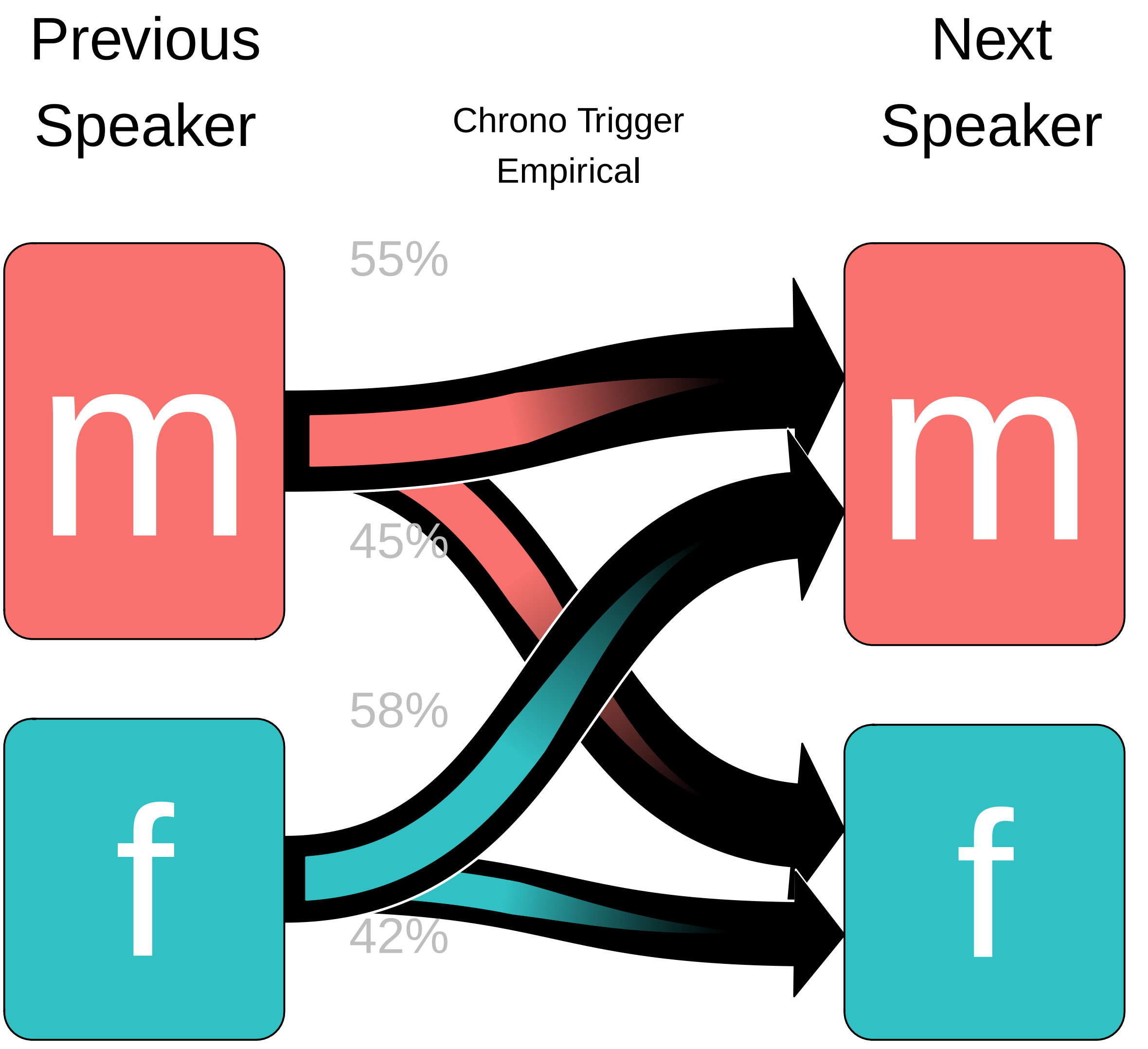
Plots for individual games

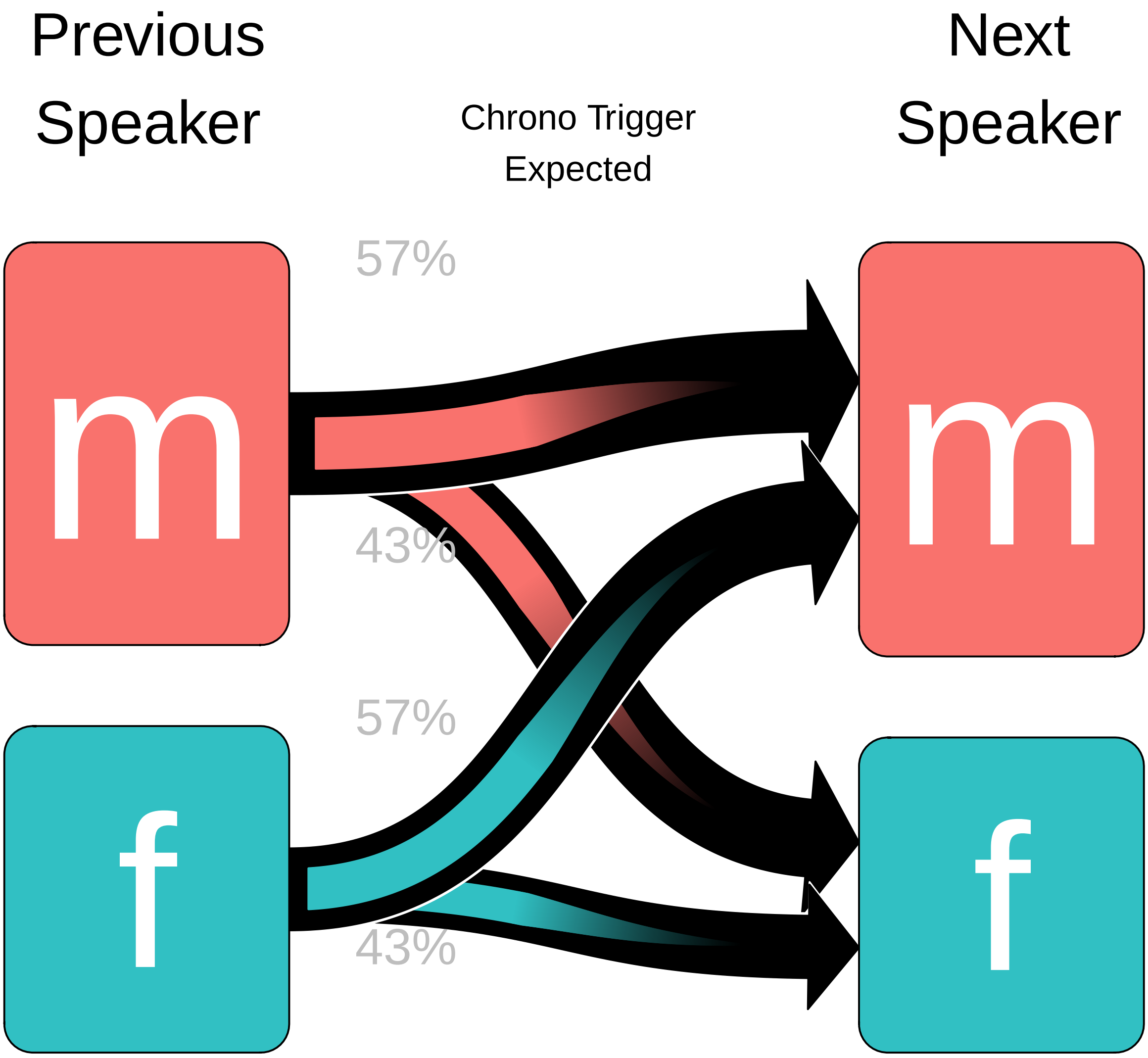
Functions for creating plots from a single game:

```
plotTransition = function(transProp, transRaw, outFileName,ys,title){
  transitions.ff10 <- transRaw %>%
    getRefClass("Transition")$new(
      label=c("Previous\nSpeaker", "Next\nSpeaker"),
      skip_shadows = TRUE,
      min_lwd = unit(0.01, "mm"), max_lwd = unit(14, "mm"),
      box_label_cex = 2,
      fill_clr = list(c("#f9726d","#31c0c3"),c("#f9726d","#31c0c3")))

  # Plot to console
  transitions.ff10$render()
  grid::grid.text(label=paste0(round(100*transProp),"%"),
                  x=c(0.35,0.35,0.35,0.35),
                  y=ys,
                  gp=gpar(fontsize=20, col="grey"))
  grid::grid.text(label=title,x=0.5,y=0.8, gp=gpar(fontsize=14))
  # Plot to pdf
  pdf(outFileName)
  transitions.ff10$render()
  grid::grid.text(label=paste0(round(100*transProp),"%"),
                  x=c(0.35,0.35,0.35,0.35),
                  y=ys,
                  gp=gpar(fontsize=20, col="grey"))
  grid::grid.text(label=title,x=0.5,y=0.8, gp=gpar(fontsize=14))
  dev.off()

}

makeTransitionGraphForOneGame = function(game,outFileName,ys=c(0.7,0.3, 0.45,0.1),title=""){
  # Empirical plot
  dx = permutationResults[permutationResults$game==game,]
  ff10t = matrix(c(dx$m2m,dx$m2f,dx$f2m,dx$f2f),ncol=2,byrow = T)
  rownames(ff10t) = c("m","f")
  colnames(ff10t) = c("m","f")
  ff10t.raw = matrix(c(dx$m2m.raw,dx$m2f.raw,dx$f2m.raw,dx$f2f.raw),ncol=2,byrow = T)
  rownames(ff10t.raw) = c("m","f")
  colnames(ff10t.raw) = c("m","f")


  plotTransition(ff10t,ff10t.raw,outFileName,ys,
                 title = paste(title,"Empirical",sep="\n"))

  # Expected plot
  expectedProp = matrix(c(dx[['m2m.mean']], dx[['m2f.mean']],
                          dx[['f2m.mean']], dx[['f2f.mean']]),nrow=2,byrow = T)
  rownames(expectedProp) = c("m","f")
  colnames(expectedProp) = c("m","f")

  totalLinesPerGender = rowSums(ff10t.raw)
  expectedRaw = round(expectedProp * totalLinesPerGender)
  outFileName2 = gsub("\\.pdf","_Expected.pdf",outFileName)
  plotTransition(expectedProp,expectedRaw,outFileName2,ys,
                 title = paste(title,"Expected",sep="\n"))

}
```

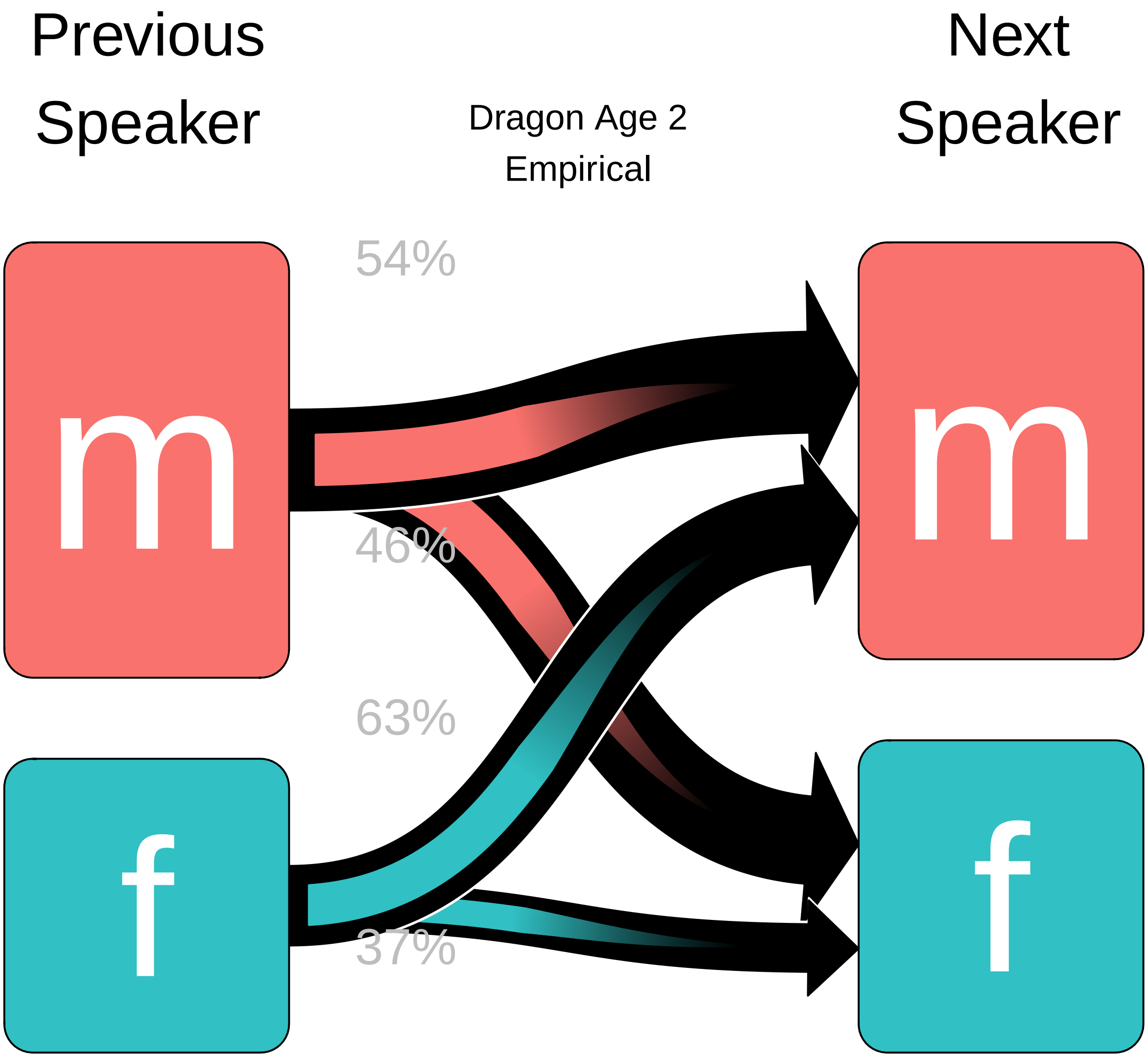
Apply to each game separately:

```
for(folder in unique(permutationResults$folder)){
  gameName = permutationResults[permutationResults$folder==folder,]$game[1]
  #print(gameName)
  gameName1 = gsub("[:']","",gameName)
  gameName2 = gsub(" +","_",gameName1)
  pdfFile = paste0("../results/graphs/transitions/Transitions_",gameName2,".pdf")
  if(gameName == "Final Fantasy XV"){
   makeTransitionGraphForOneGame(gameName, pdfFile,ys=c(0.55,0.12, 0.35,0.02),"FF XV")
  } else{
    makeTransitionGraphForOneGame(gameName, pdfFile,title=gameName1)
  }
}
```

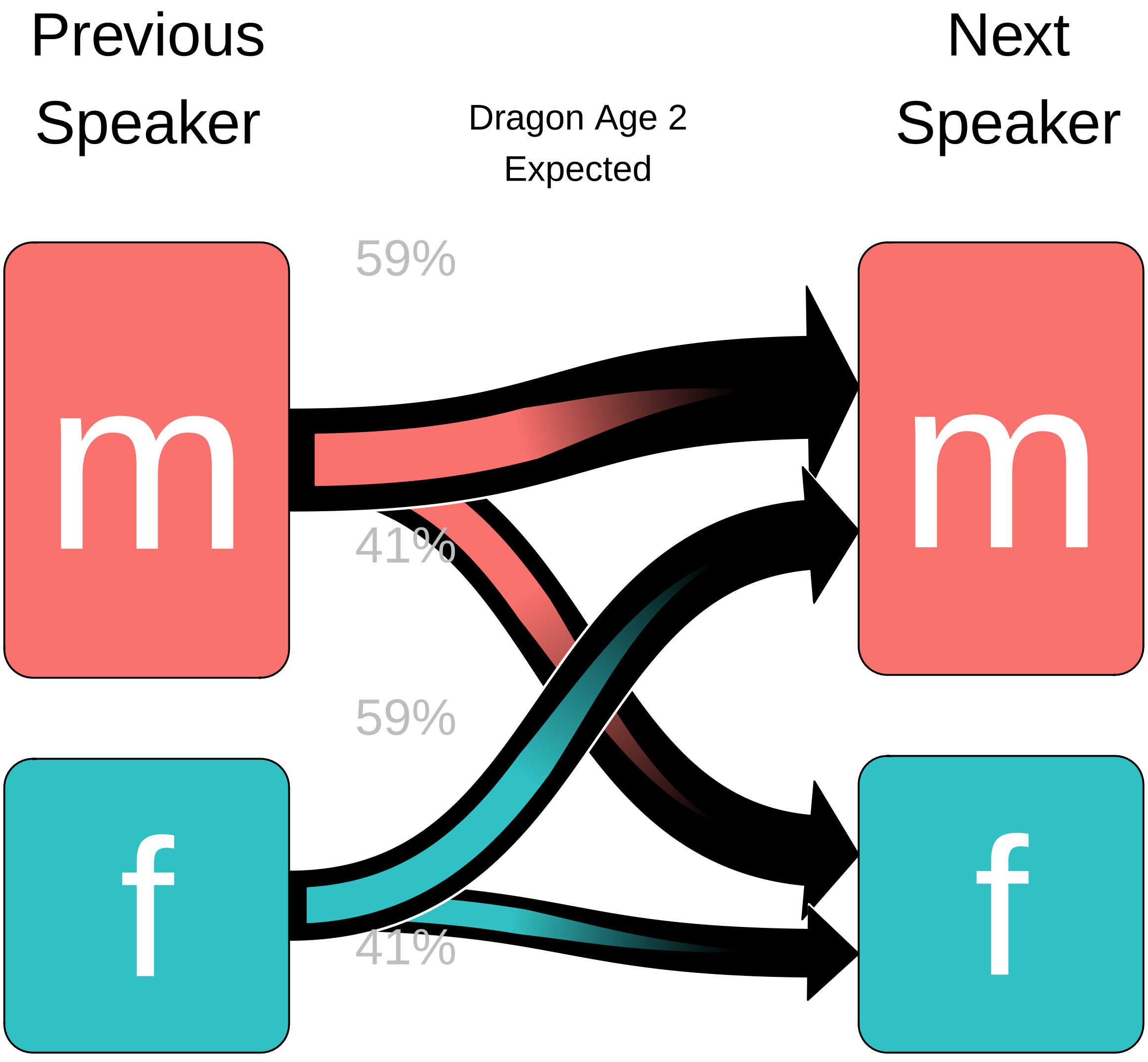


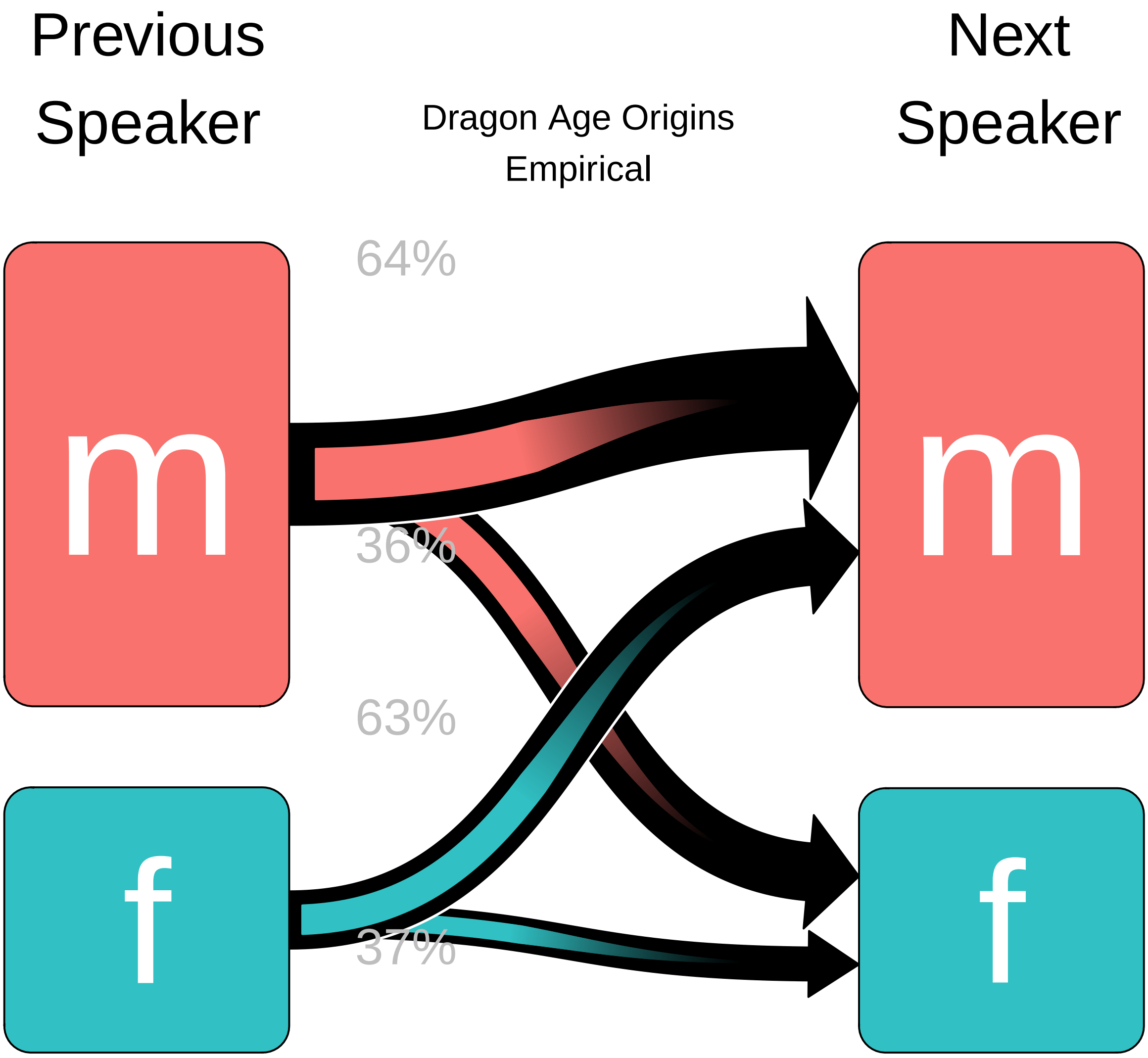


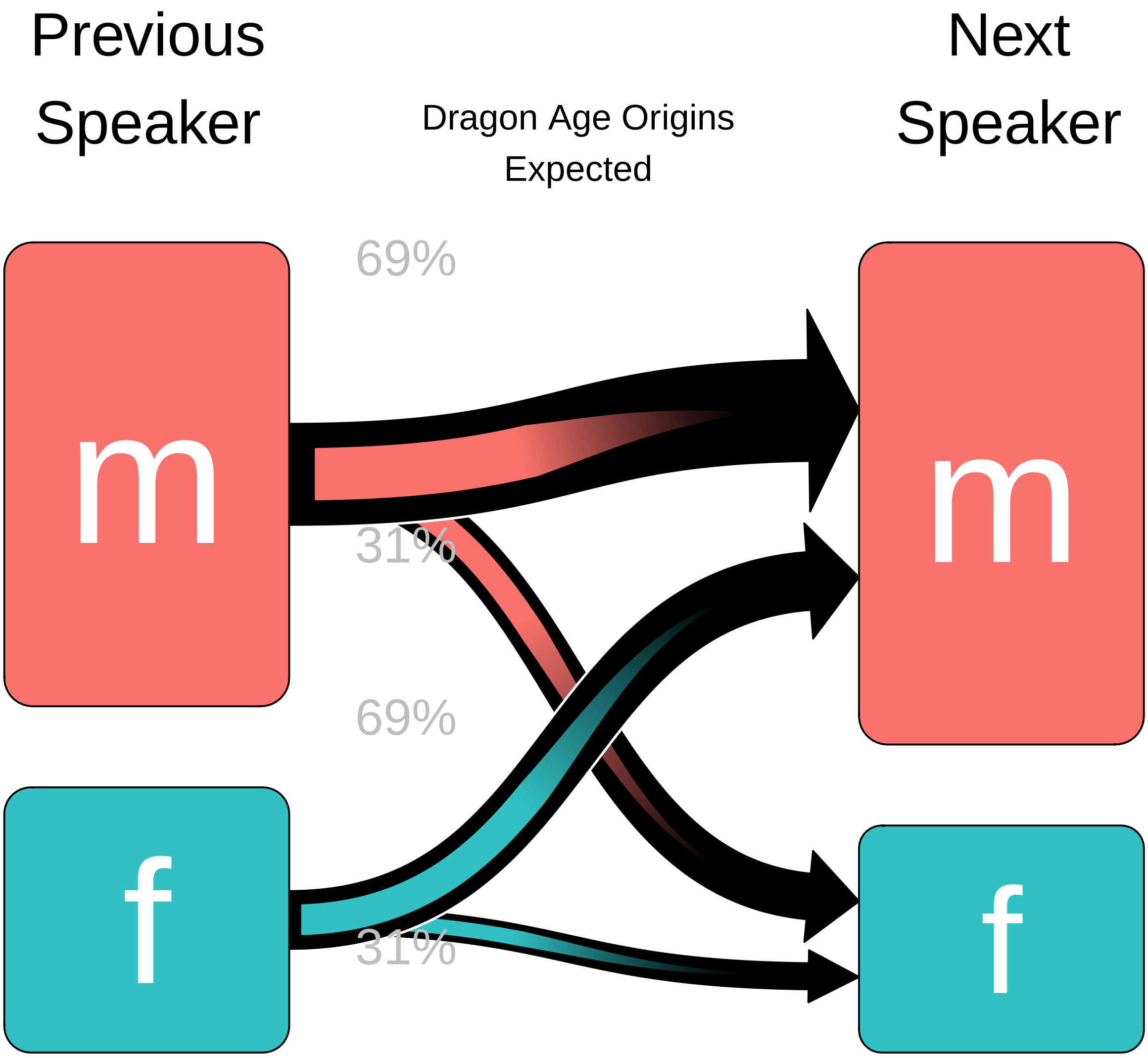
Previous Speaker
Chrono Trigger Expected
Next Speaker
m
f
57%
43%
57%
43%
m
f

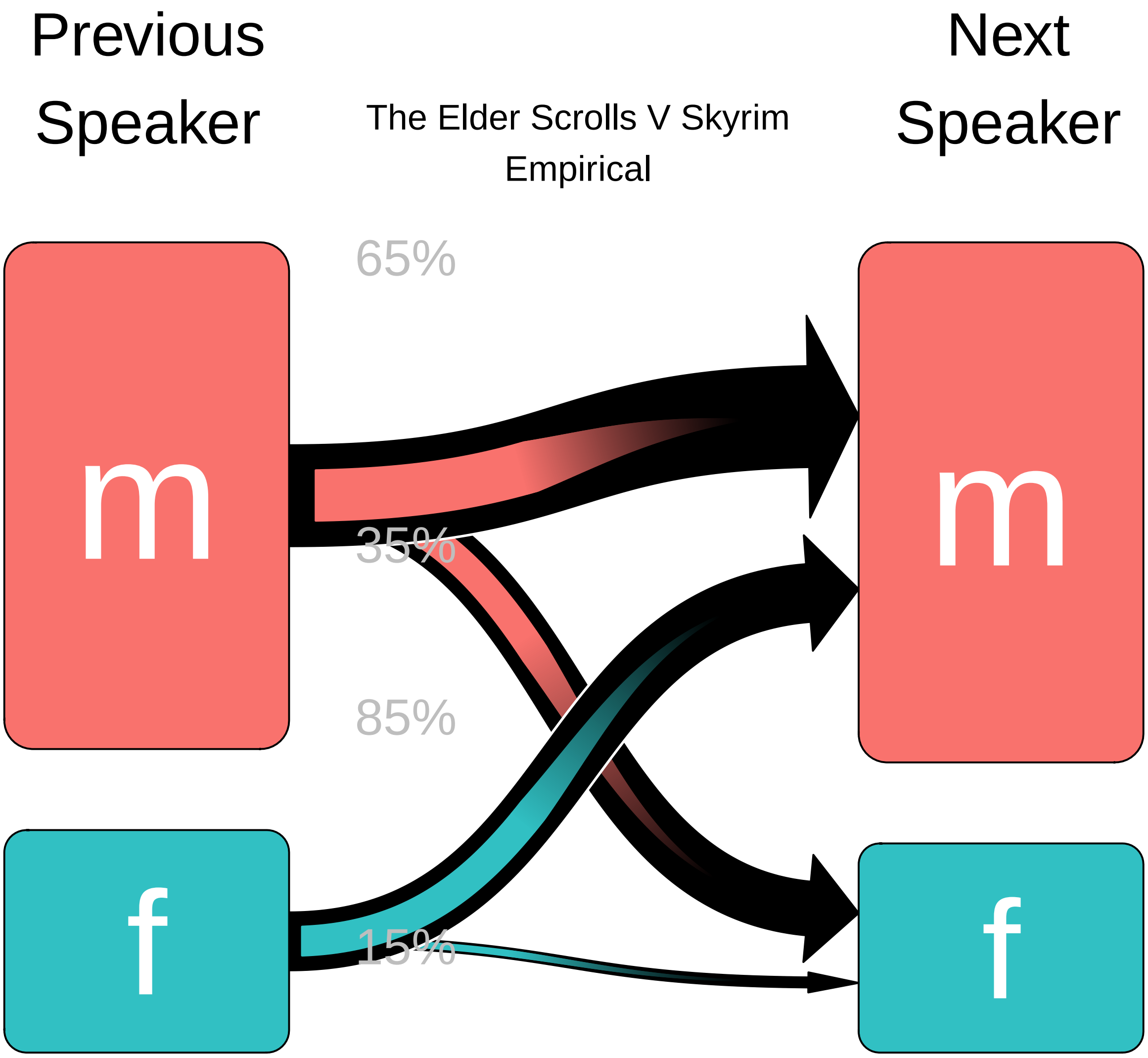


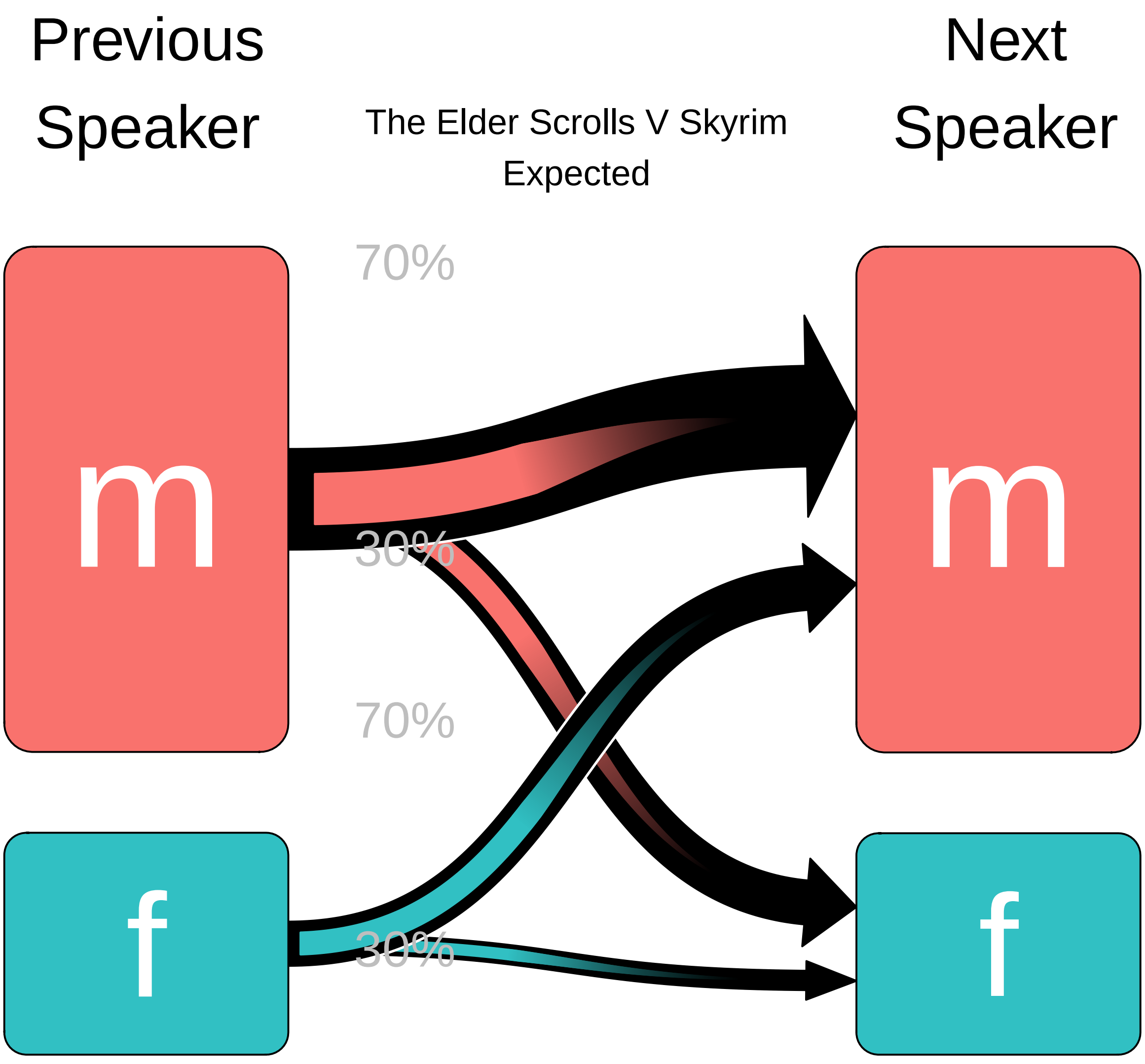
Previous
Speaker
Dragon Age 2
Empirical
Next
Speaker
54%
46%
63%
37%
m
f
m
f

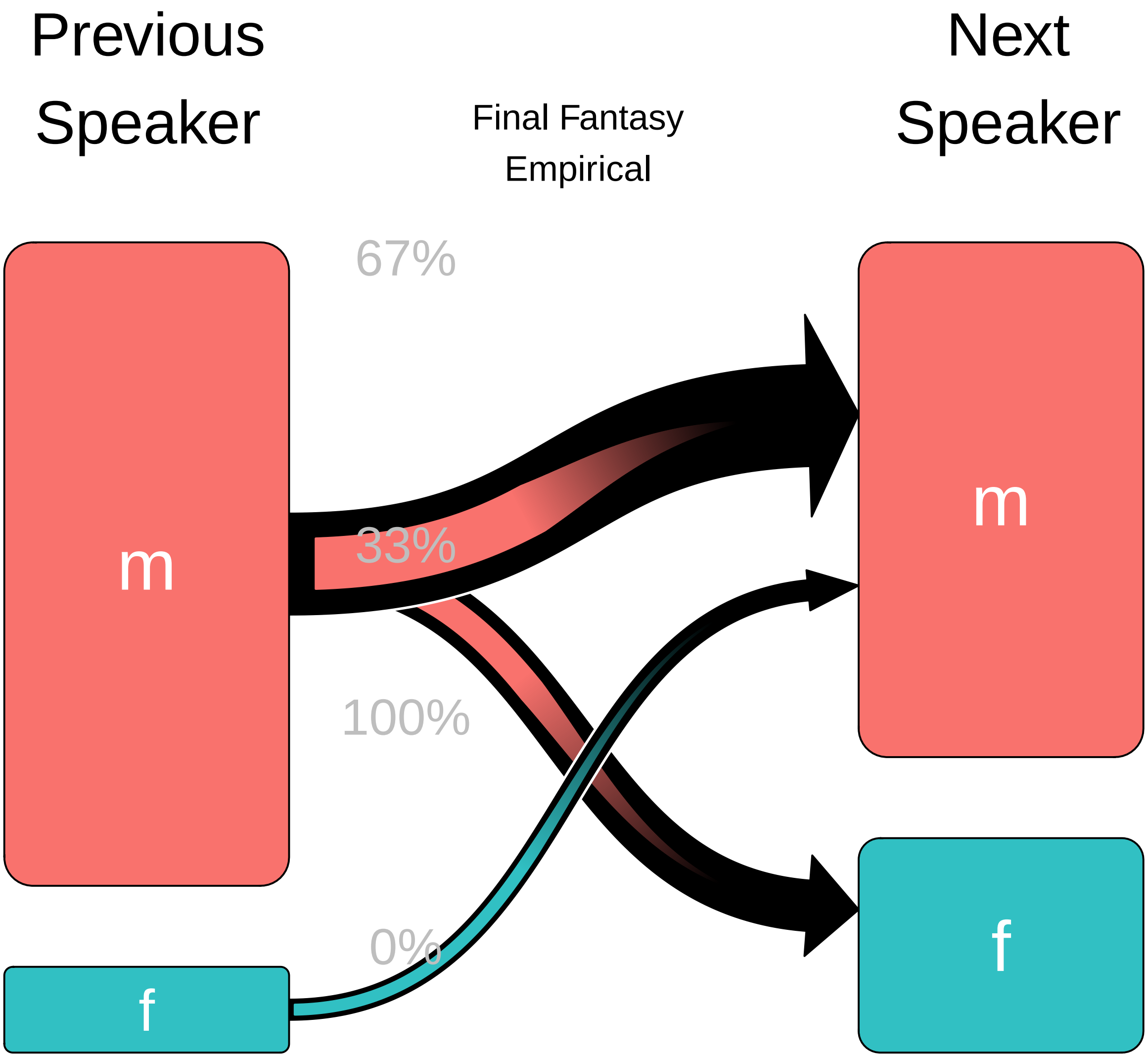

Previous
Speaker
Dragon Age 2
Expected
Next
Speaker
59%
41%
59%
41%
m
f
m
f


Previous
Speaker
Dragon Age Origins
Empirical
Next
Speaker
64%
m
m
36%
63%
f
f
37%

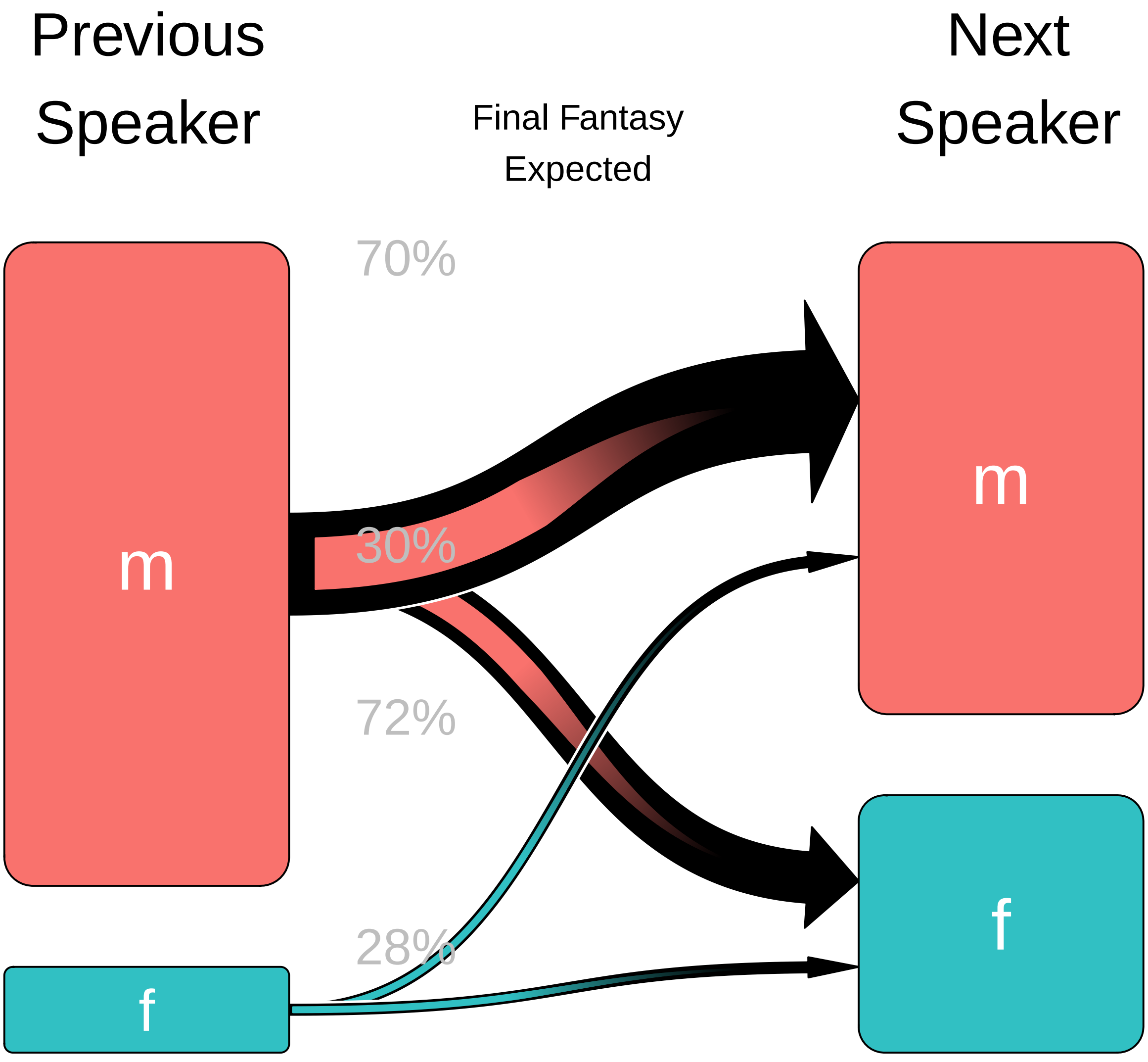

Previous
Speaker
Dragon Age Origins
Expected
Next
Speaker
69%
31%
69%
31%
m
f
m
f

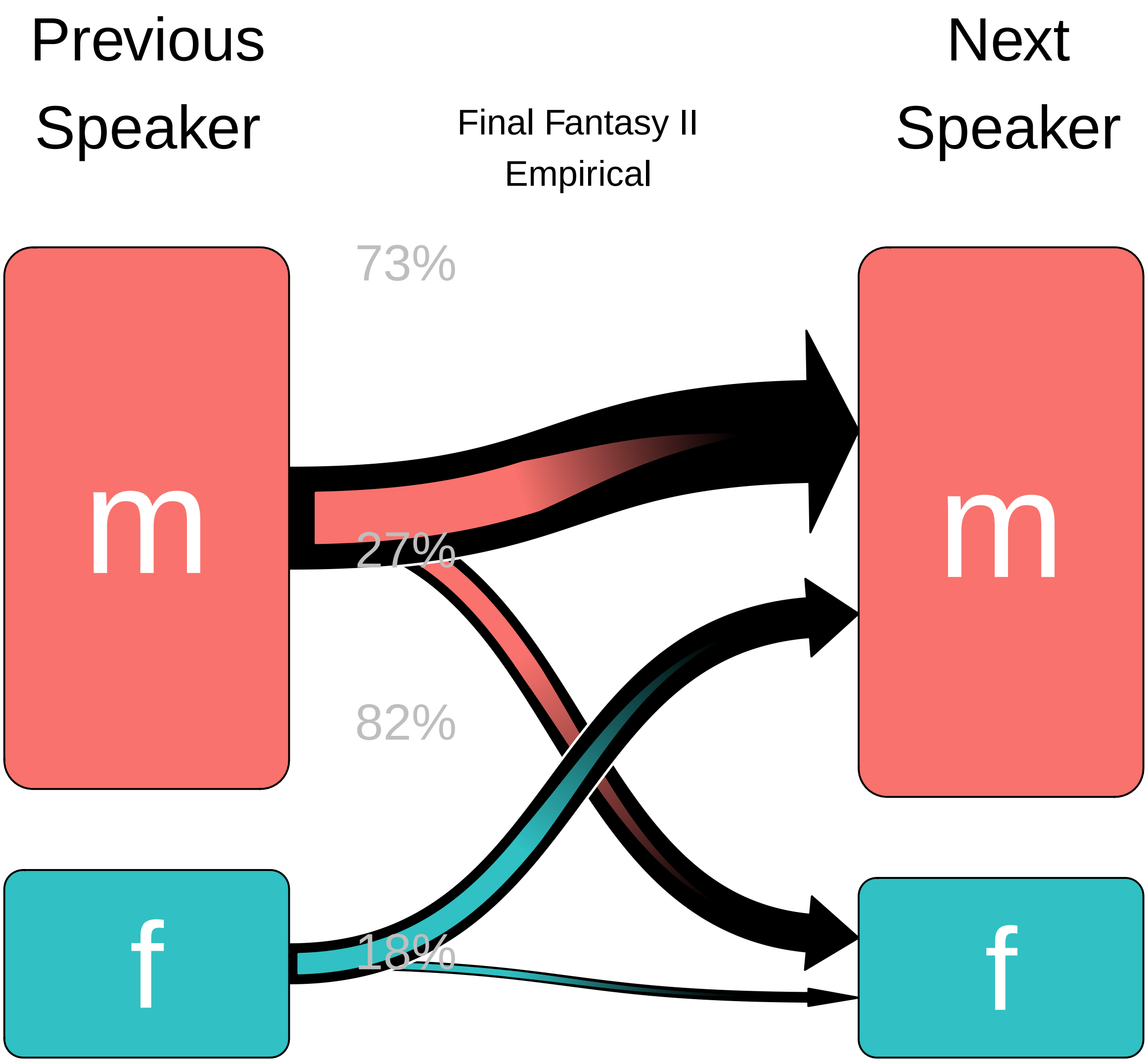

Previous
Speaker
The Elder Scrolls V Skyrim
Empirical
Next
Speaker
65%
m
35%
85%
m
f
15%
f

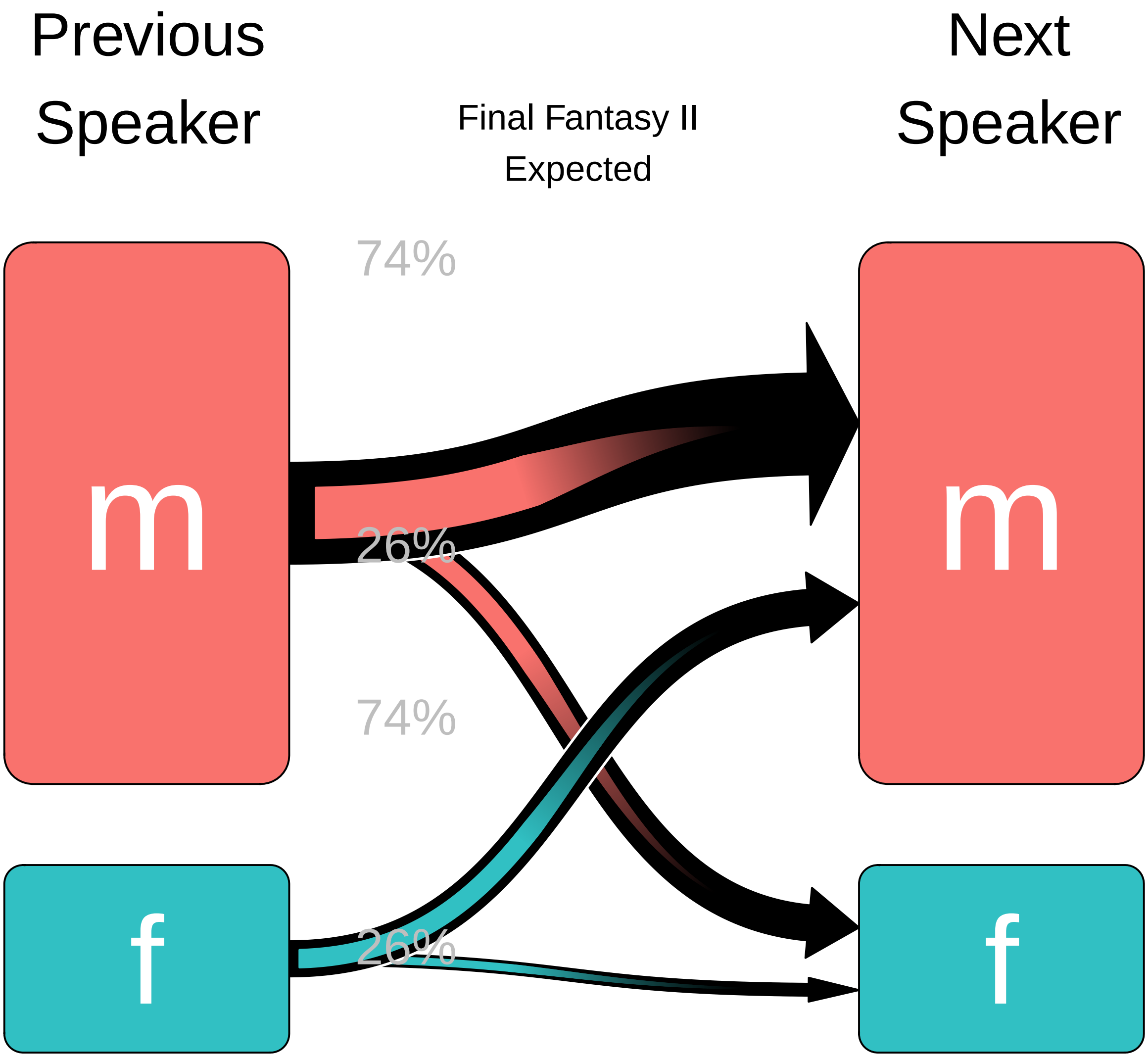

Previous Speaker
The Elder Scrolls V Skyrim
Expected
Next Speaker
70%
m
30%
70%
f
30%
m
f

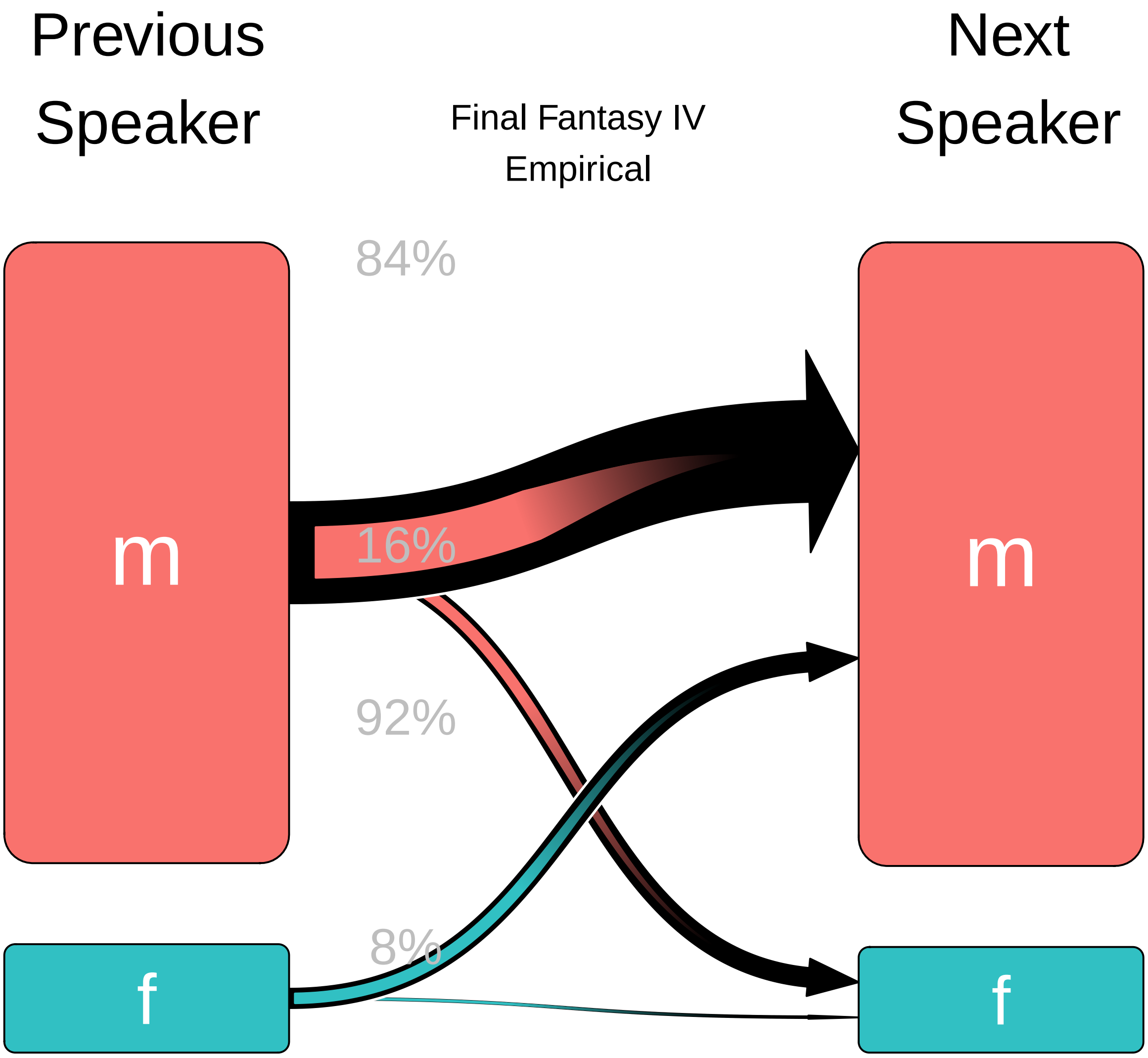


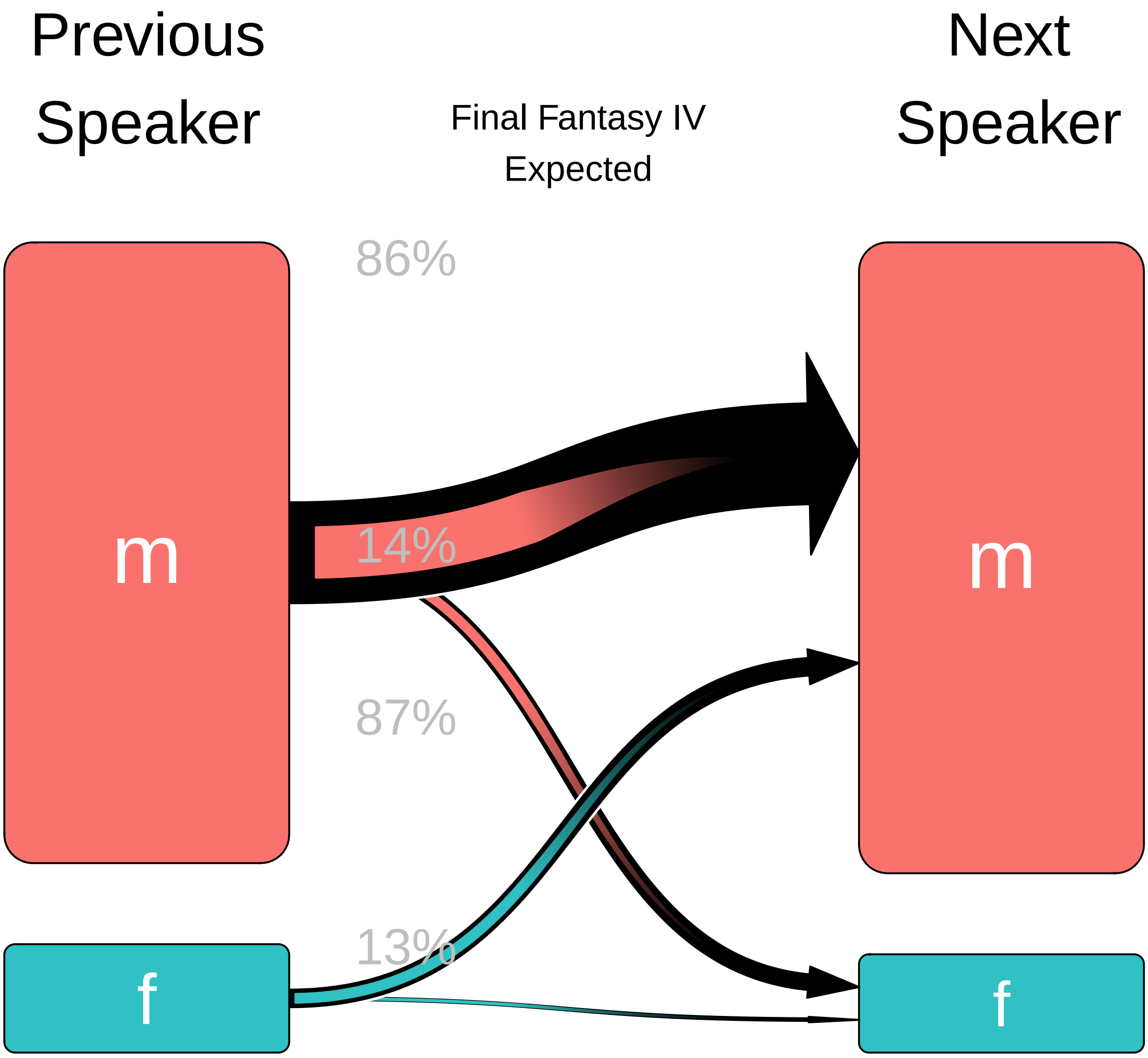
Previous
Speaker
Final Fantasy
Empirical
Next
Speaker
67%
m
33%
m
100%
0%
f
f

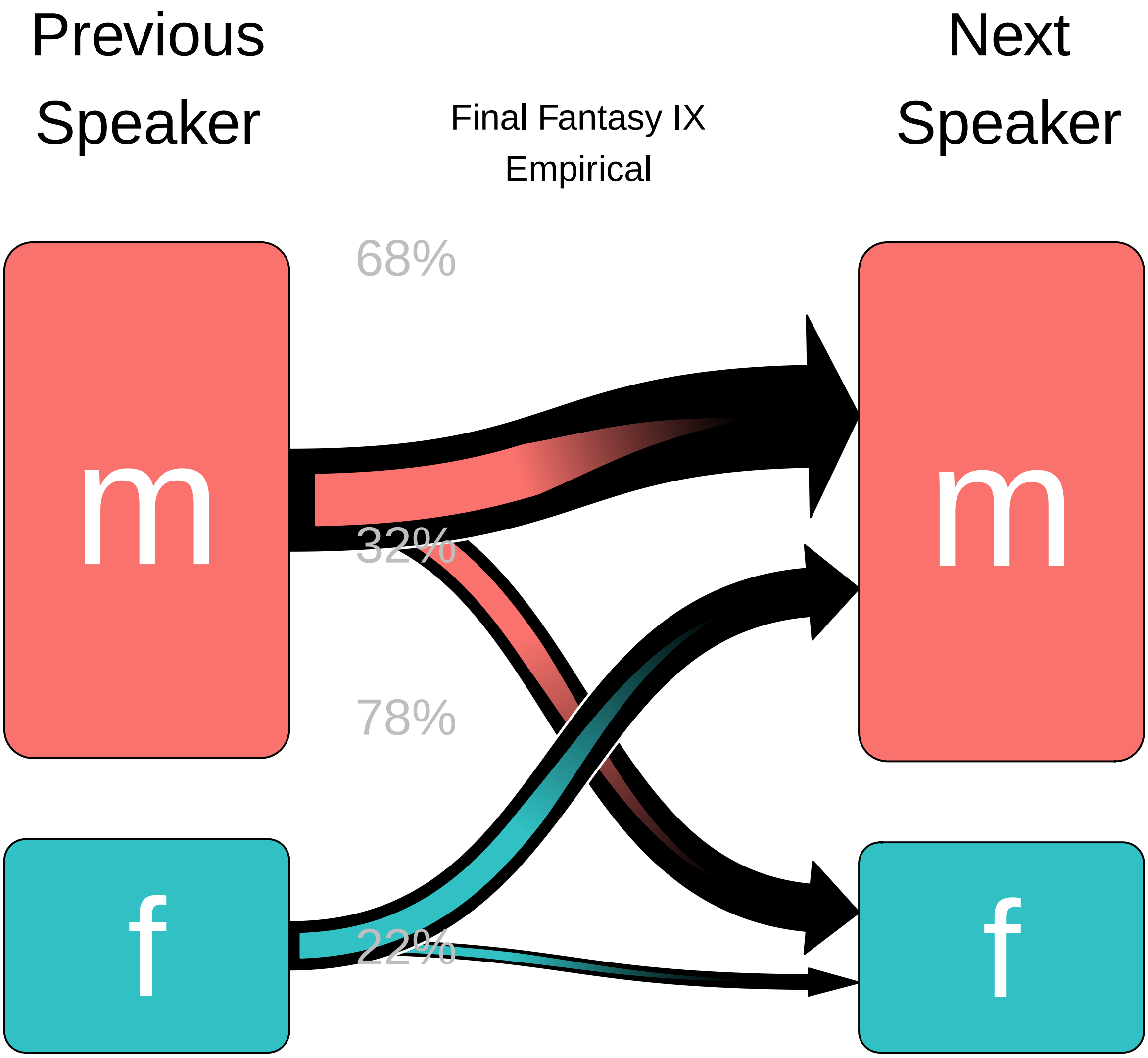

Previous
Speaker
Final Fantasy
Expected
Next
Speaker
70%
m
30%
72%
28%
f
m
f


Previous
Speaker
Final Fantasy II
Empirical
Next
Speaker
73%
m
27%
82%
m
f
18%
f

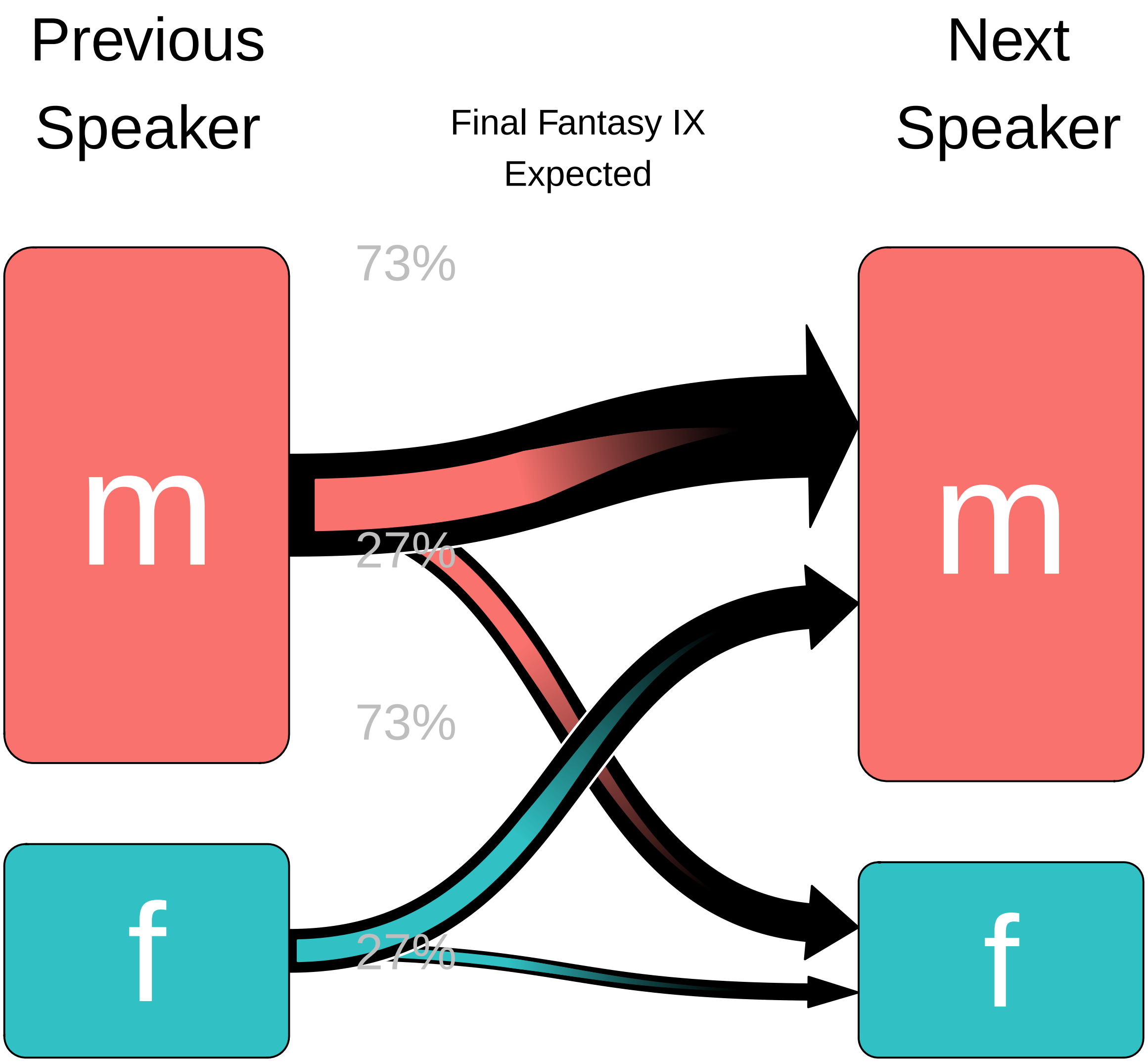

Previous
Speaker
Final Fantasy II
Expected
Next
Speaker
74%
26%
74%
26%
m
m
f
f

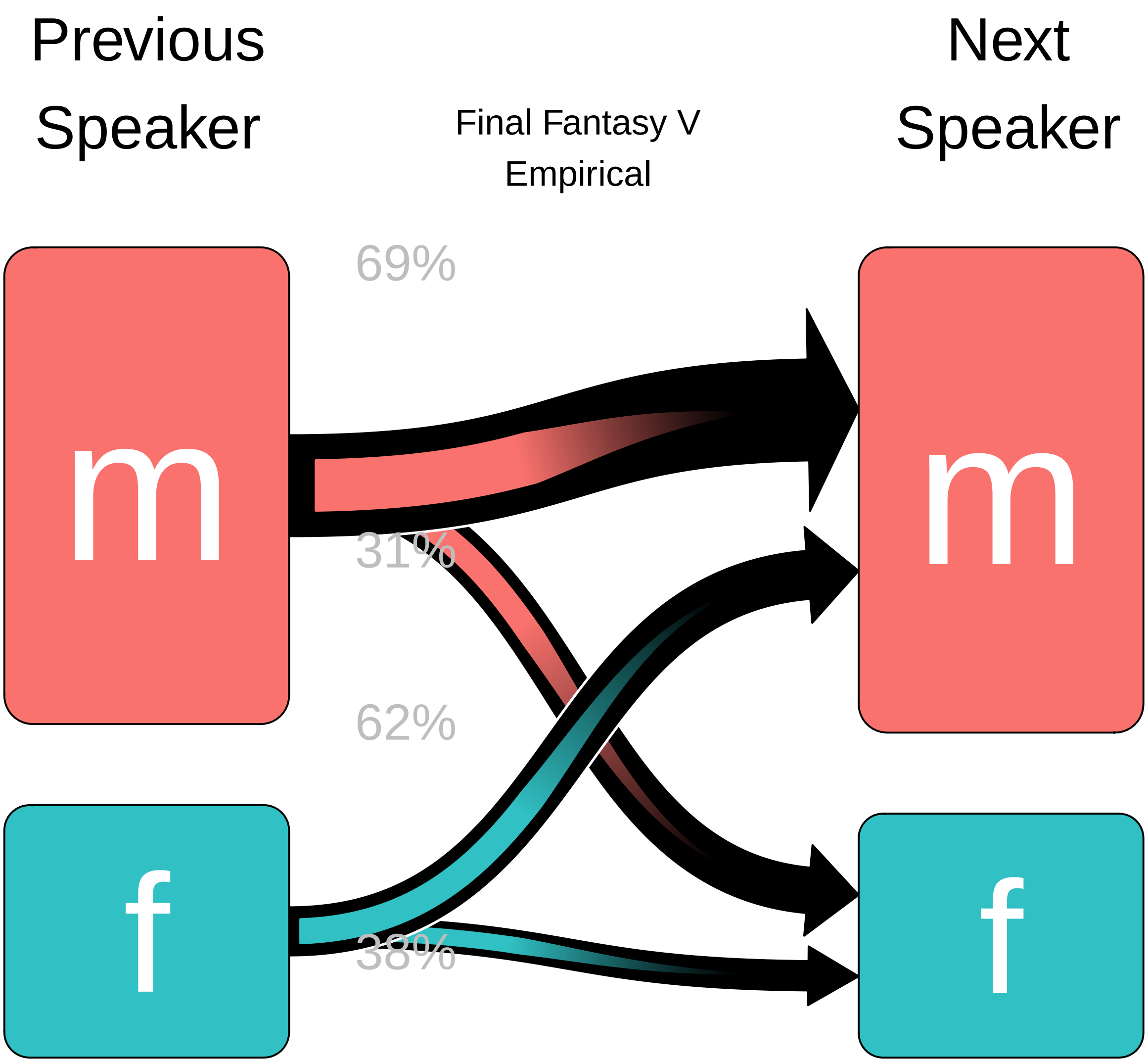

Previous
Speaker
Final Fantasy IV
Empirical
Next
Speaker
84%
m
16%
92%
8%
m
f
f

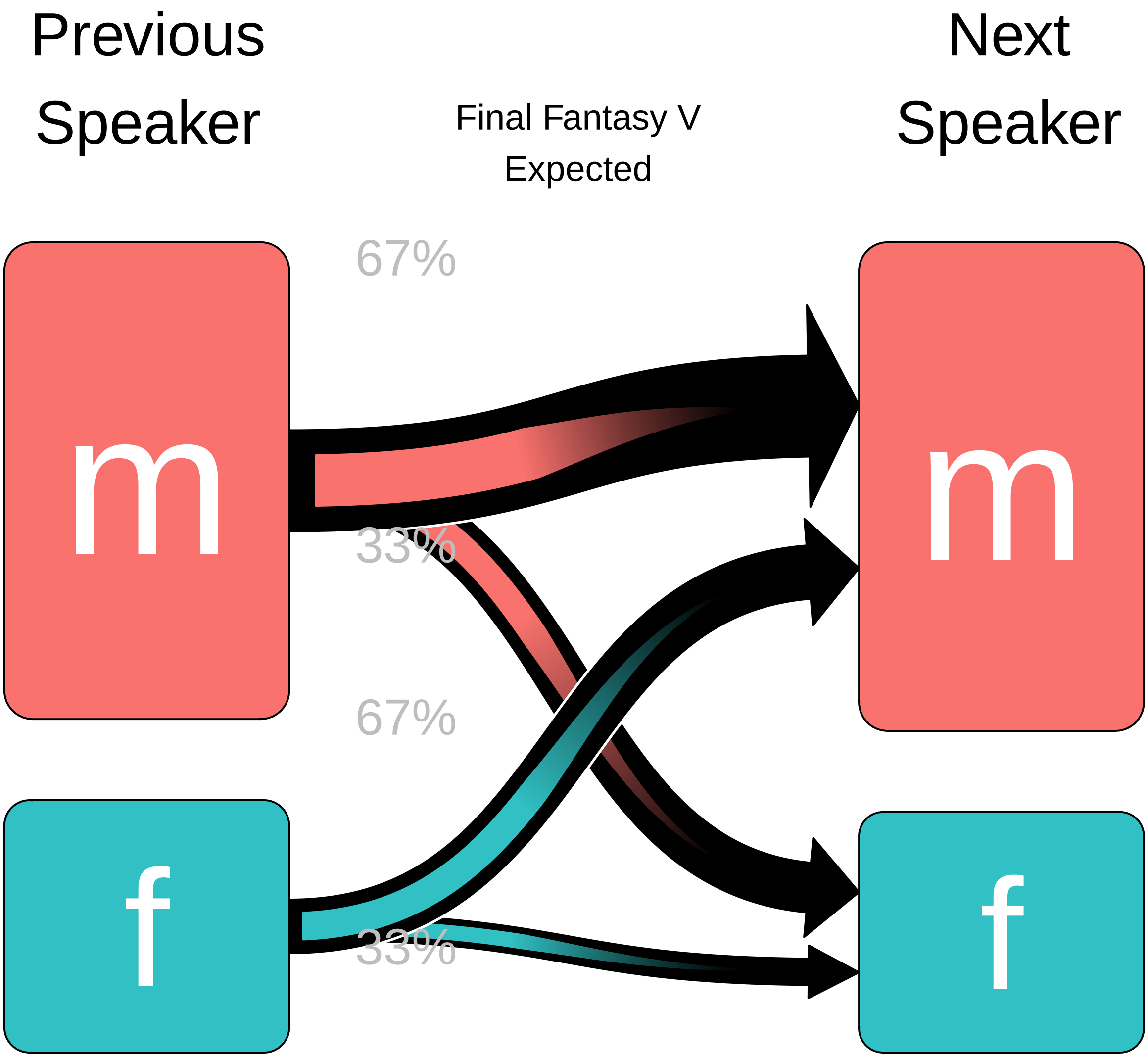

Previous
Speaker
Final Fantasy IV
Expected
Next
Speaker
86%
m
14%
87%
m
13%
f
f

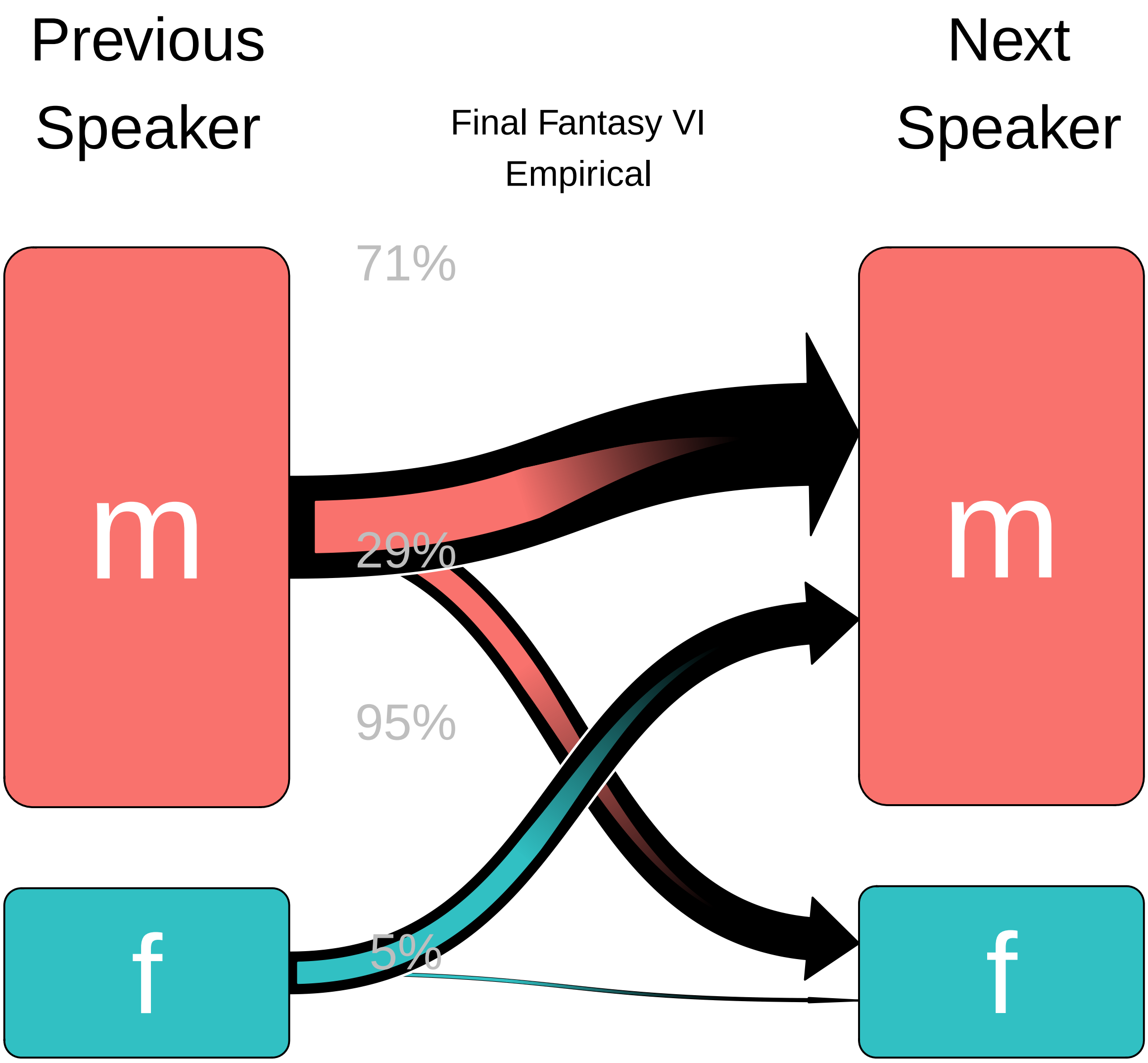


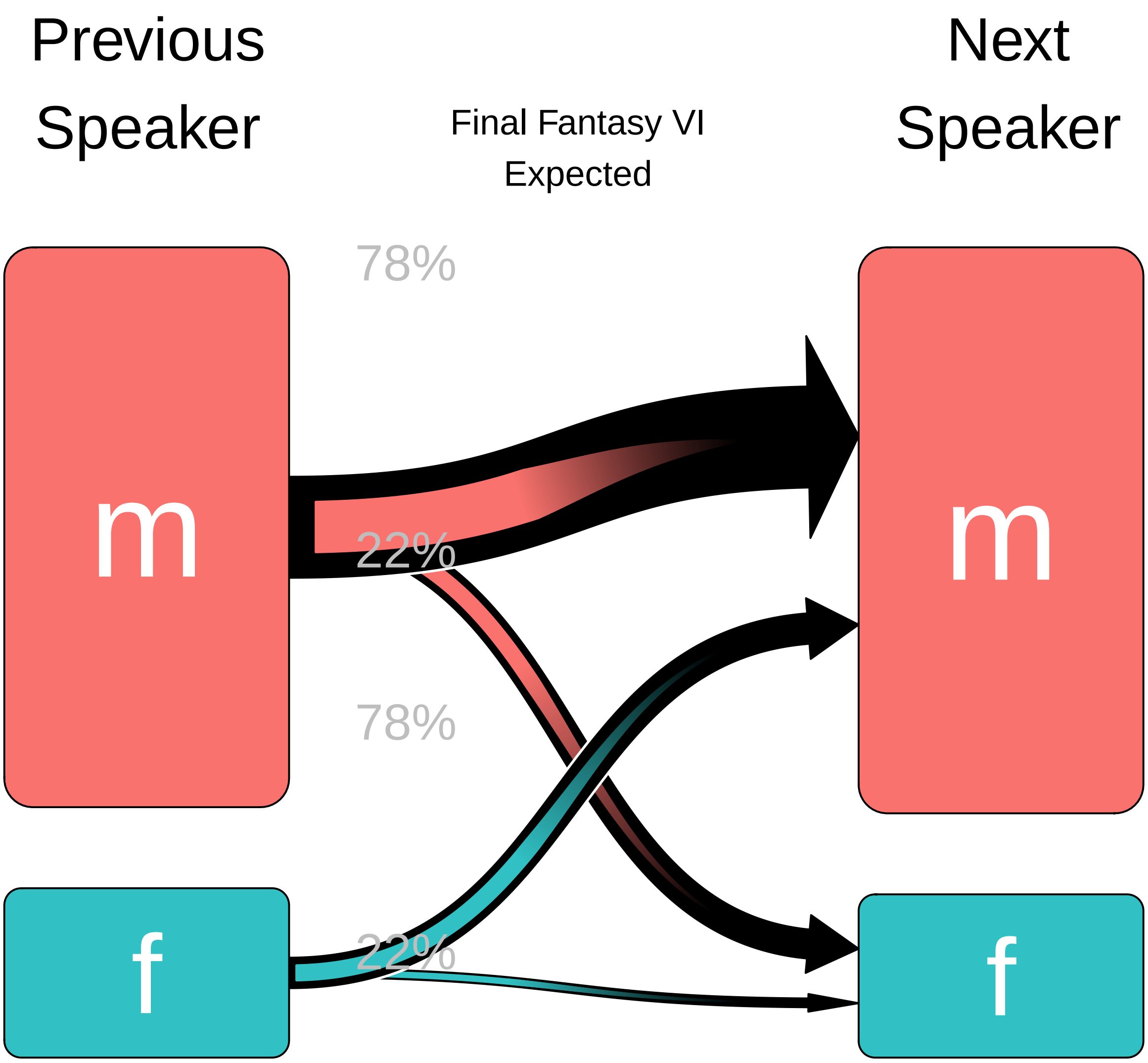
Previous
Speaker
Final Fantasy IX
Empirical
Next
Speaker
68%
32%
78%
22%
m
f
m
f

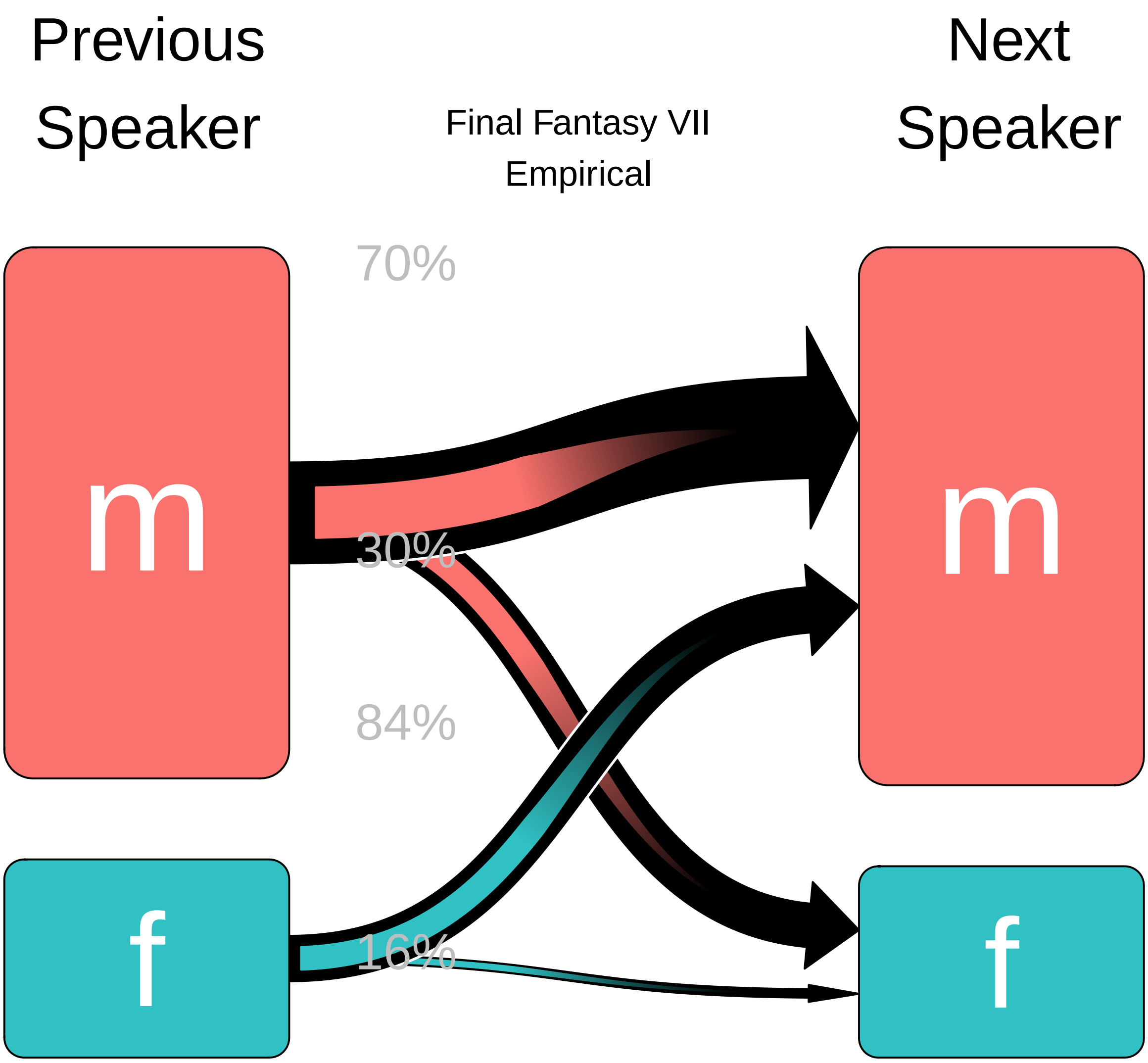

Previous
Speaker
Final Fantasy IX
Expected
Next
Speaker
73%
m
27%
73%
m
f
27%
f

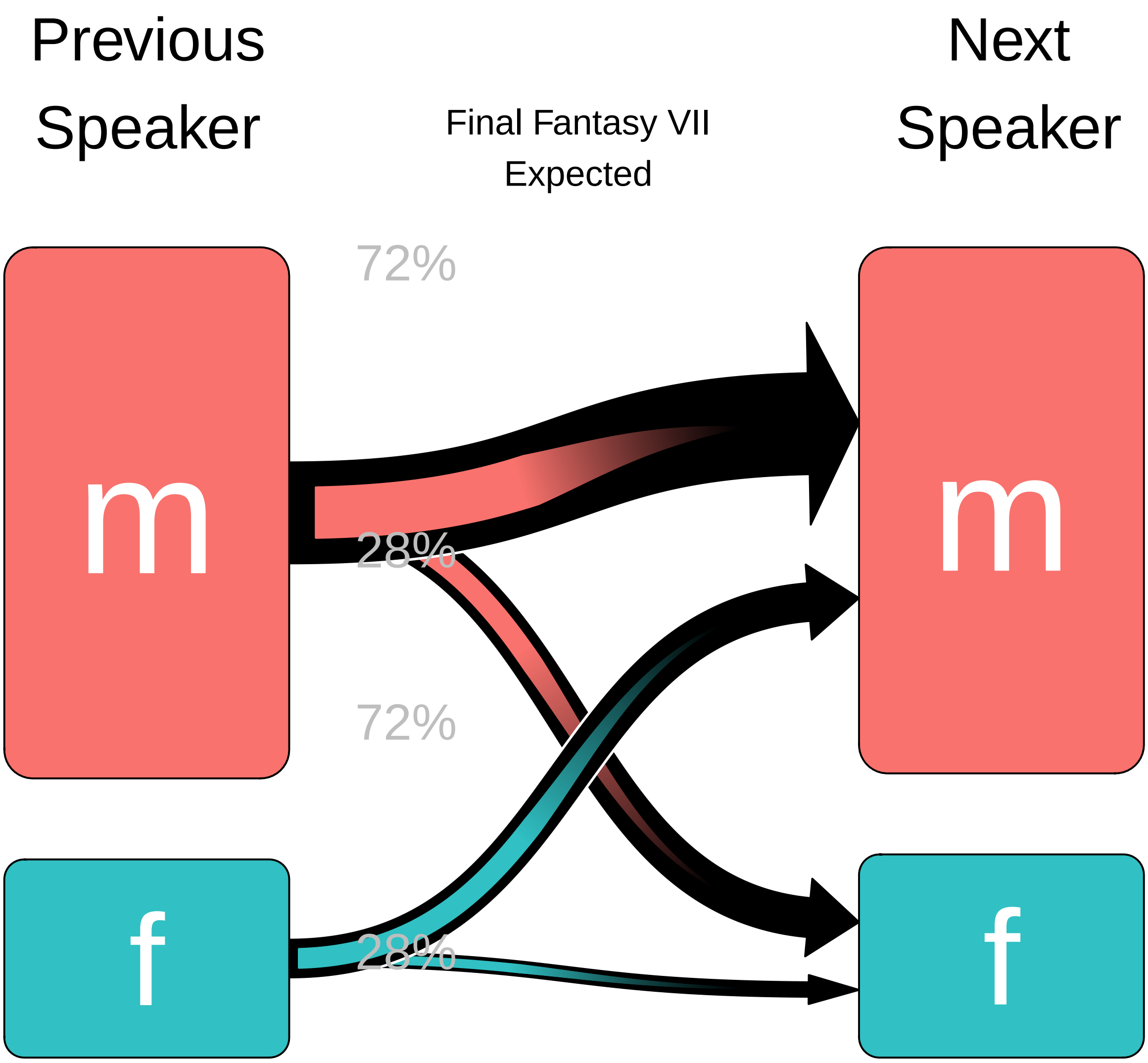

Previous Speaker
Final Fantasy V
Empirical
Next Speaker
69%
31%
62%
38%
m
f
m
f

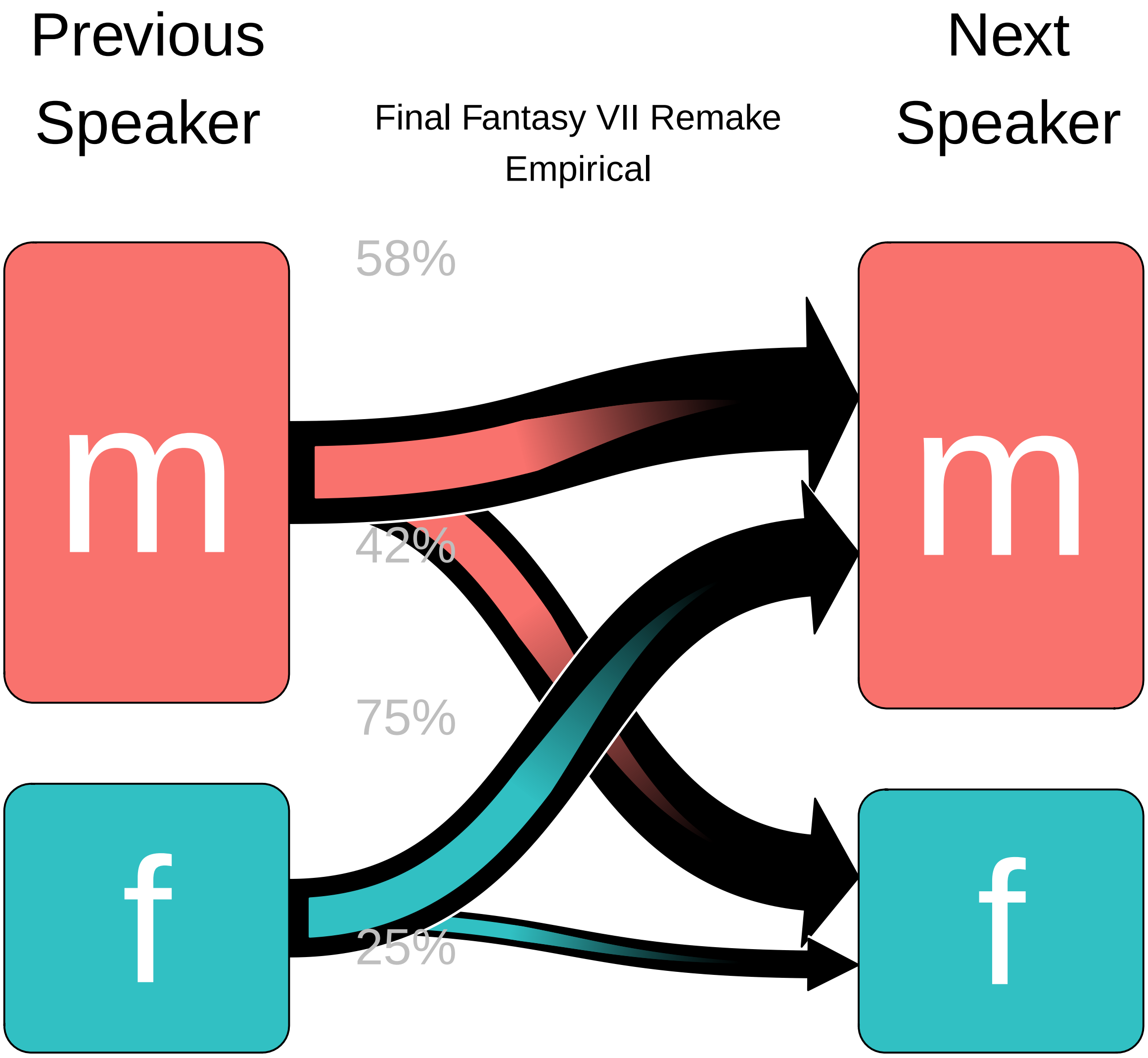

Previous
Speaker
Final Fantasy V
Expected
Next
Speaker
67%
m
33%
67%
m
f
33%
f

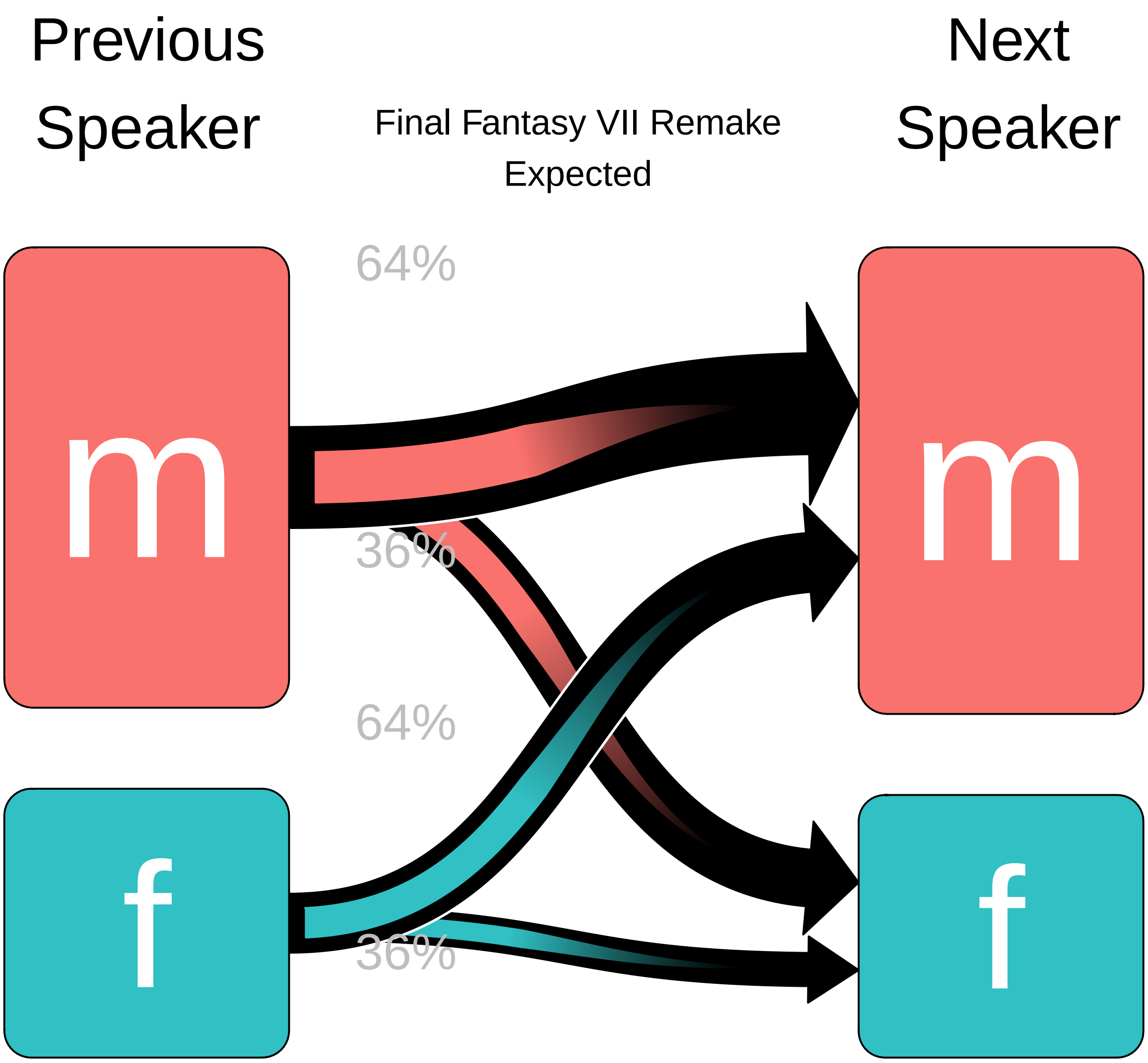

Previous
Speaker
Final Fantasy VI
Empirical
Next
Speaker
71%
m
29%
95%
m
f
5%
f

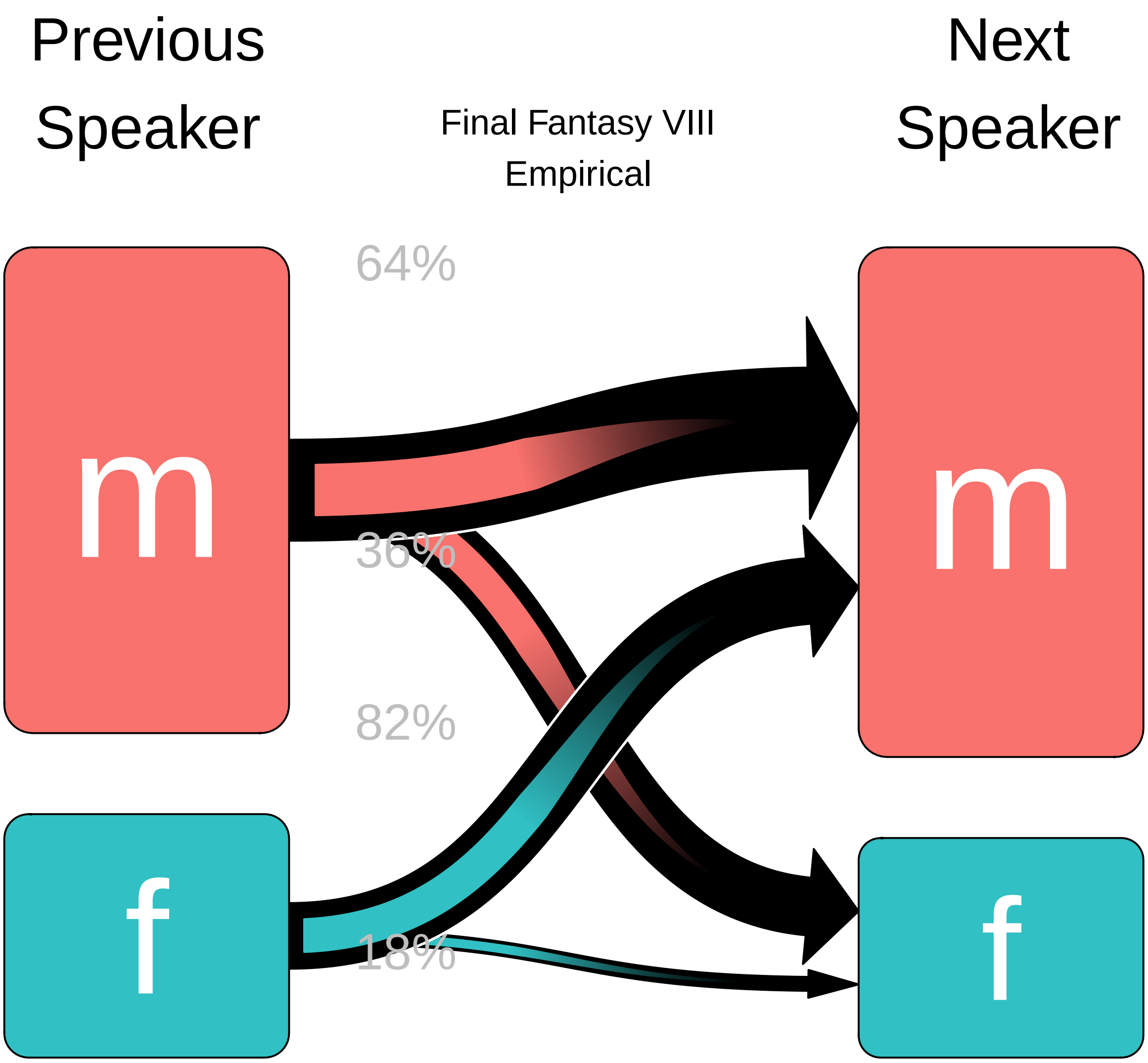

Previous
Speaker
Final Fantasy VI
Expected
Next
Speaker
78%
m
22%
78%
m
f
22%
f

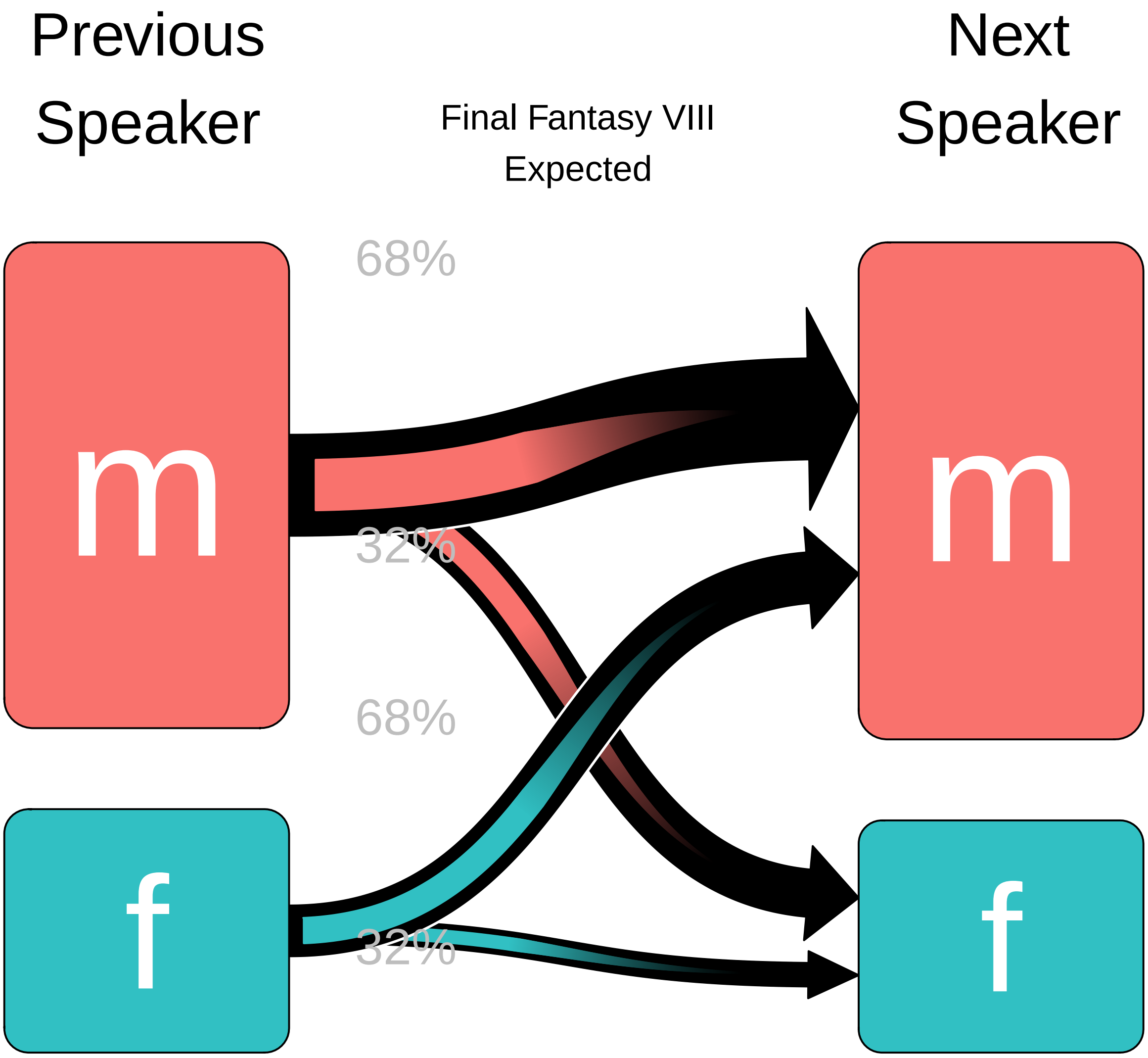

Previous
Speaker
Final Fantasy VII
Empirical
Next
Speaker
70%
m
30%
84%
m
f
16%
f

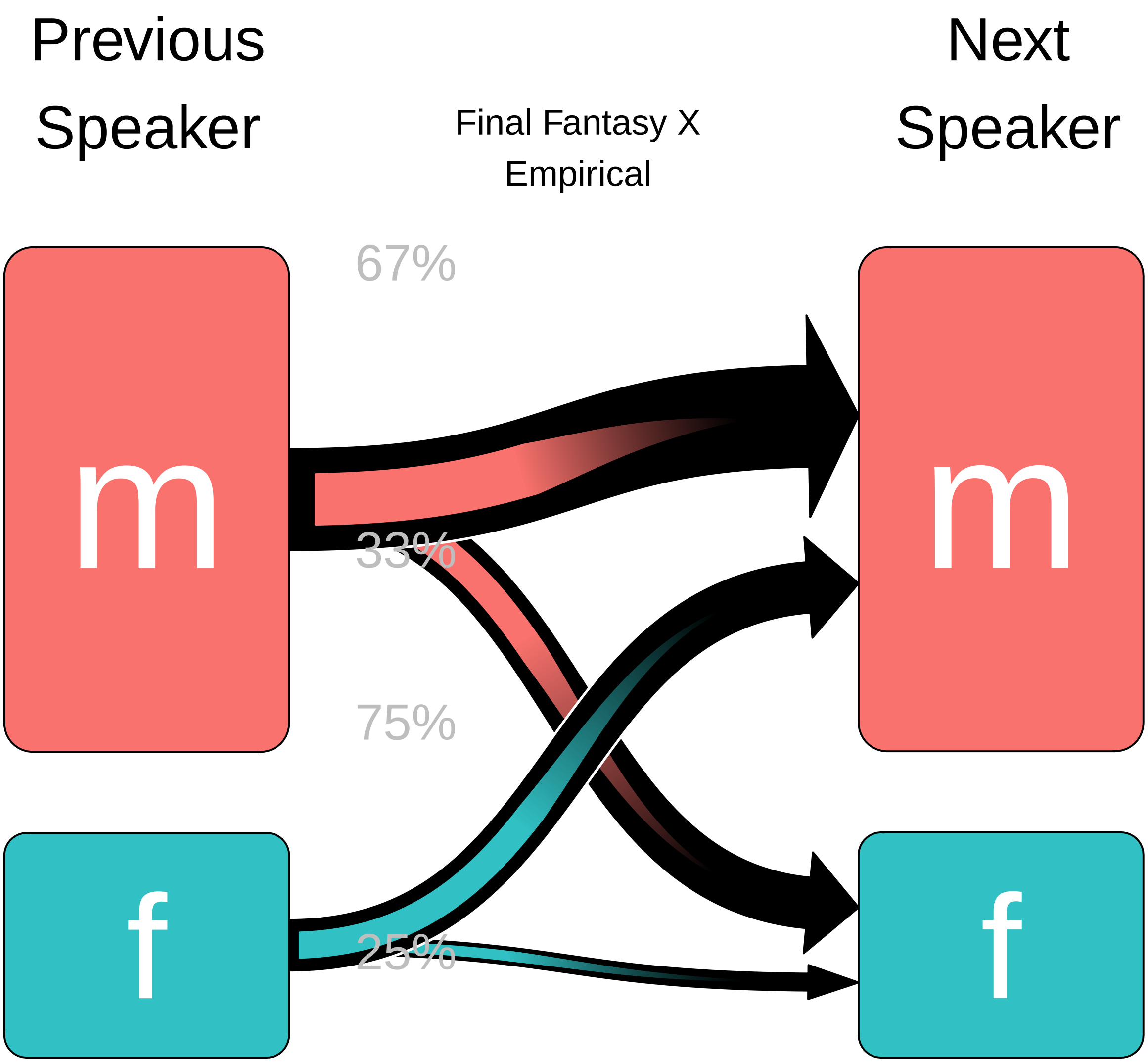

Previous
Speaker
Final Fantasy VII
Expected
Next
Speaker
72%
28%
72%
28%
m
m
f
f

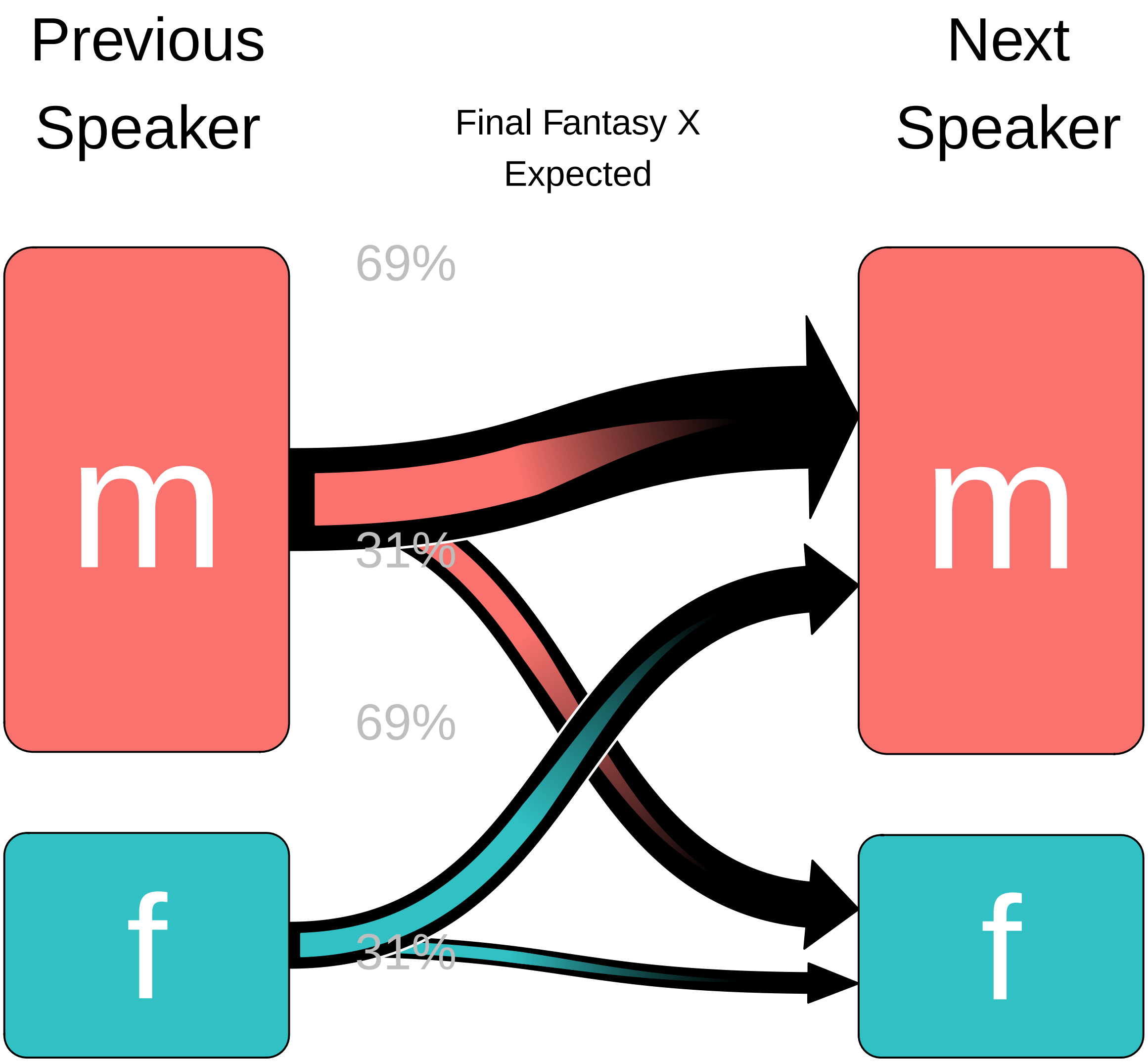

Previous
Speaker
Final Fantasy VII Remake
Empirical
Next
Speaker
58%
m
m
42%
75%
f
f
25%

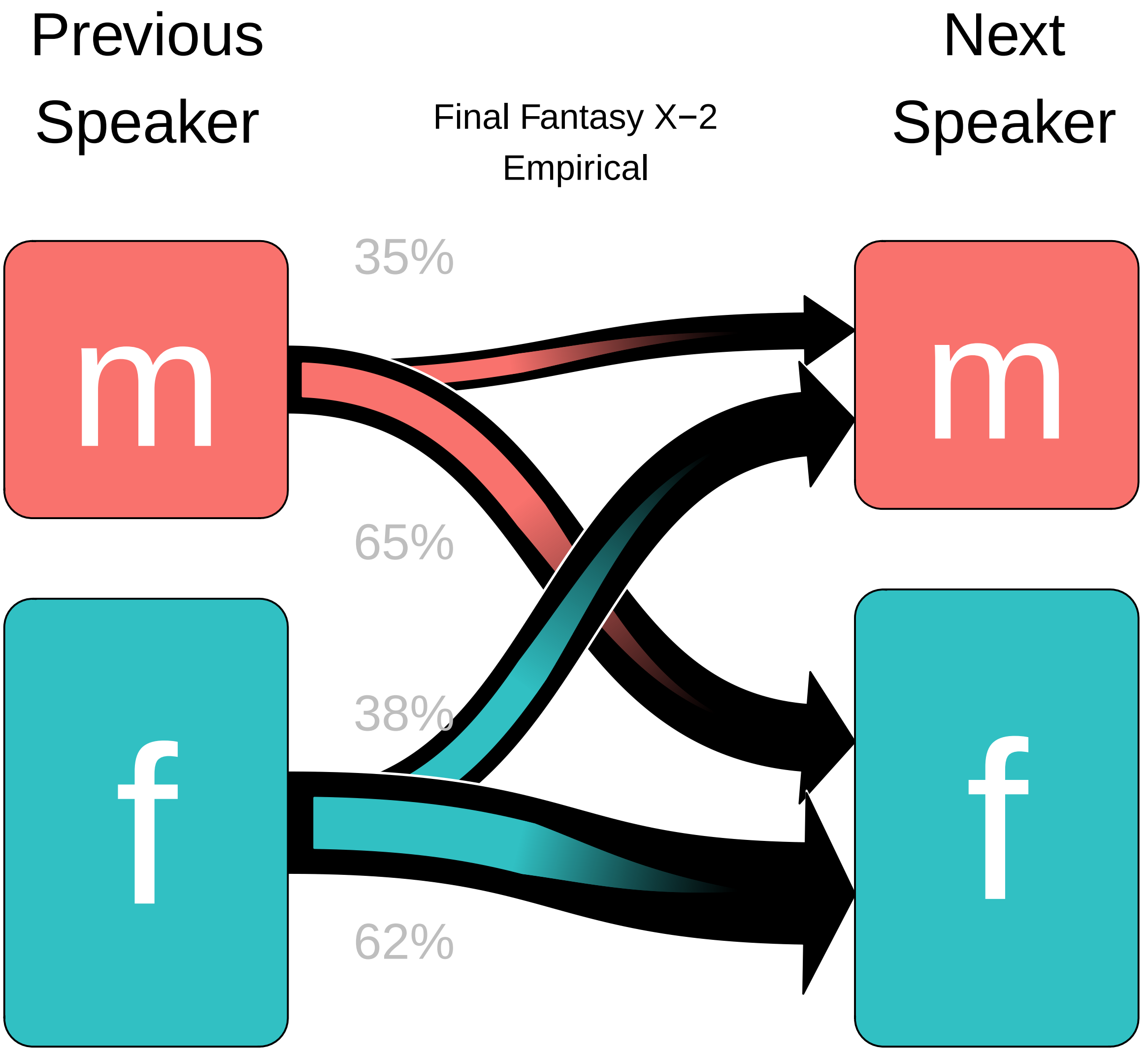

Previous
Speaker
Final Fantasy VII Remake
Expected
Next
Speaker
64%
m
36%
64%
f
36%
m
f

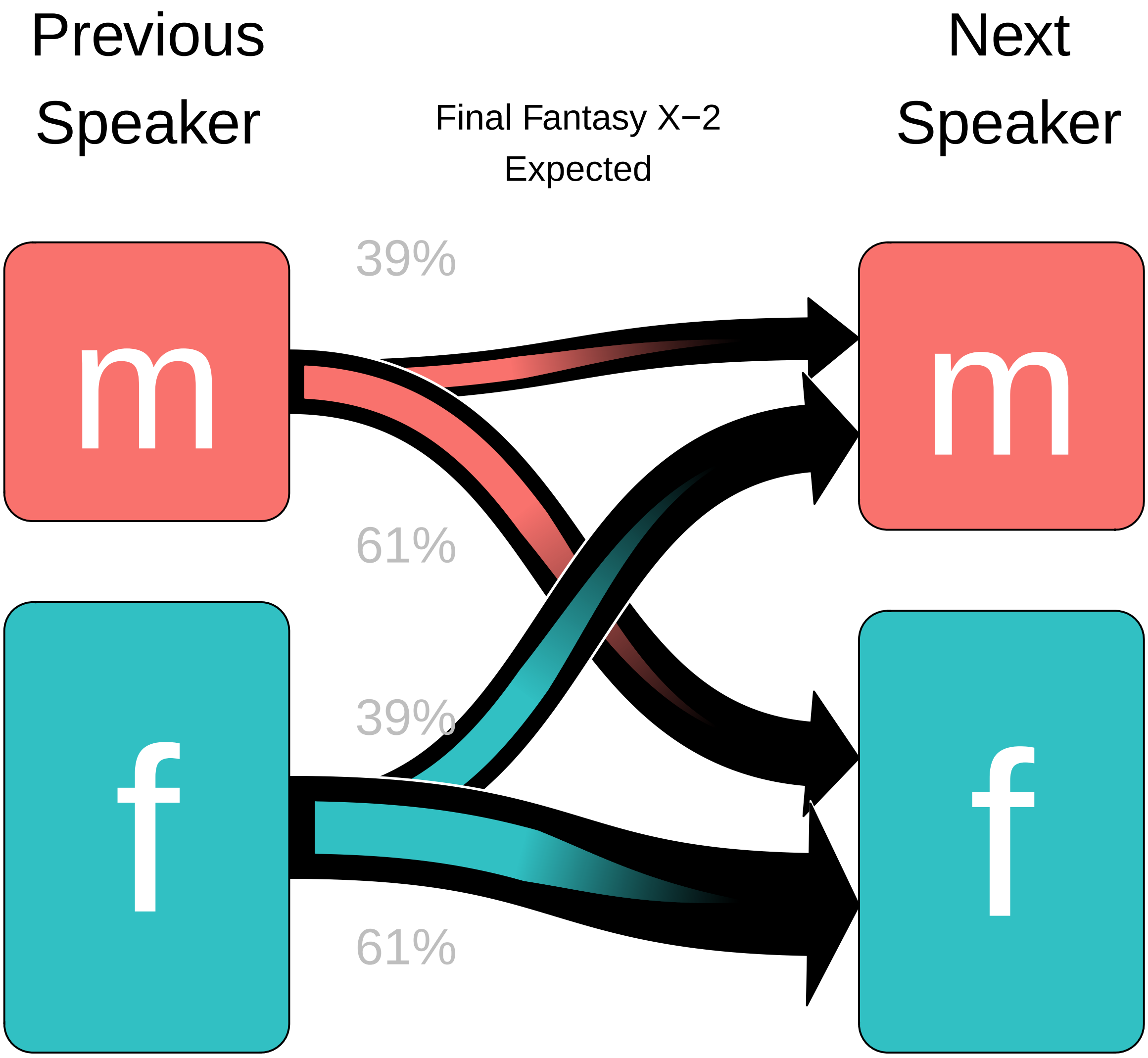

Previous
Speaker
Final Fantasy VIII
Empirical
Next
Speaker
64%
36%
82%
18%
m
f
m
f

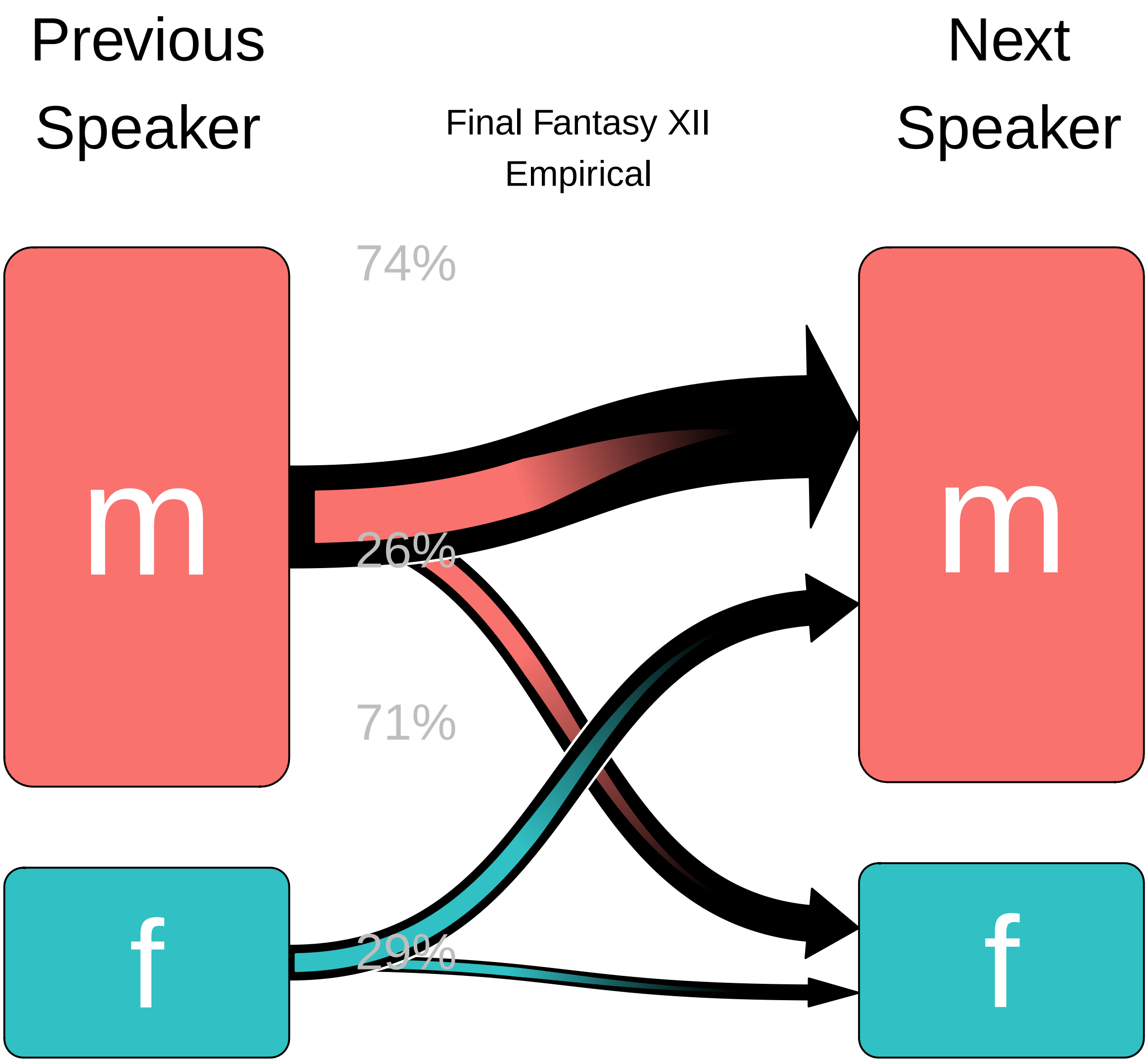

Previous
Speaker
Final Fantasy VIII
Expected
Next
Speaker
68%
m
32%
68%
m
f
32%
f

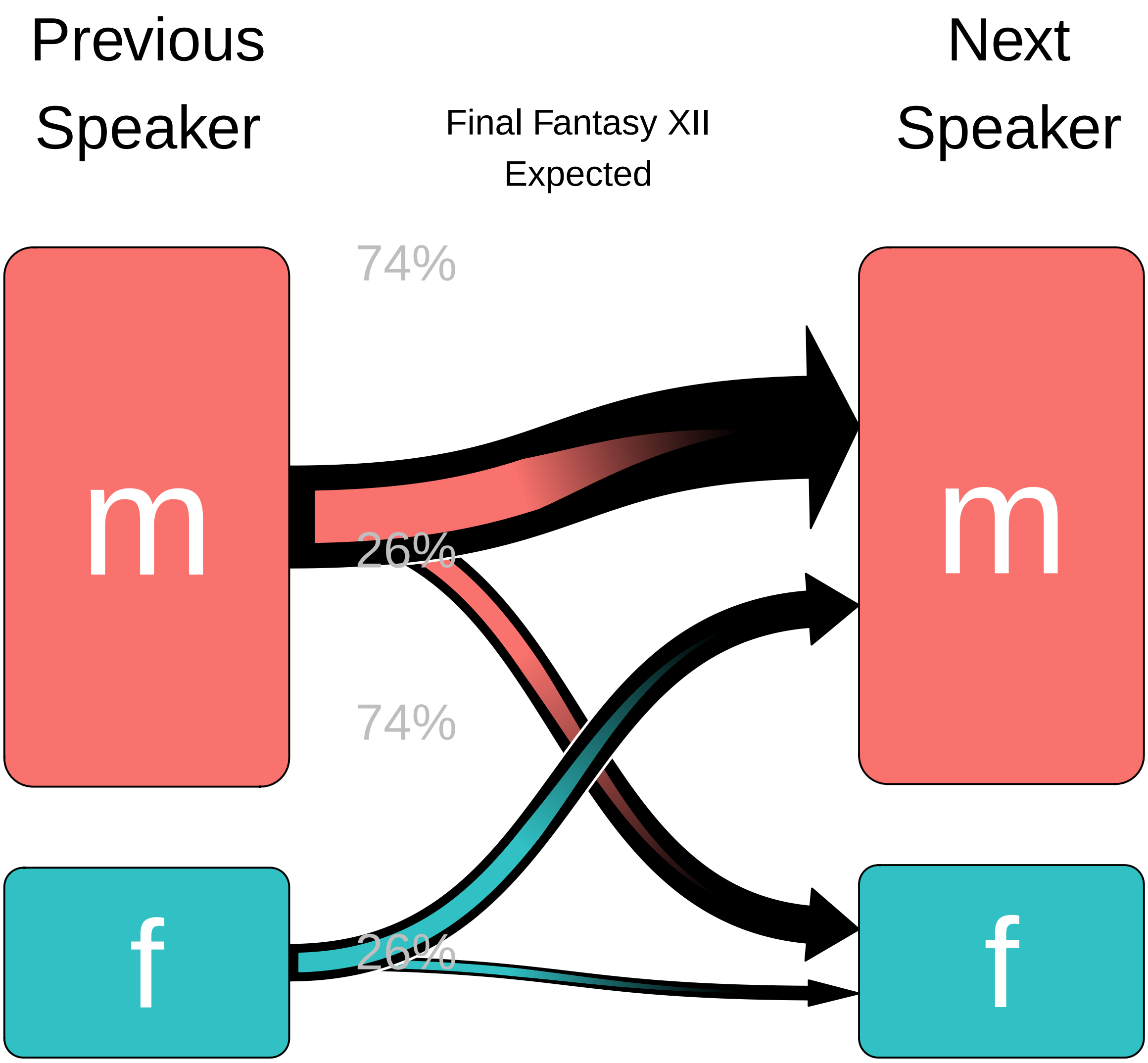

Previous
Speaker
Final Fantasy X
Empirical
Next
Speaker
67%
33%
75%
25%
m
f
m
f


Previous
Speaker
Final Fantasy X
Expected
Next
Speaker
69%
m
31%
69%
m
f
31%
f

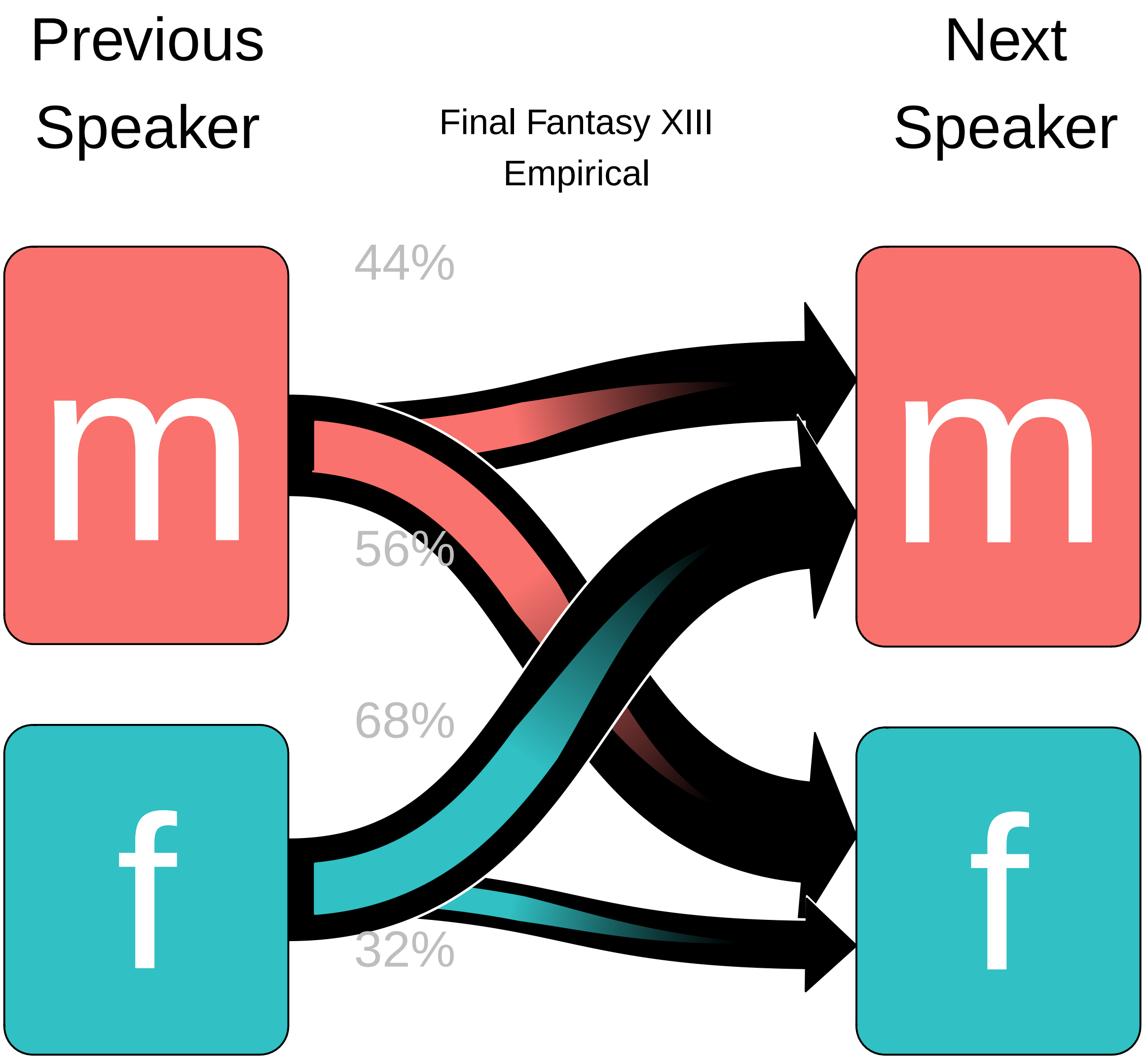

Previous
Speaker
Final Fantasy X−2
Empirical
Next
Speaker
m
f
m
f
35%
65%
38%
62%

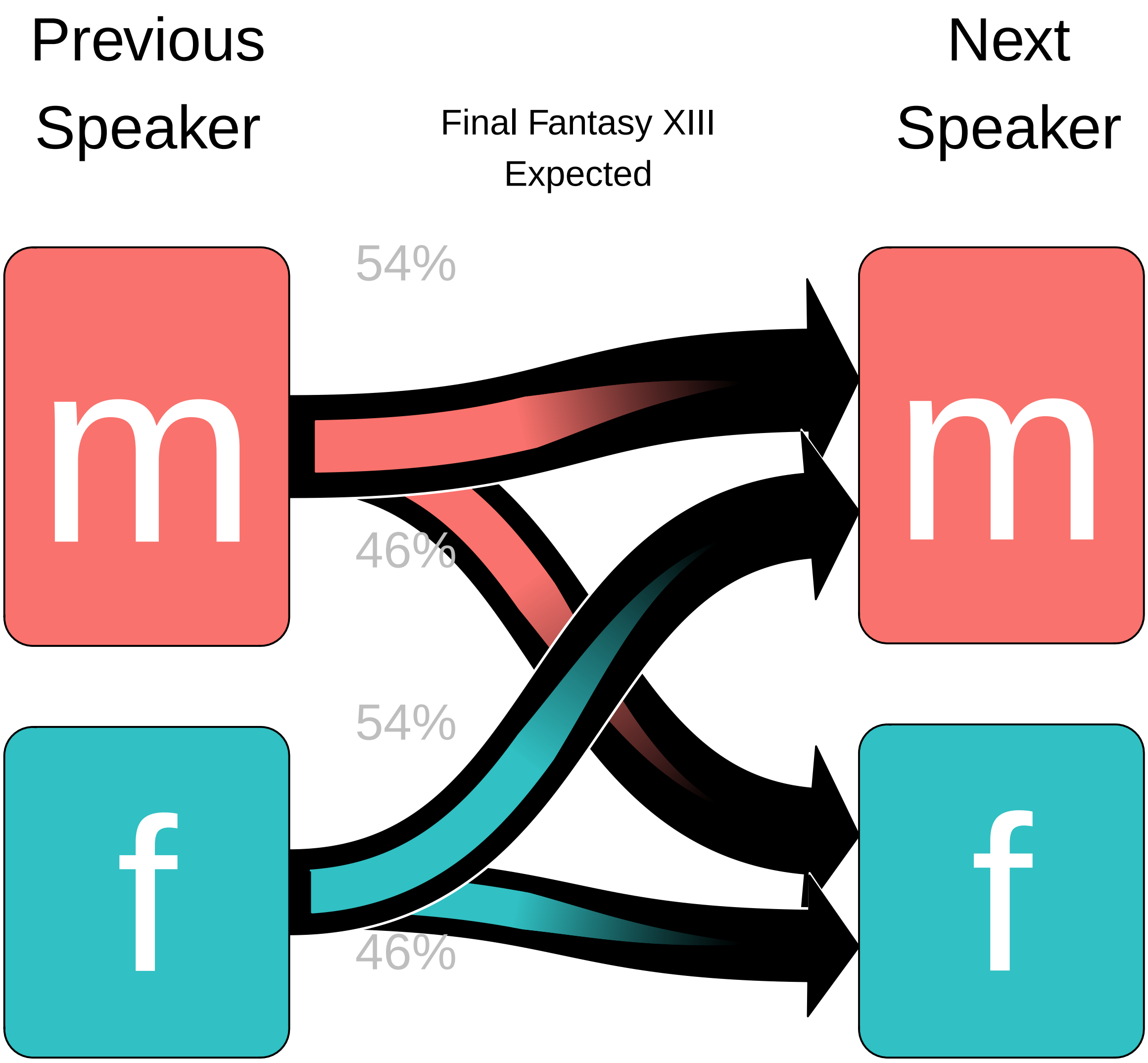

Previous
Speaker
Final Fantasy X−2
Expected
Next
Speaker
39%
m
61%
39%
f
61%
m
f


Previous
Speaker
Final Fantasy XII
Empirical
Next
Speaker
74%
m
26%
71%
m
f
29%
f

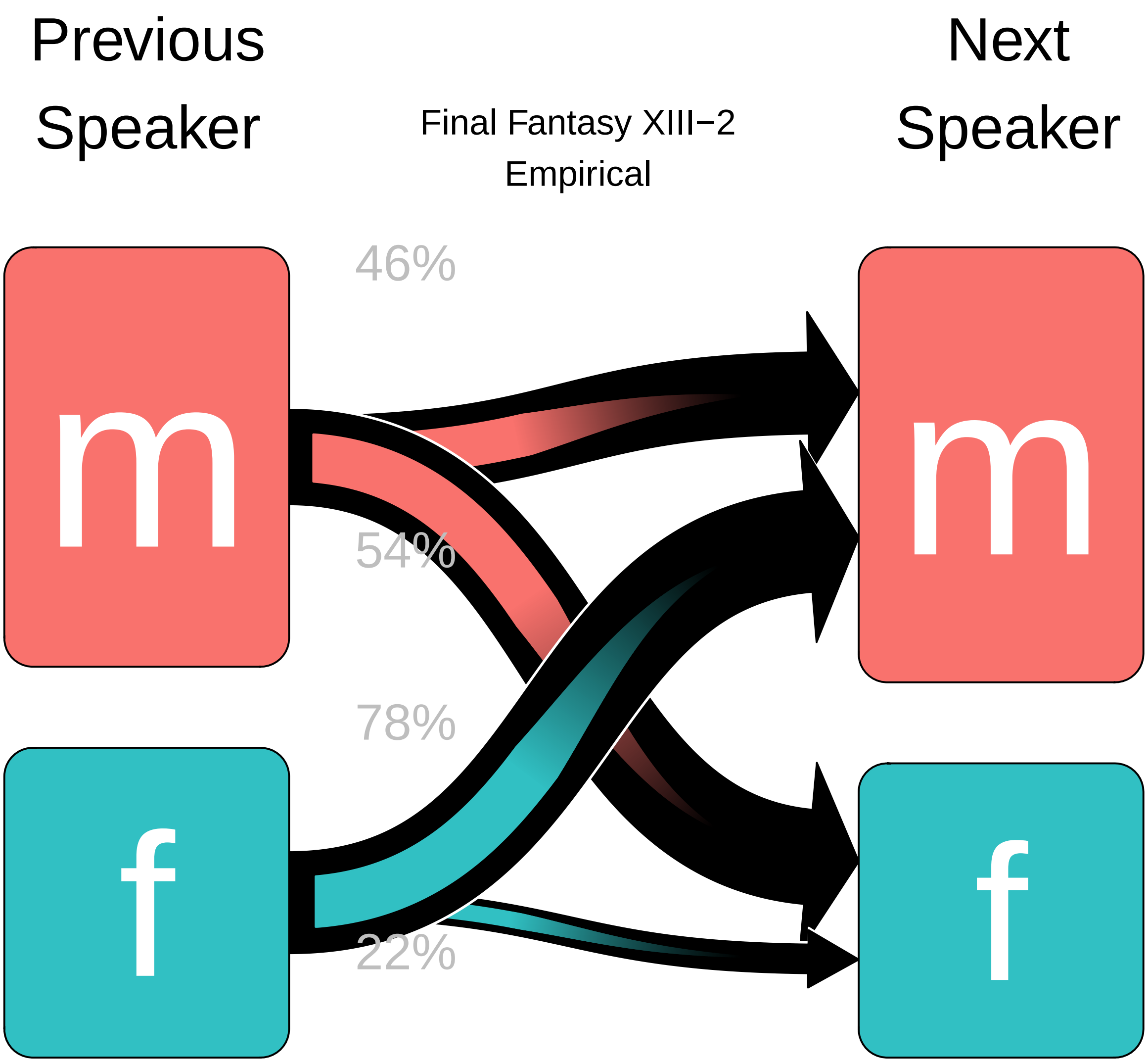


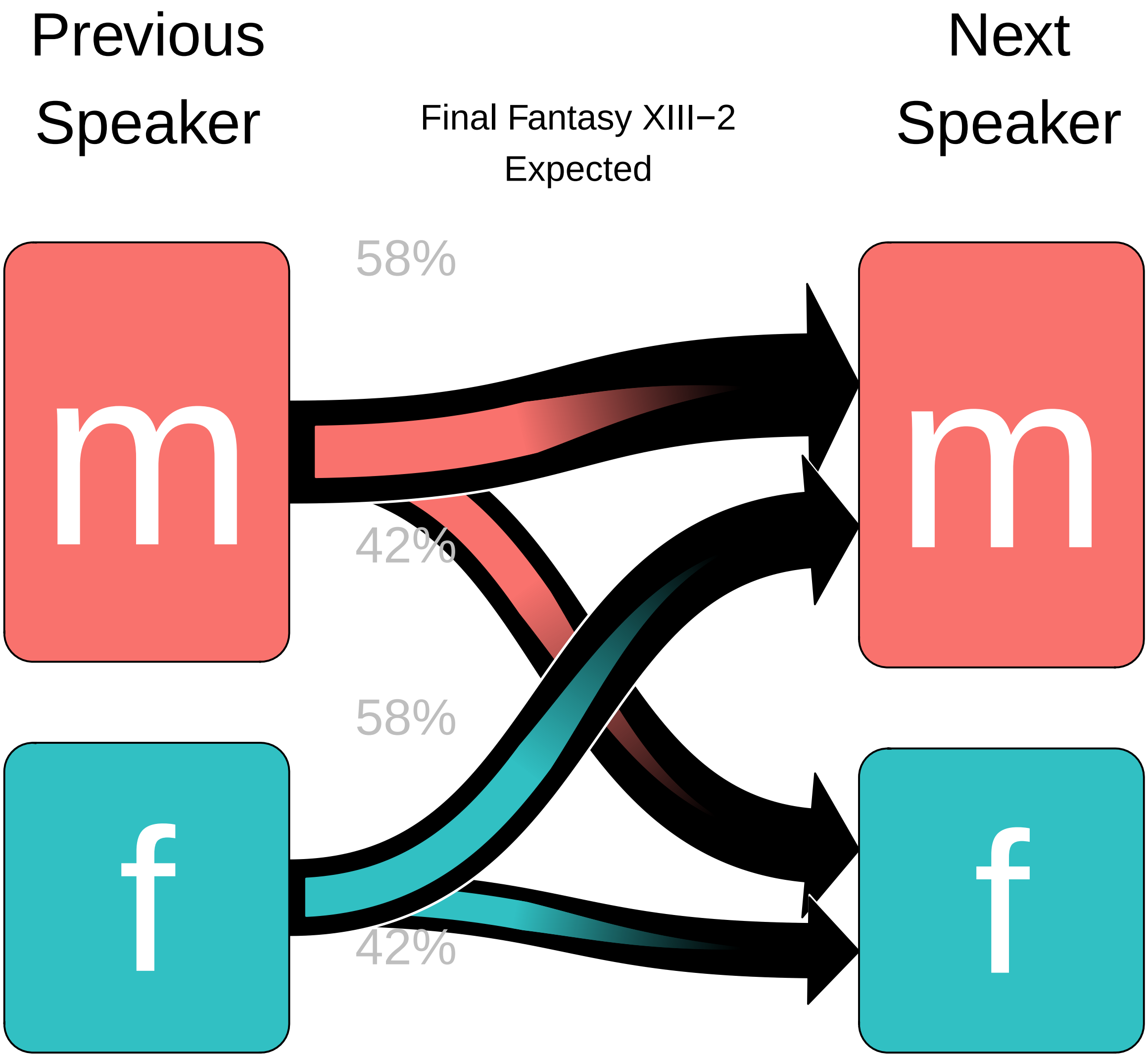
Previous Speaker
Final Fantasy XII
Expected
Next Speaker
74%
m
26%
74%
f
26%
m
f

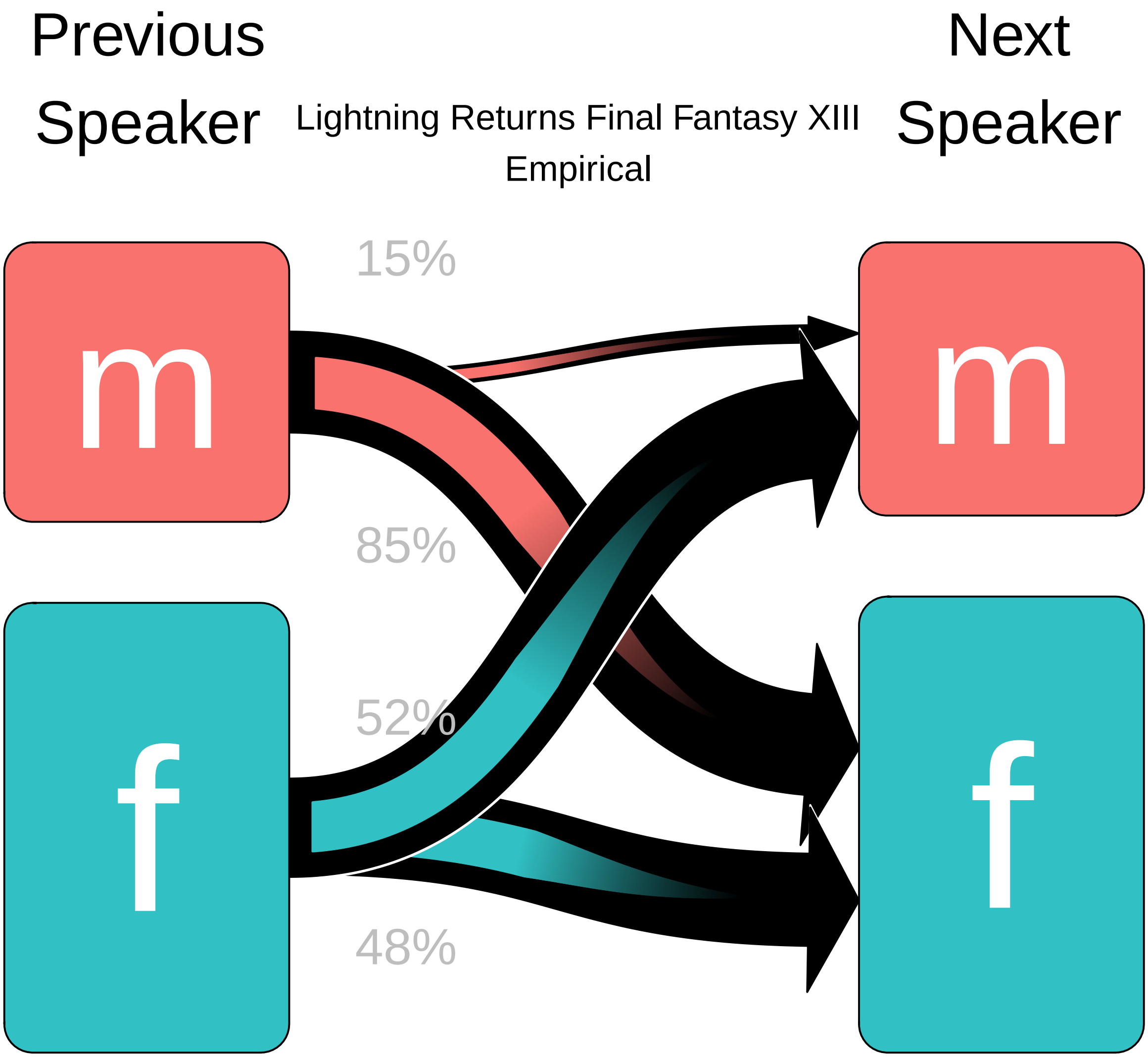

Previous
Speaker
Final Fantasy XIII
Empirical
Next
Speaker
44%
56%
68%
32%
m
f
m
f

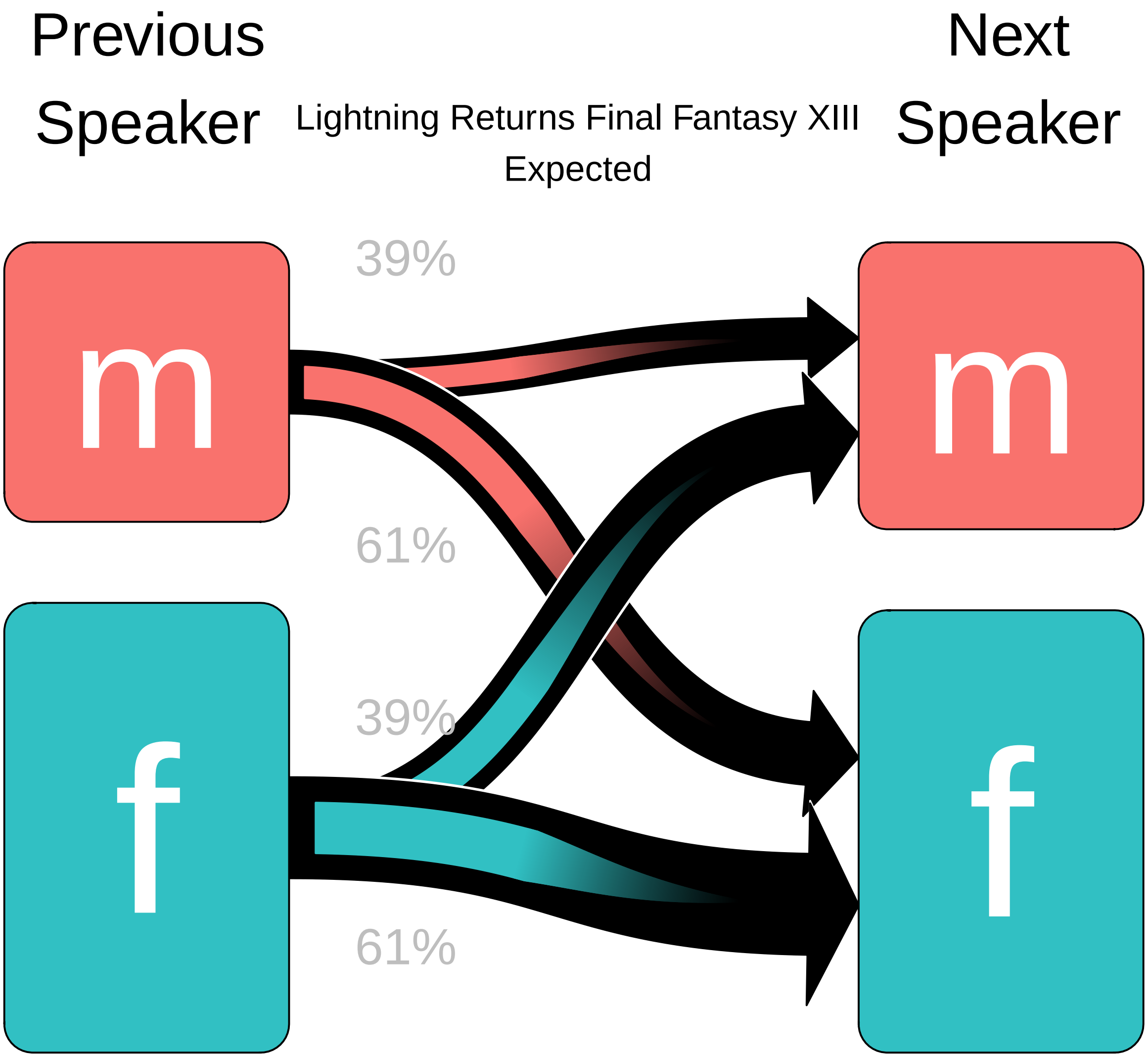


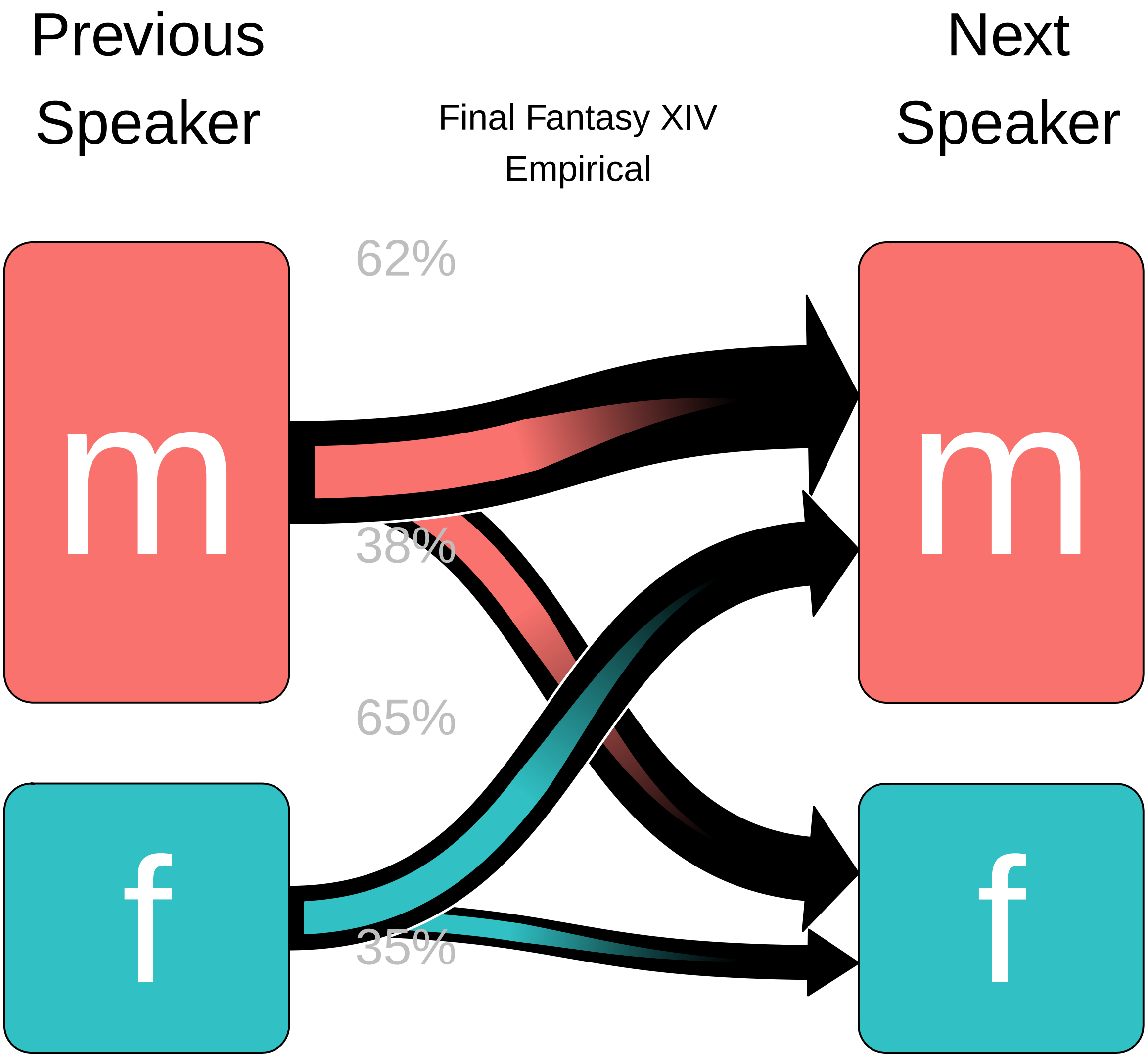
Previous
Speaker
Final Fantasy XIII
Expected
Next
Speaker
54%
46%
54%
46%
m
f
m
f

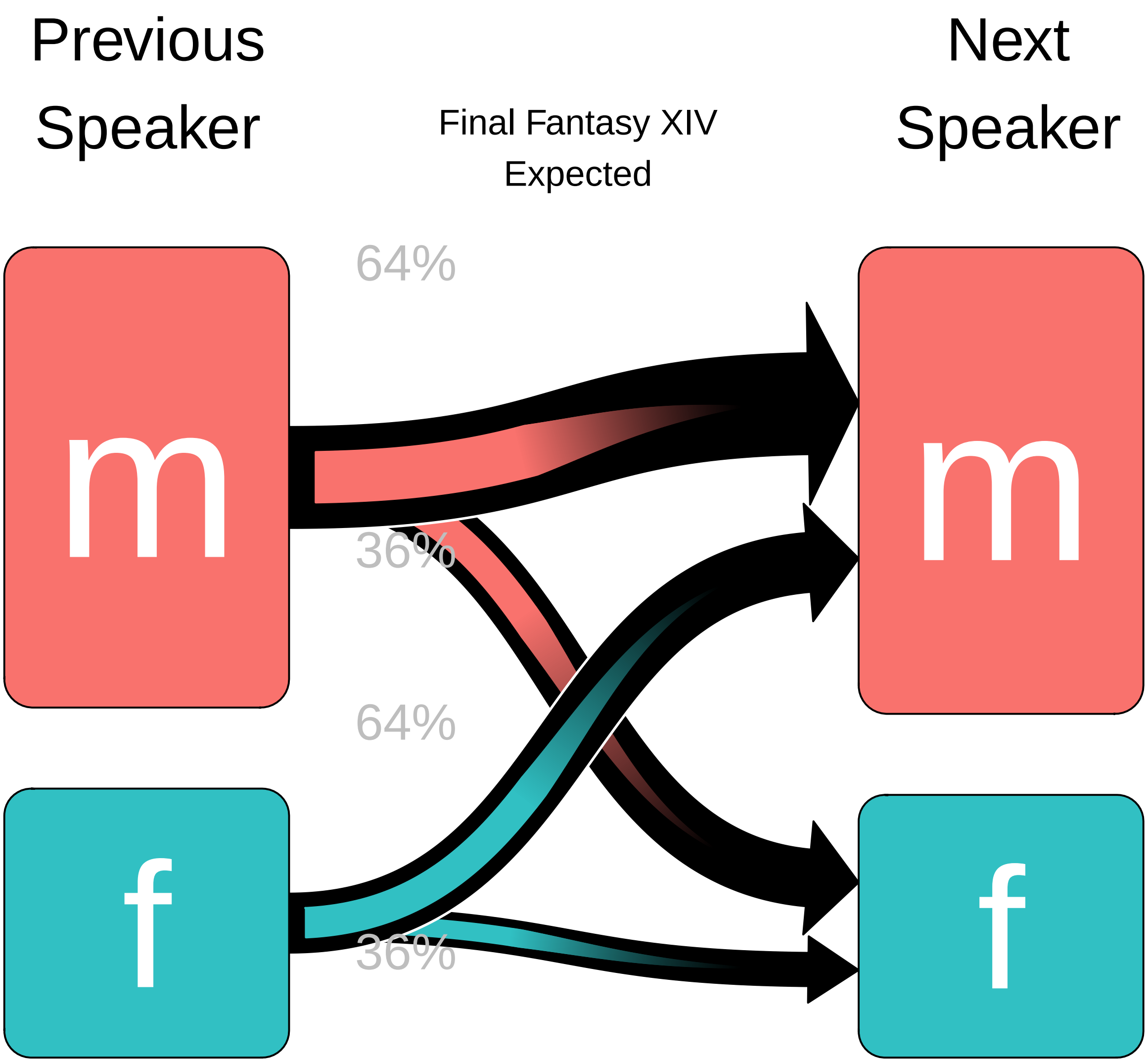

Previous
Speaker
Final Fantasy XIII−2
Empirical
Next
Speaker
46%
54%
78%
22%
m
f
m
f

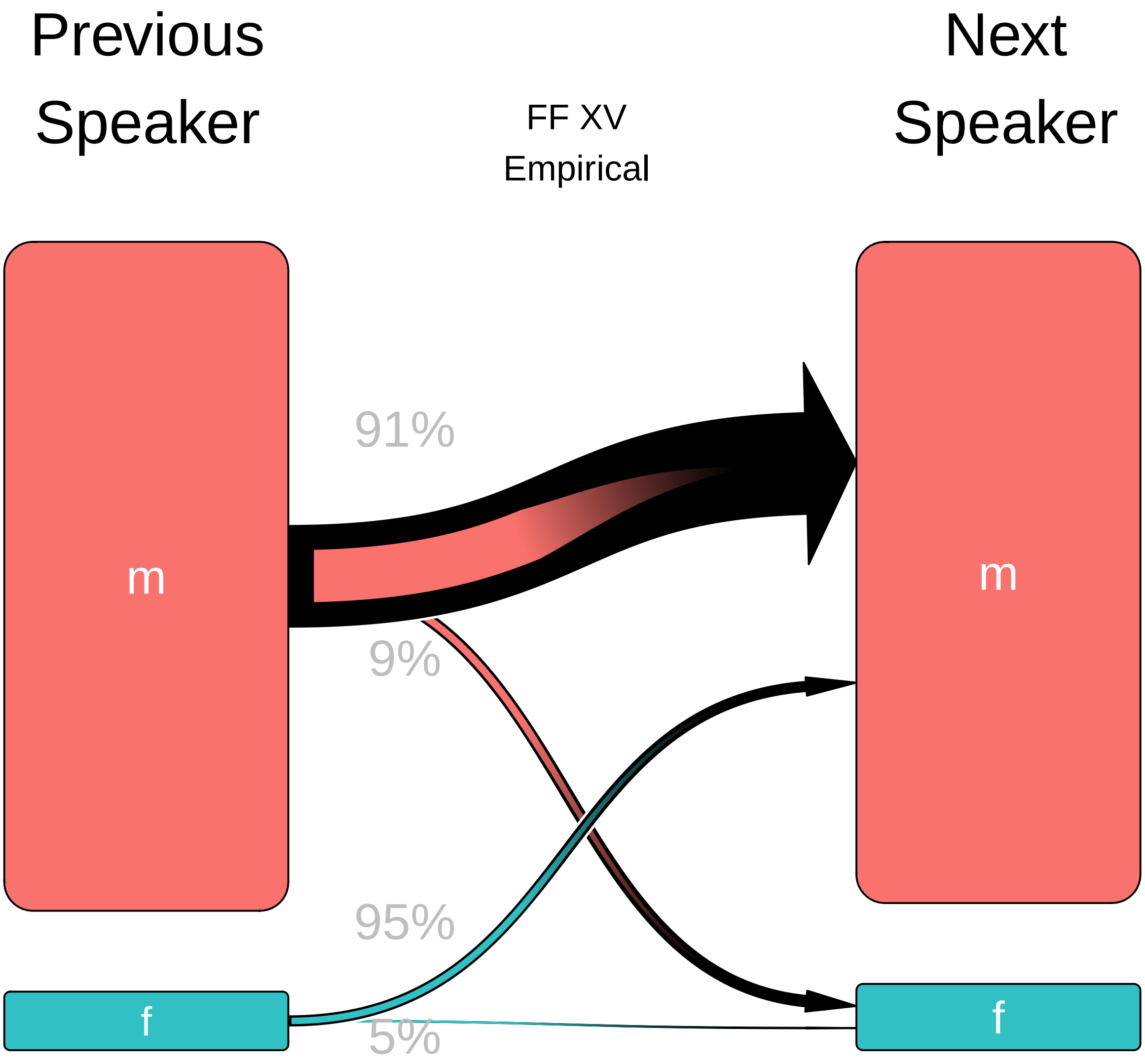

Previous
Speaker
Final Fantasy XIII−2
Expected
Next
Speaker
58%
42%
58%
42%
m
f
m
f

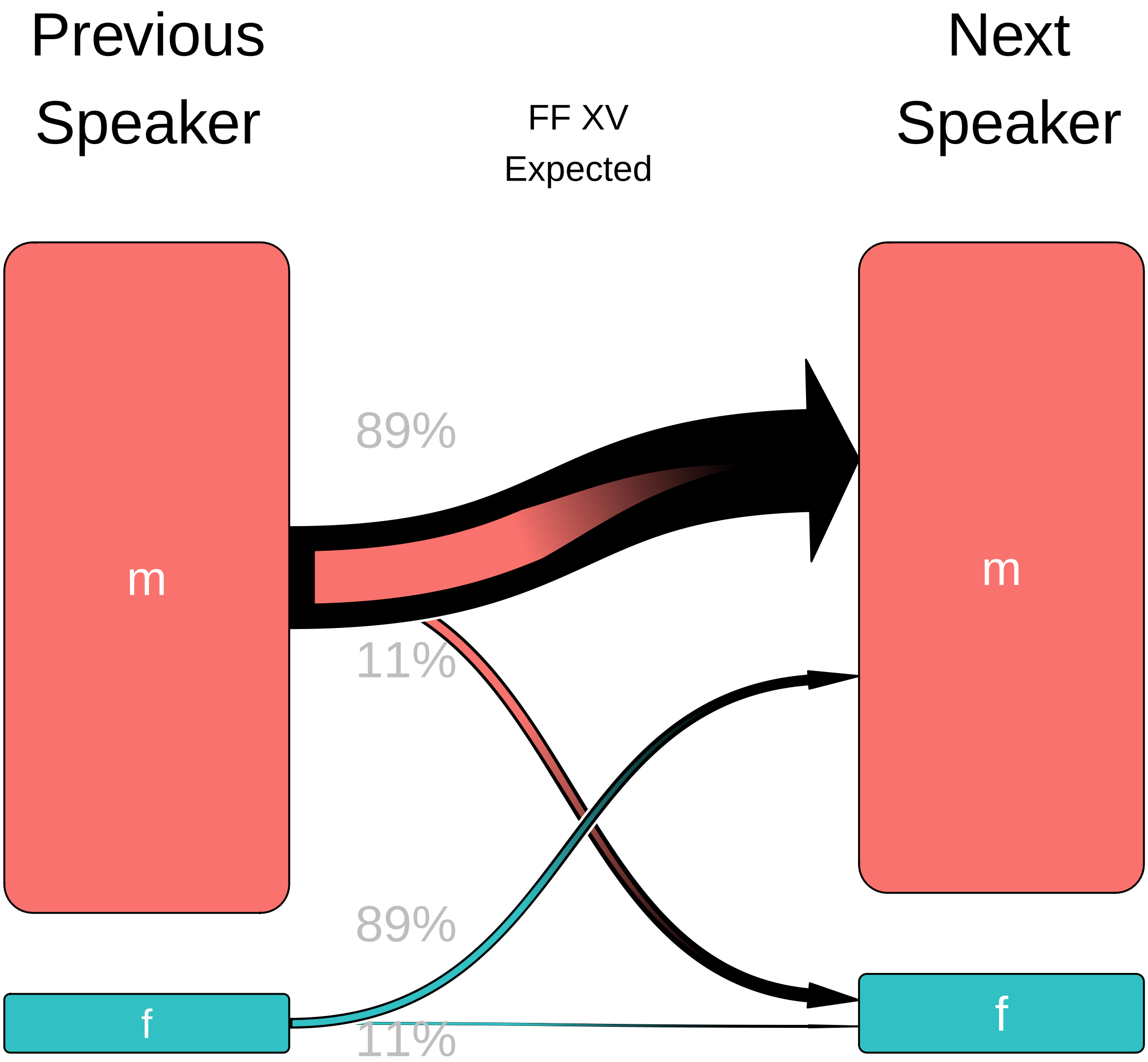

Previous
Speaker
Lightning Returns Final Fantasy XIII
Empirical
Next
Speaker
15%
85%
52%
48%
m
f
m
f

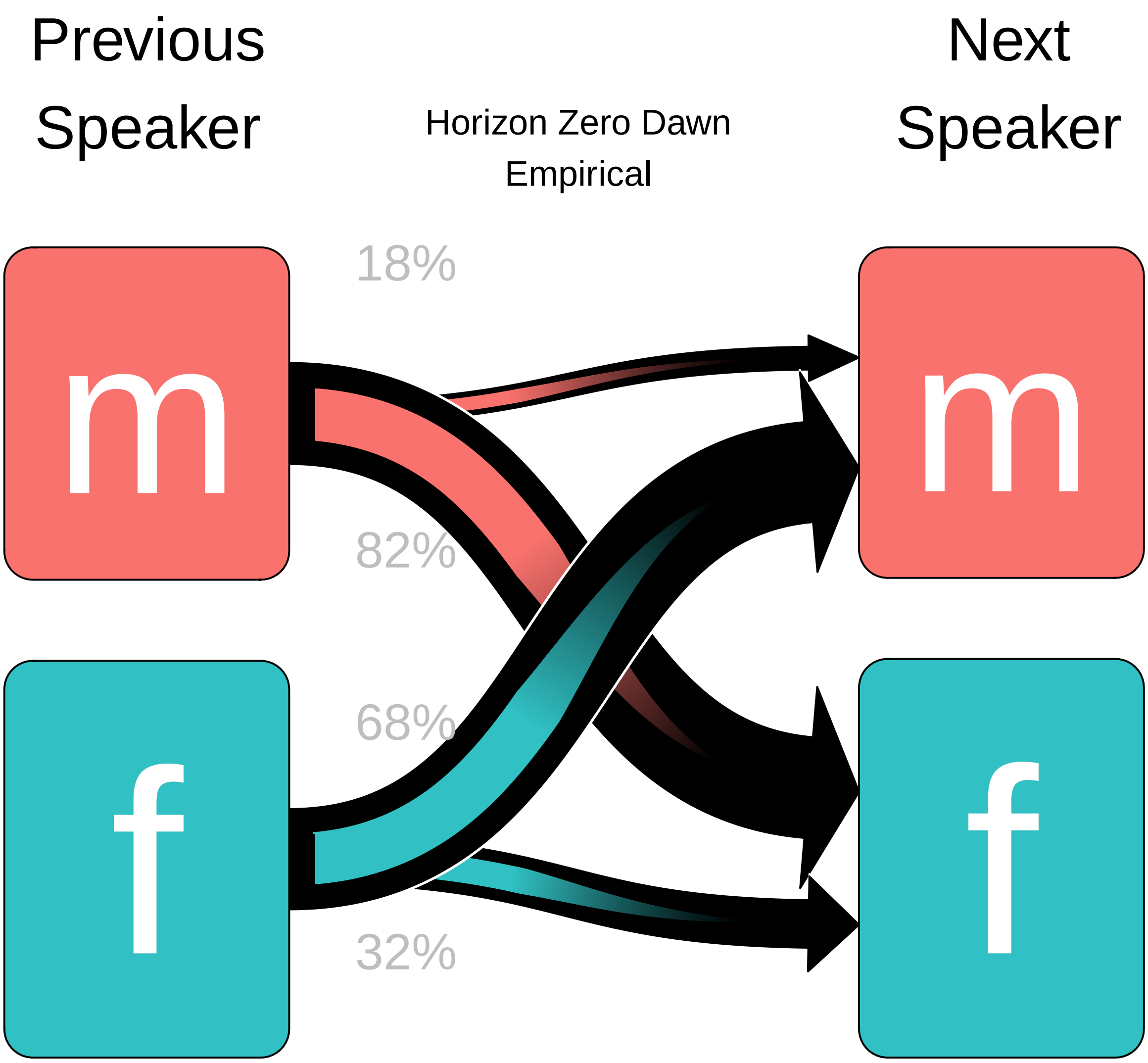

Previous Speaker
Lightning Returns Final Fantasy XIII
Expected
Next Speaker
m
f
39%
61%
39%
61%
m
f

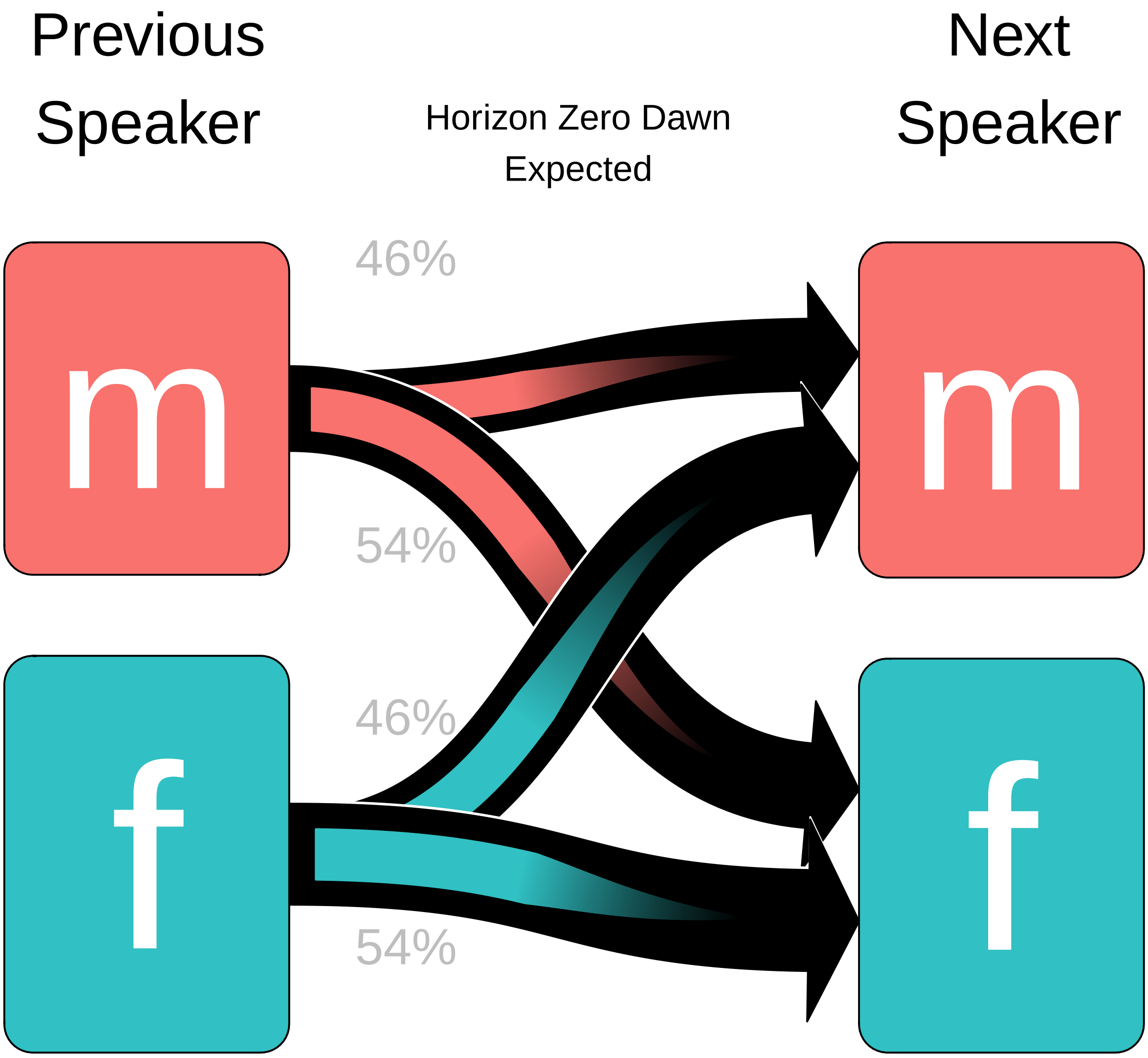

Previous
Speaker
Final Fantasy XIV
Empirical
Next
Speaker
62%
38%
65%
35%
m
f
m
f

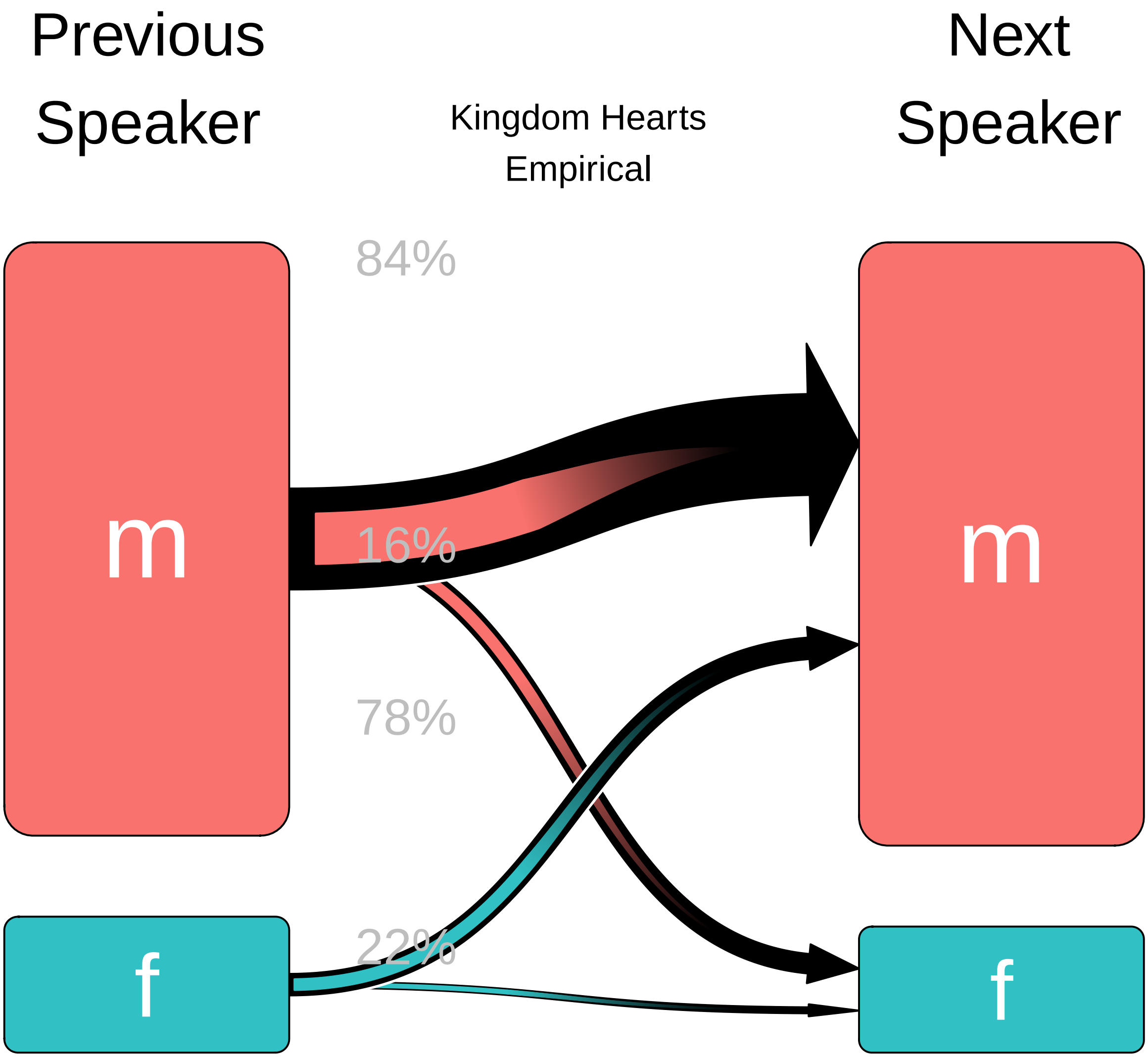

Previous
Speaker
Final Fantasy XIV
Expected
Next
Speaker
64%
m
36%
64%
f
36%
m
f

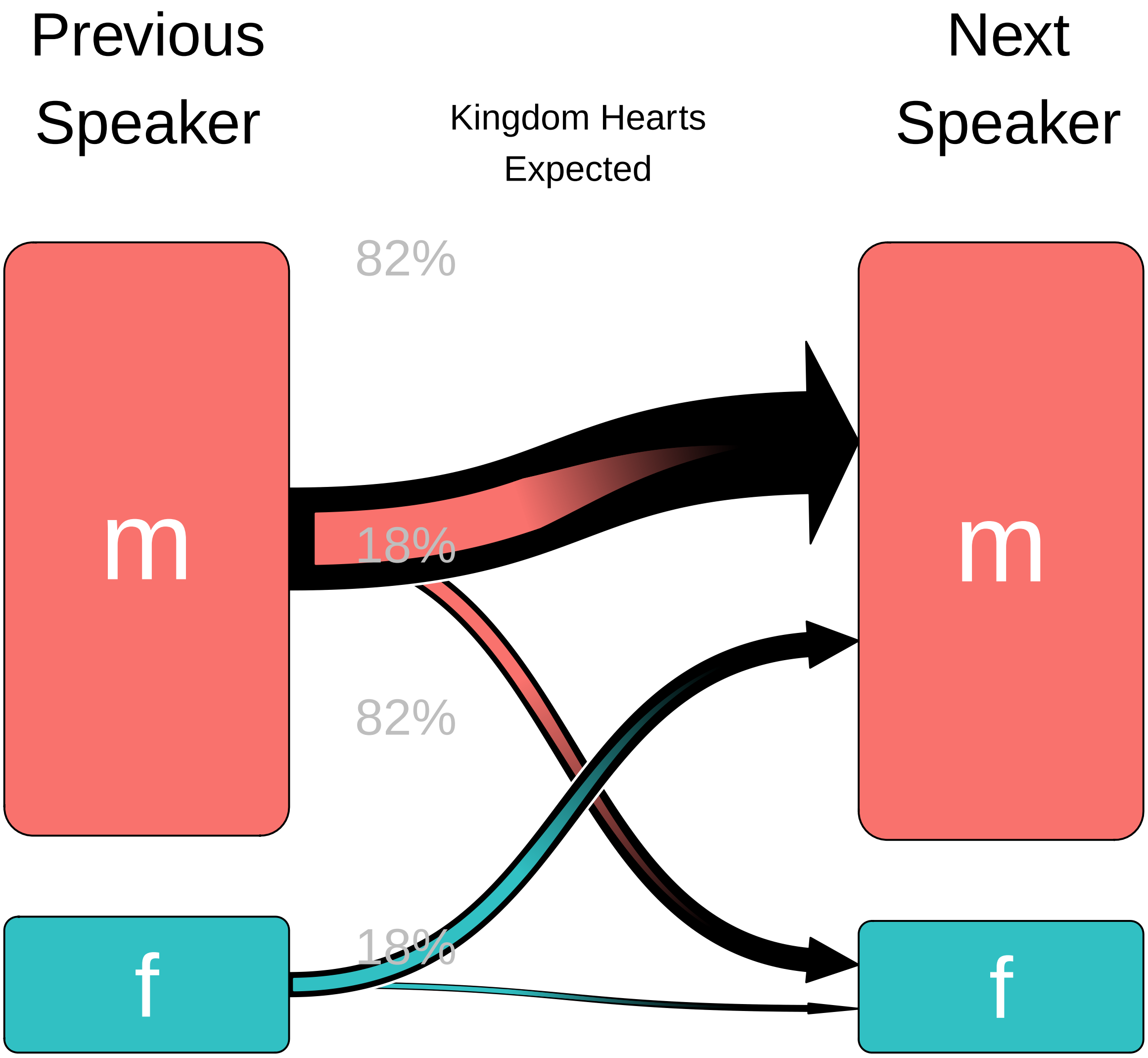

Previous Speaker
FF XV
Empirical
Next Speaker
m
91%
9%
95%
5%
f
m
f


Previous
Speaker
FF XV
Expected
Next
Speaker
m
89%
11%
89%
11%
m
f
f

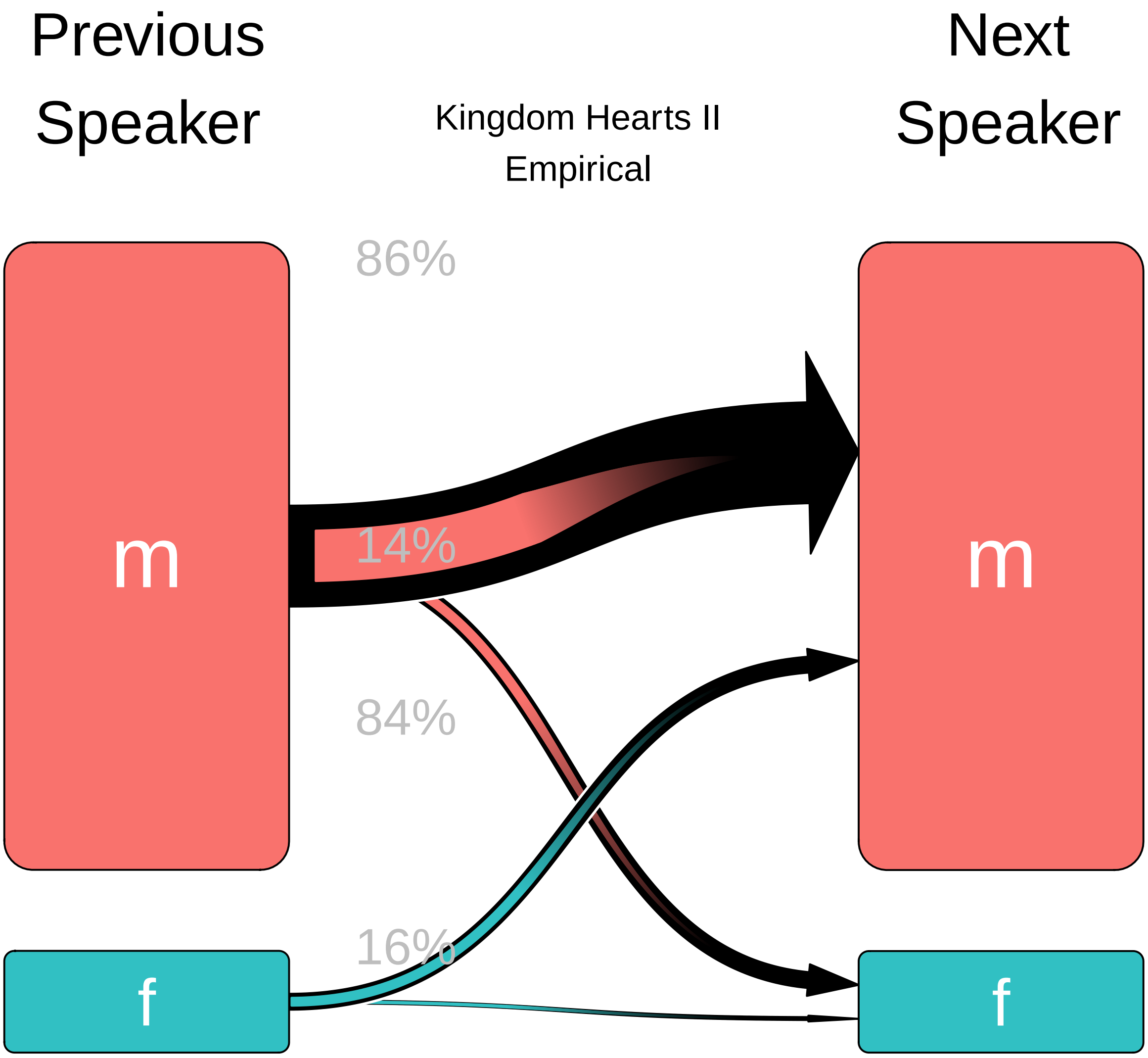

Previous
Speaker
Horizon Zero Dawn
Empirical
Next
Speaker
18%
m
m
82%
68%
f
f
32%

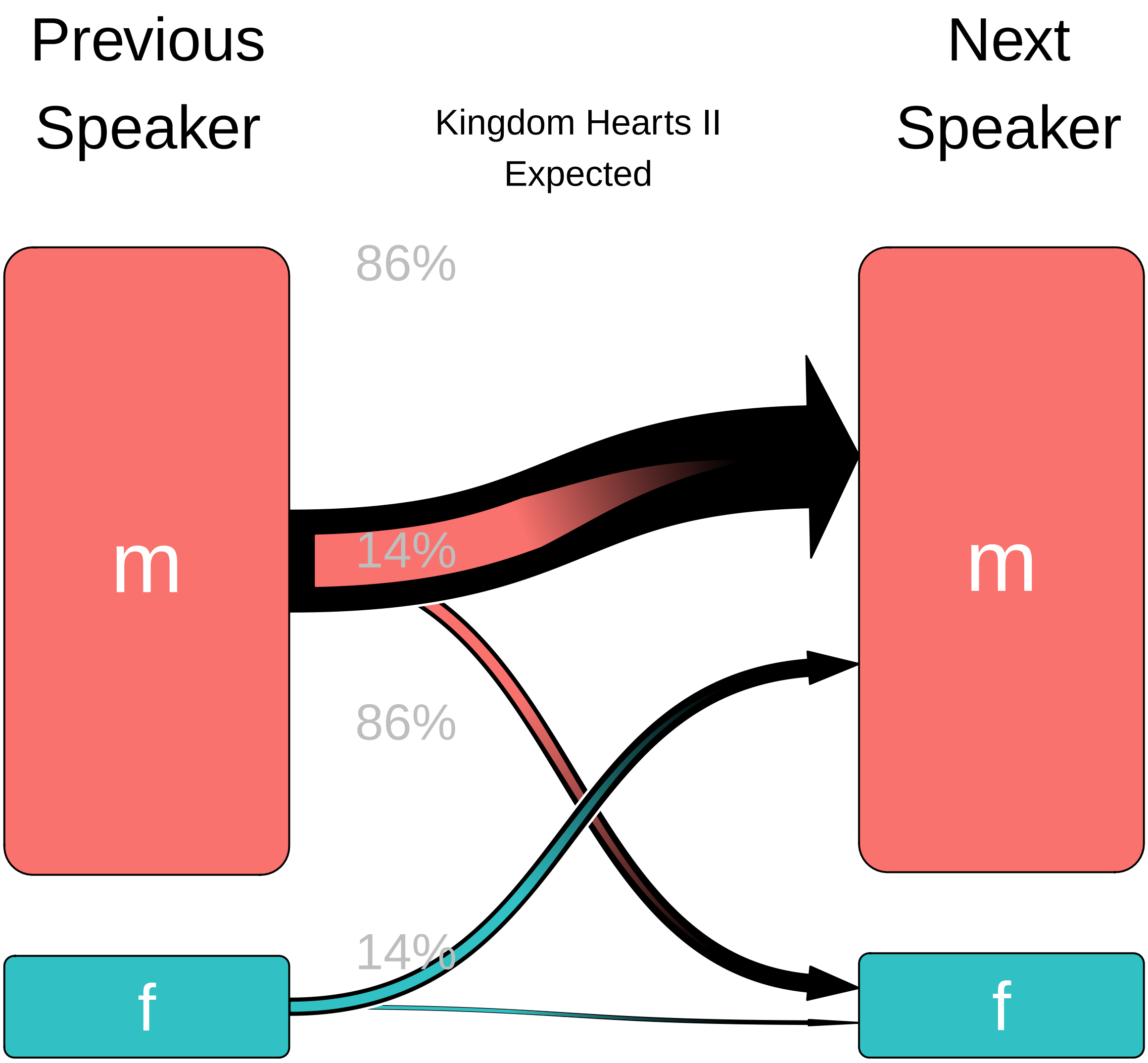

Previous
Speaker
Horizon Zero Dawn
Expected
Next
Speaker
46%
54%
46%
54%
m
f
m
f

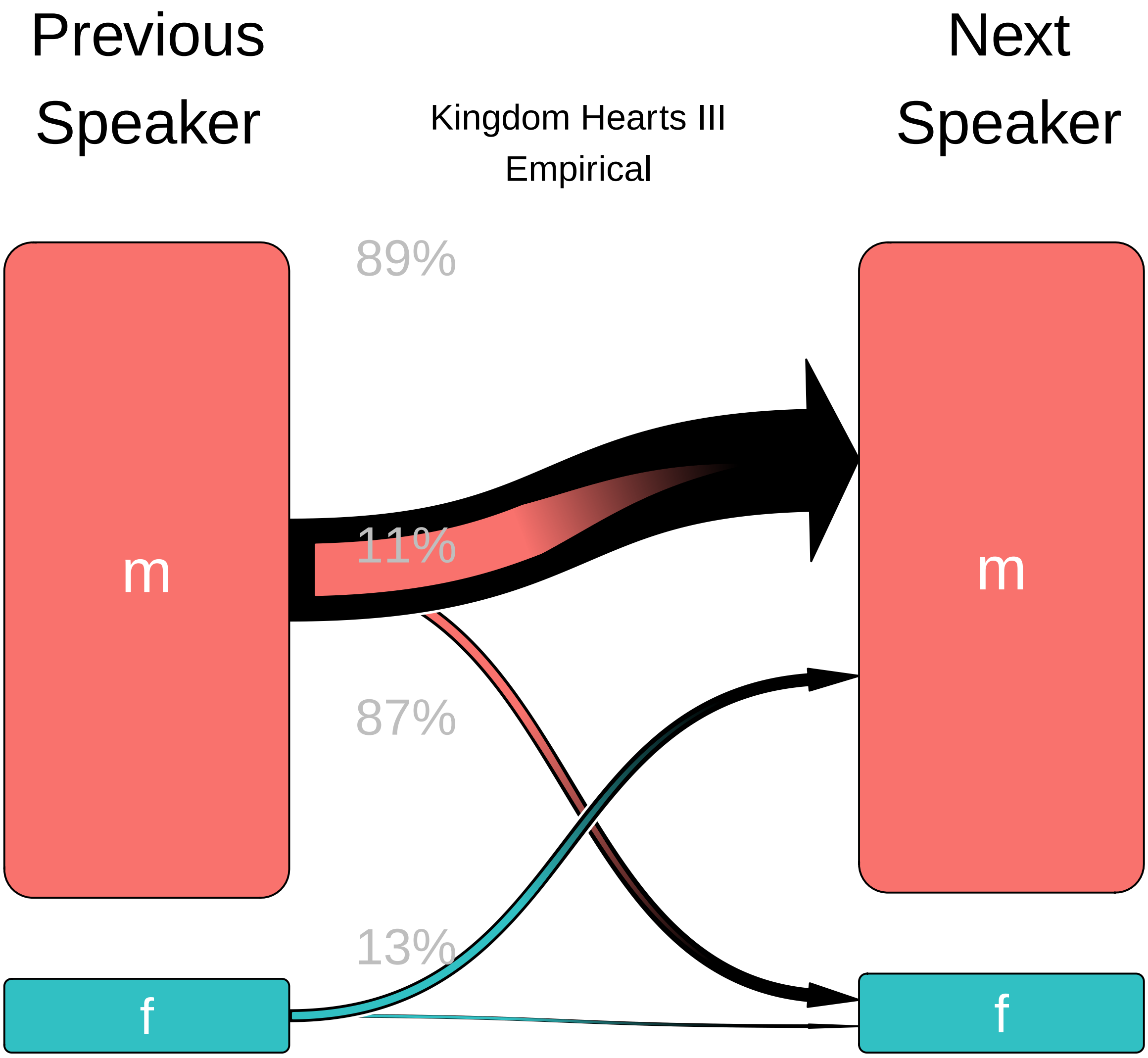

Previous
Speaker
Kingdom Hearts
Empirical
Next
Speaker
84%
m
16%
78%
m
f
22%
f



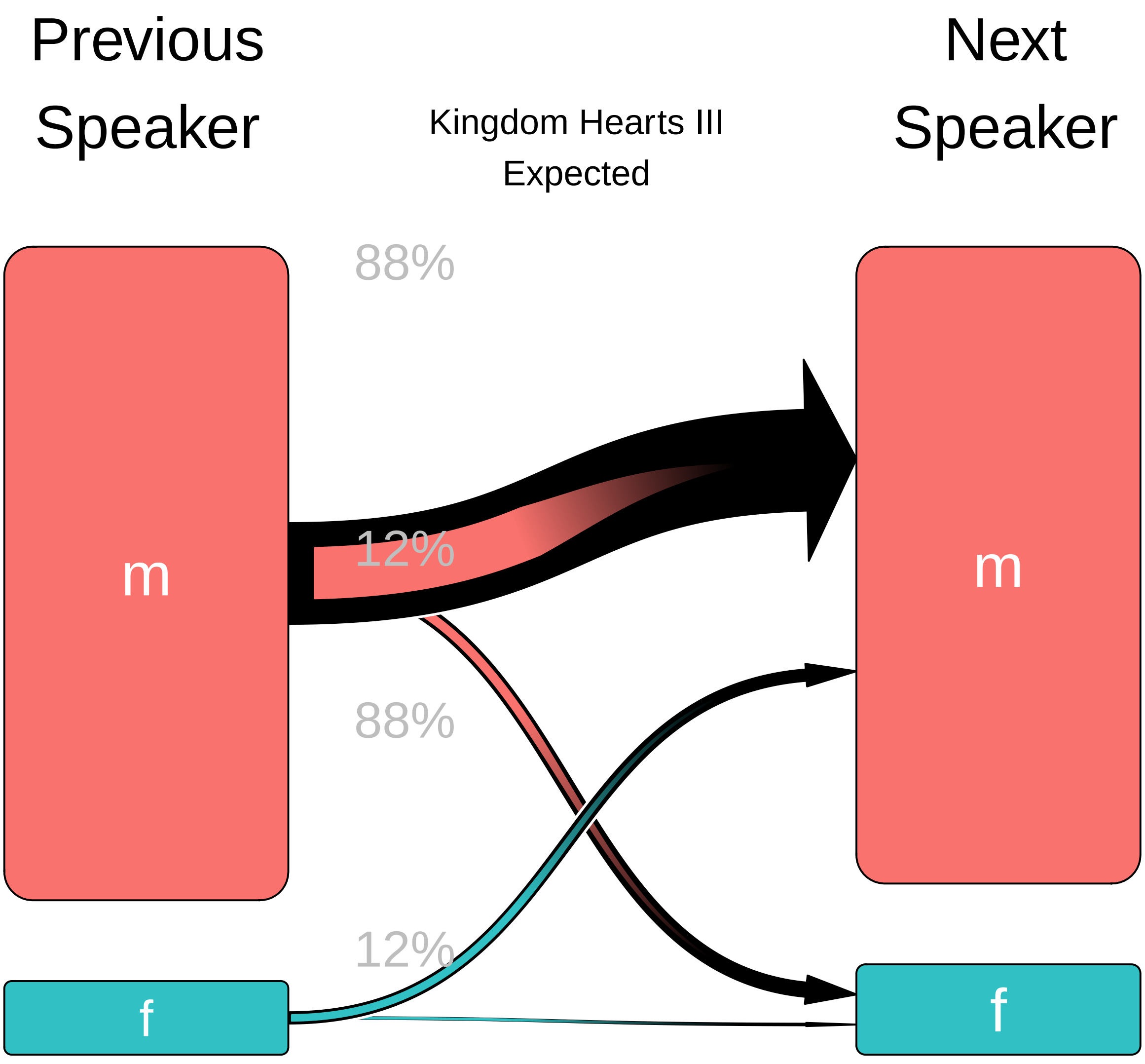
Previous
Speaker
Kingdom Hearts
Expected
Next
Speaker
82%
18%
82%
18%
m
m
f
f

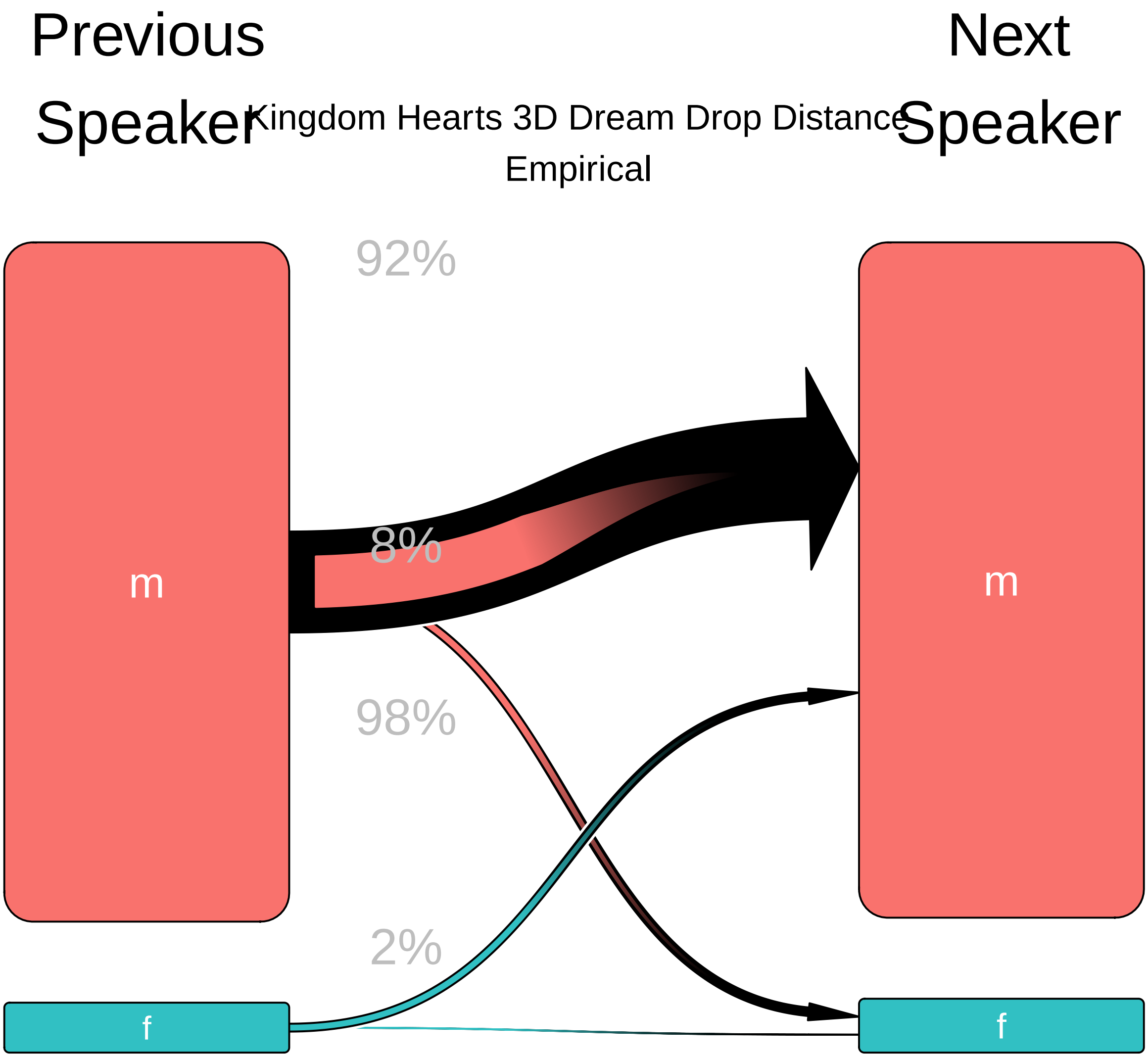

Previous
Speaker
Kingdom Hearts II
Empirical
Next
Speaker
86%
m
14%
84%
16%
f
m
f

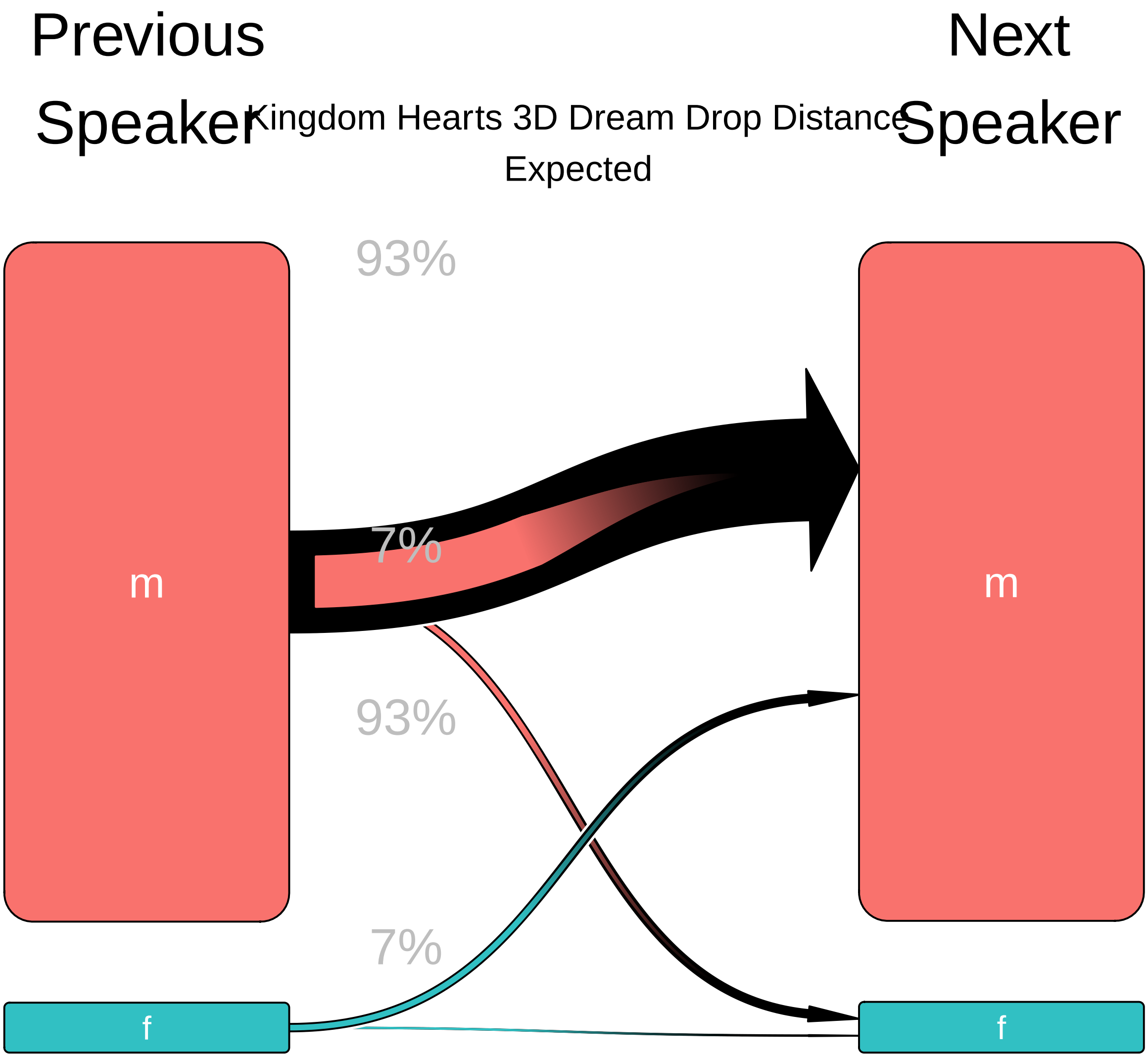

Previous
Speaker
Kingdom Hearts II
Expected
Next
Speaker
86%
m
14%
m
86%
14%
f
f

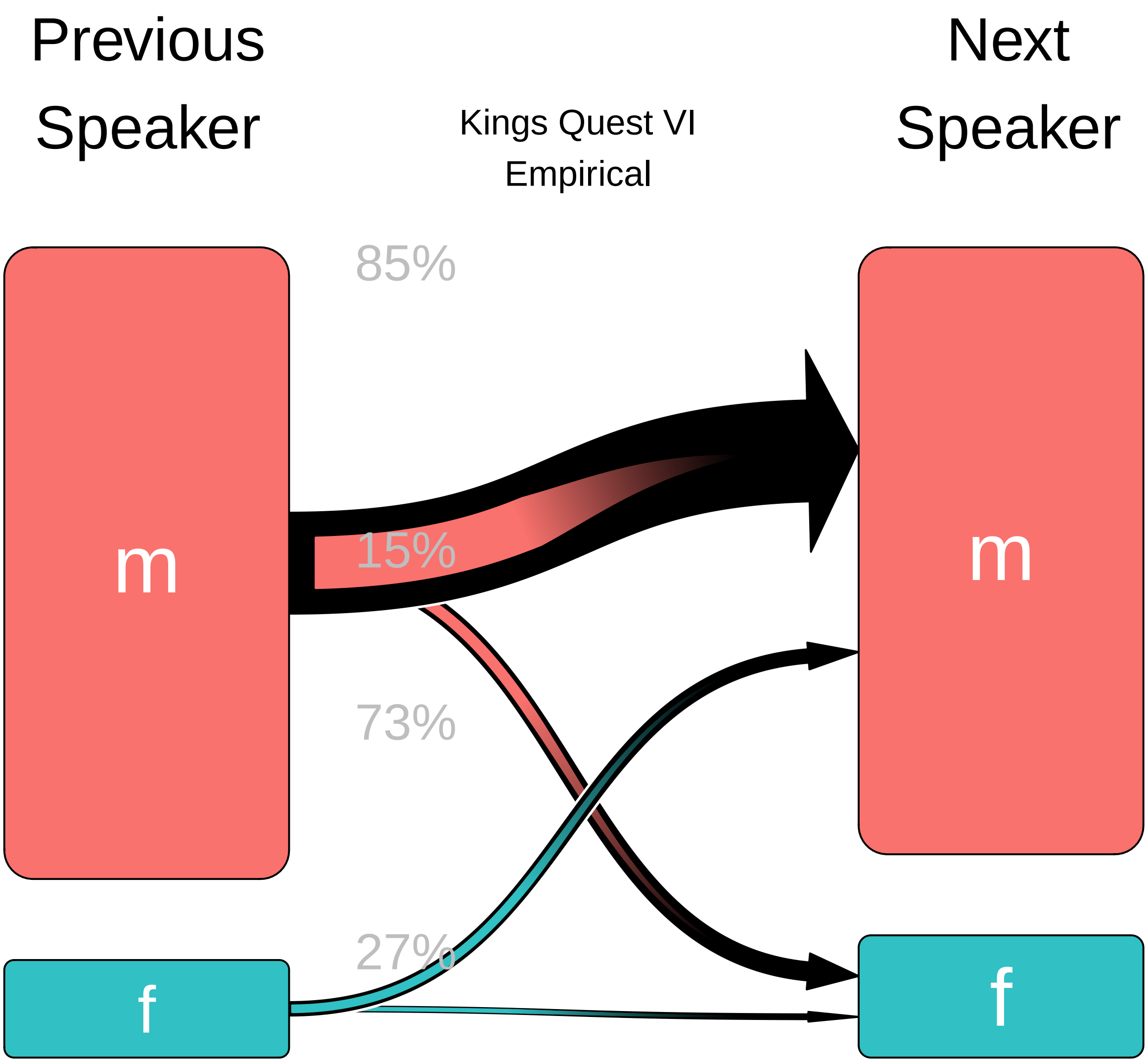


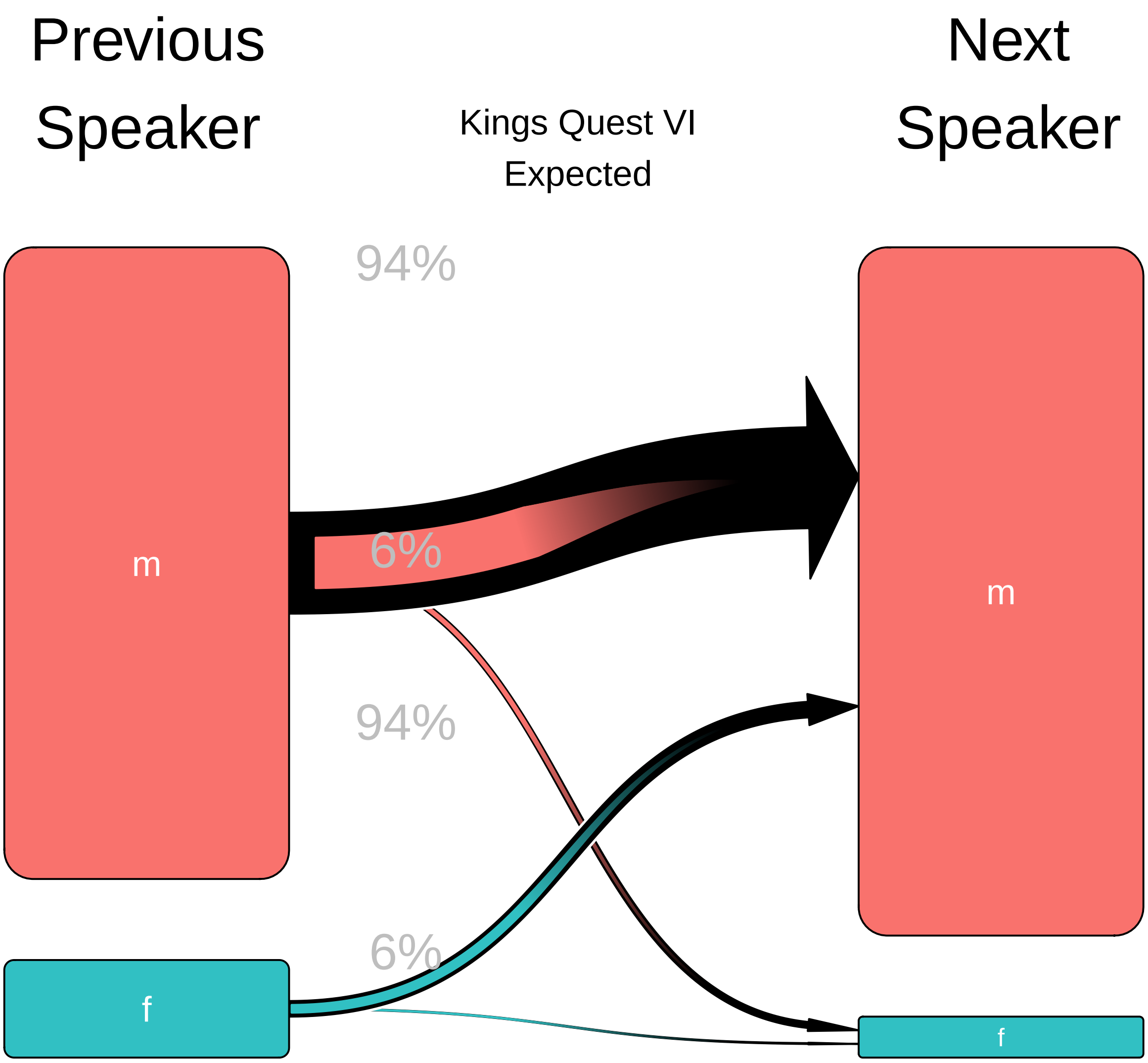
Previous Speaker
Kingdom Hearts III
Empirical
Next Speaker
89%
m
11%
87%
13%
f
m
f

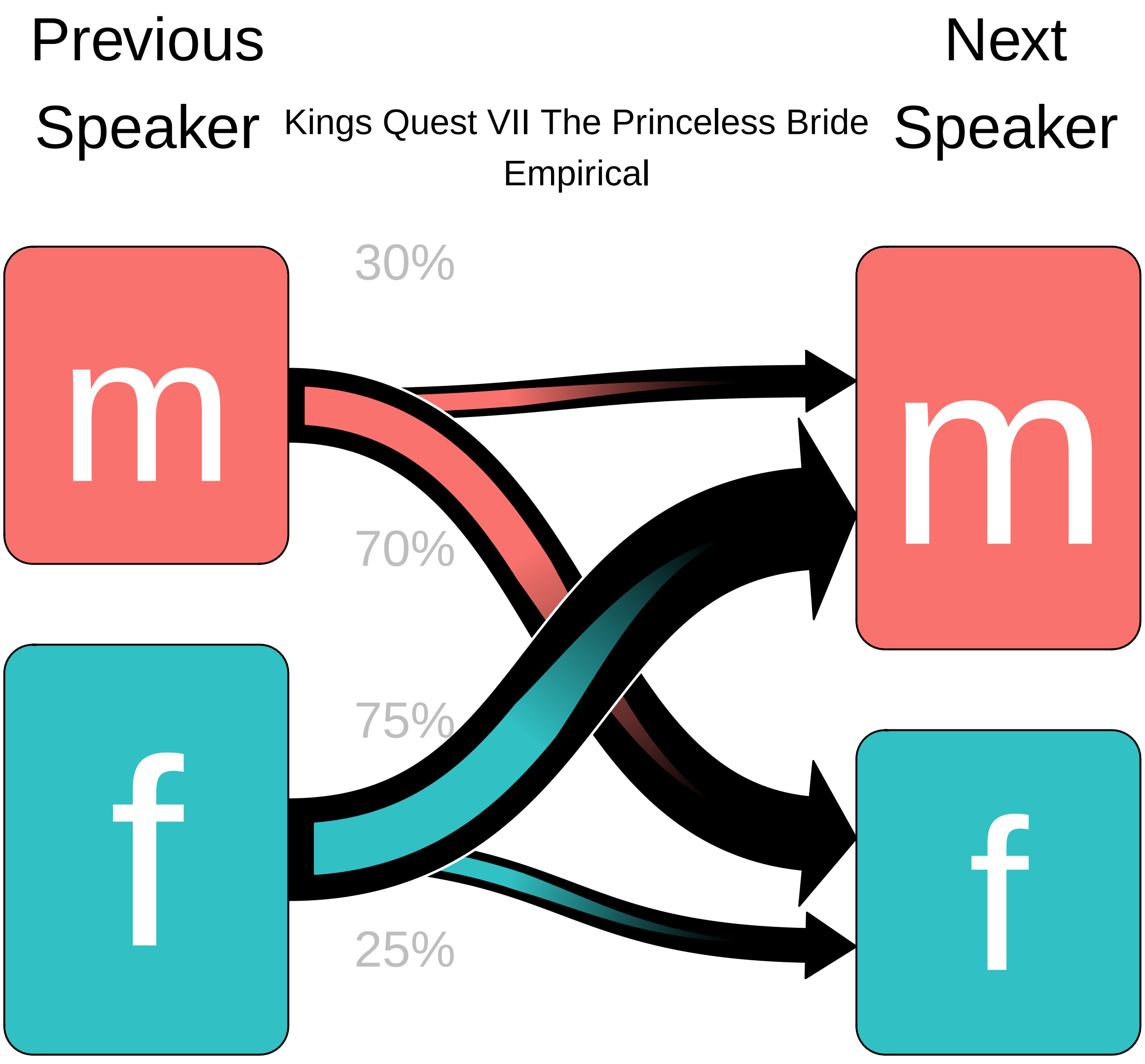

Previous
Speaker
Kingdom Hearts III
Expected
Next
Speaker
88%
12%
88%
12%
m
m
f
f

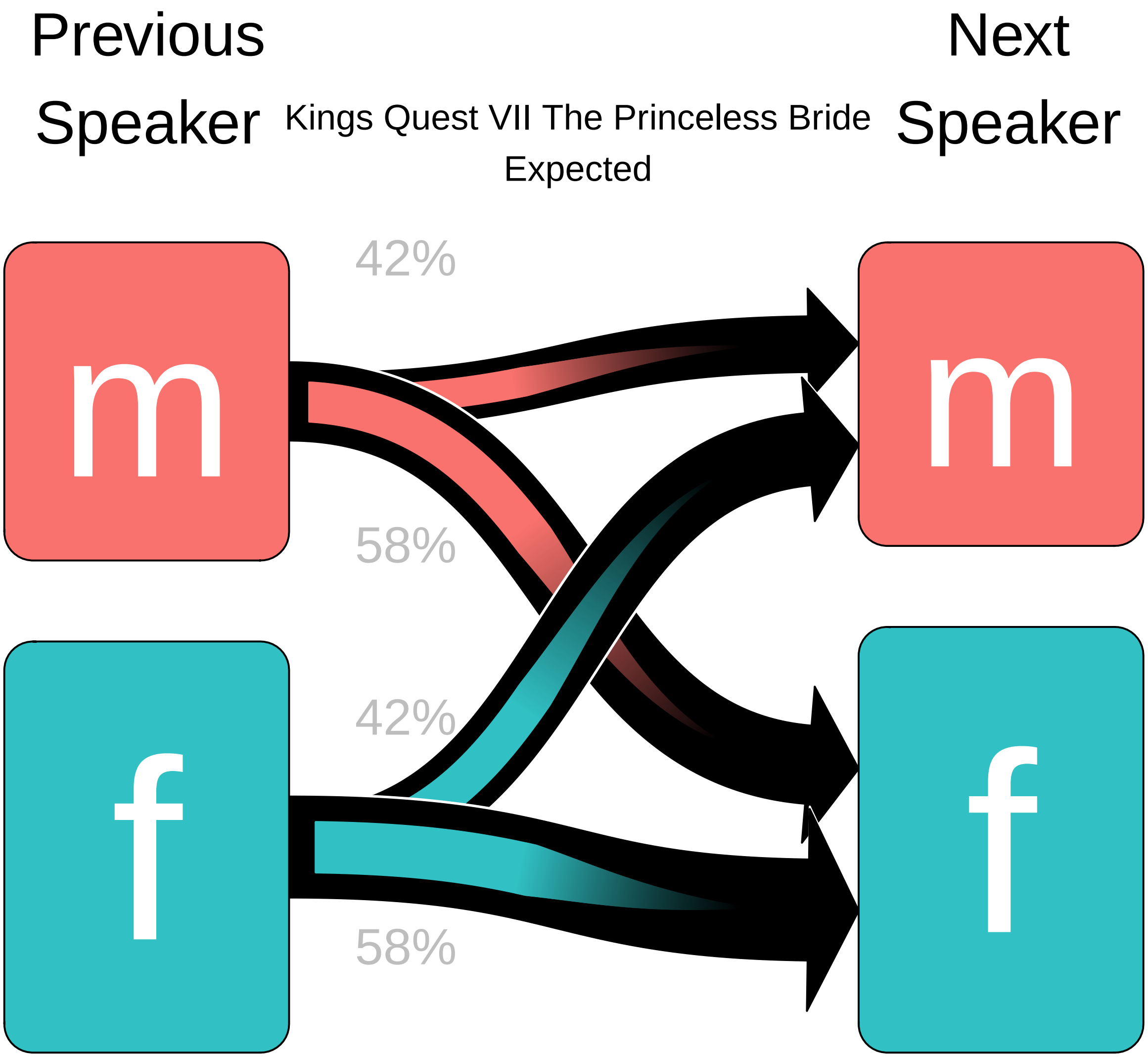

Previous Speaker
Kingdom Hearts 3D Dream Drop Distance
Empirical
Next Speaker
92%
8%
98%
2%
m
f
m
f


Previous Speaker
Next Speaker
Kingdom Hearts 3D Dream Drop Distance
Expected
93%
7%
93%
7%
m
m
f
f

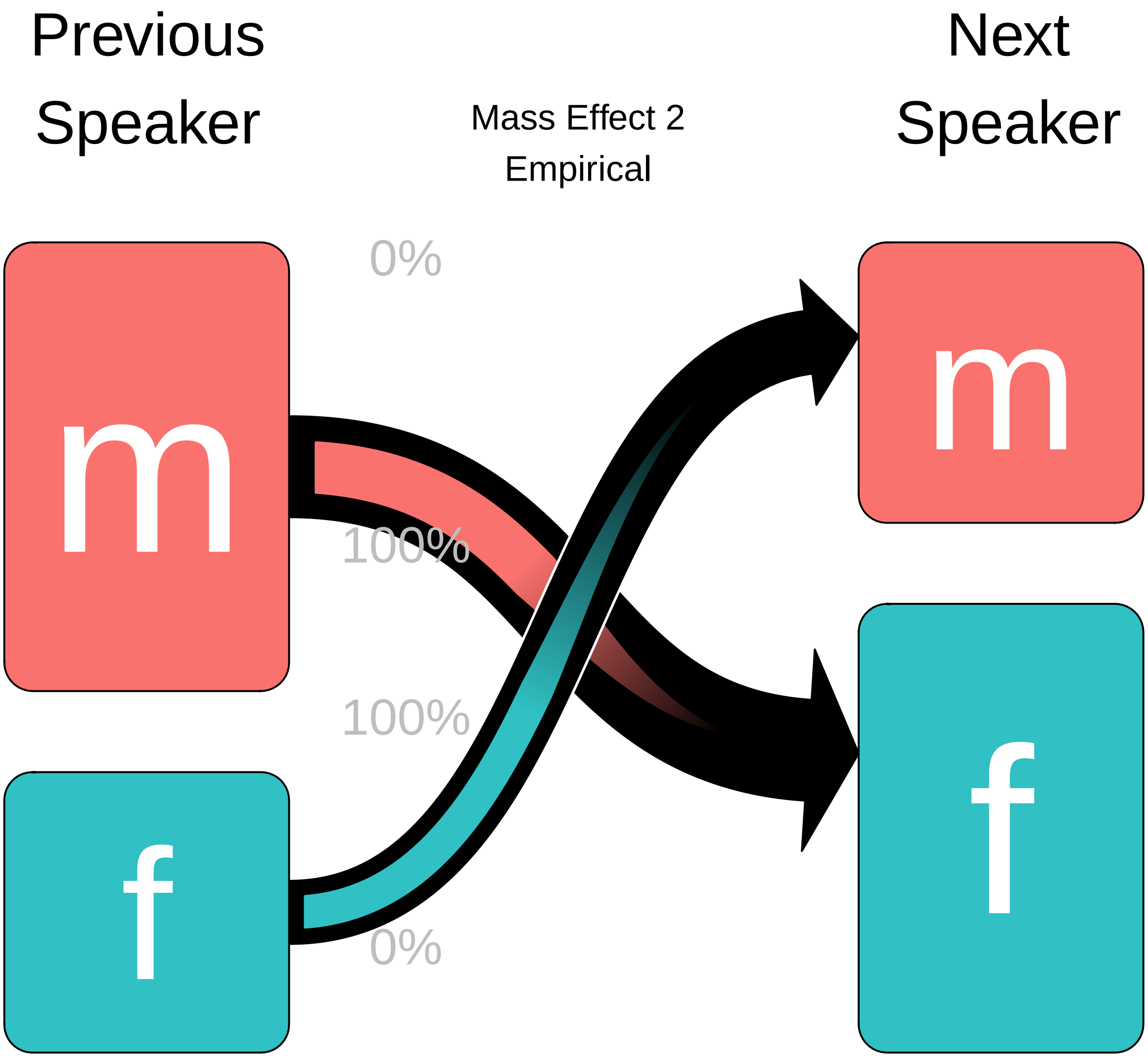

Previous
Speaker
Kings Quest VI
Empirical
Next
Speaker
85%
15%
73%
27%
m
f
m
f

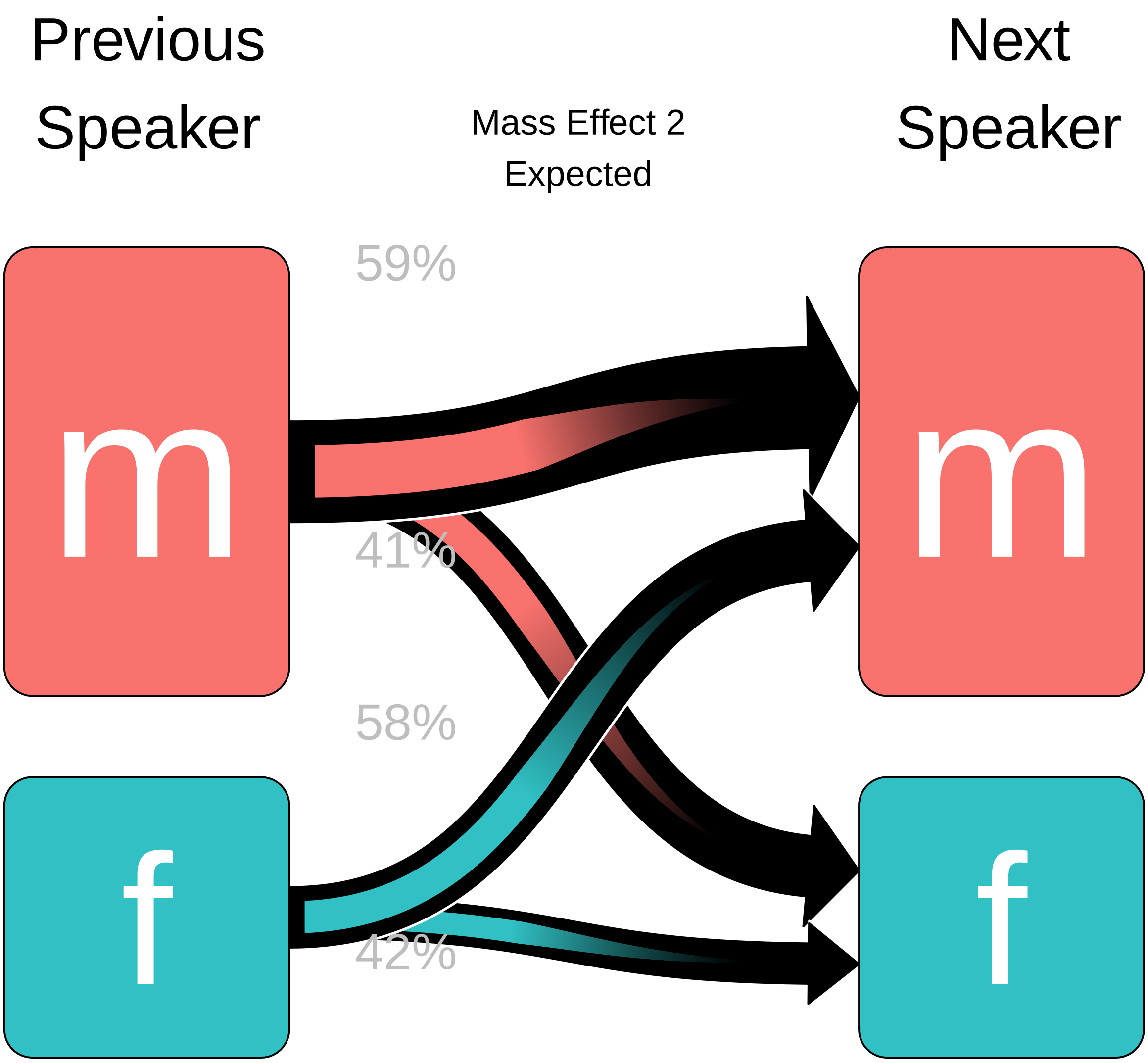

Previous Speaker
Kings Quest VI Expected
Next Speaker
94%
6%
94%
6%
m
f
m
f

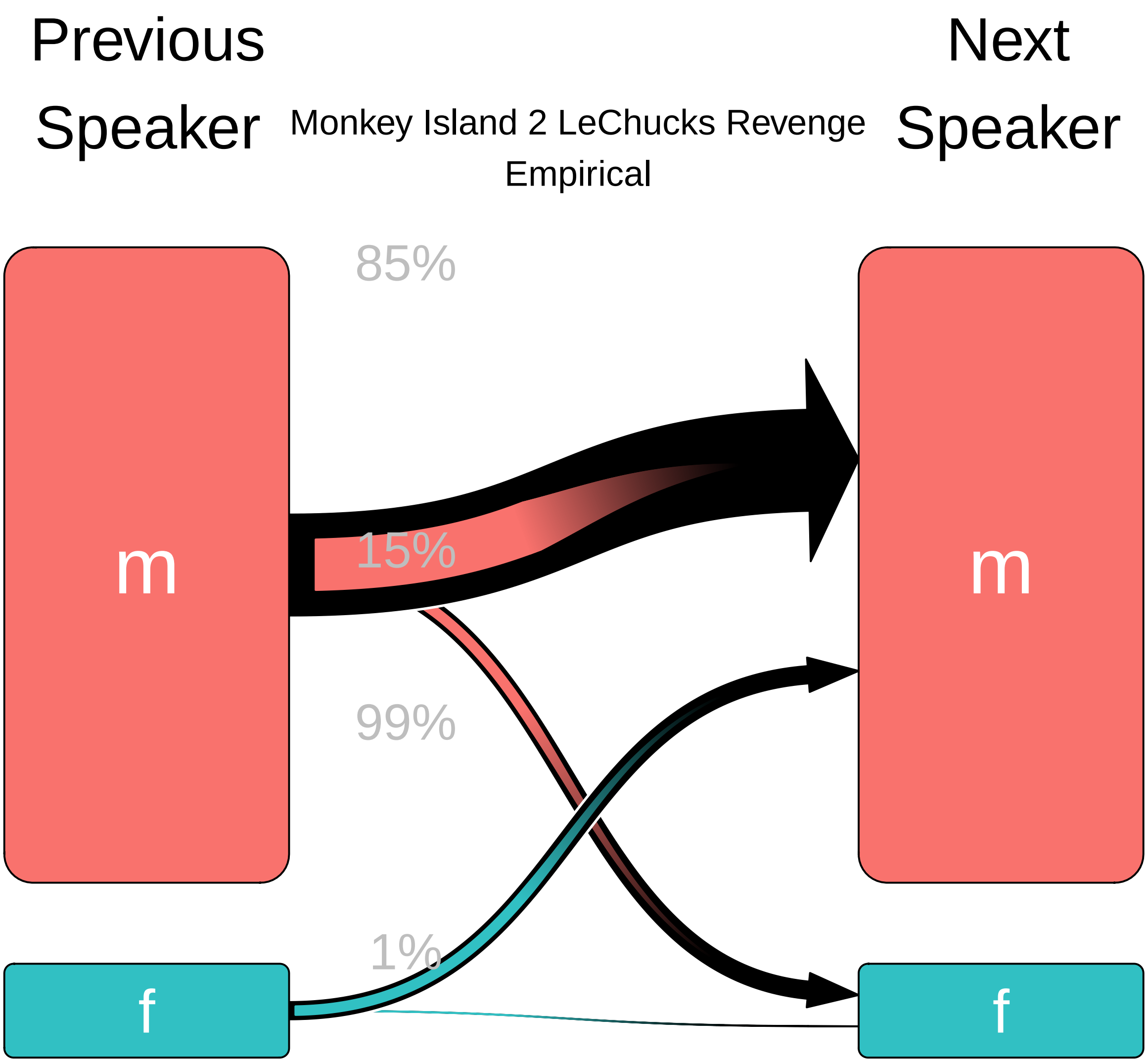

Previous Speaker
Kings Quest VII The Princeless Bride
Empirical
Next Speaker
m
f
30%
70%
75%
25%
m
f

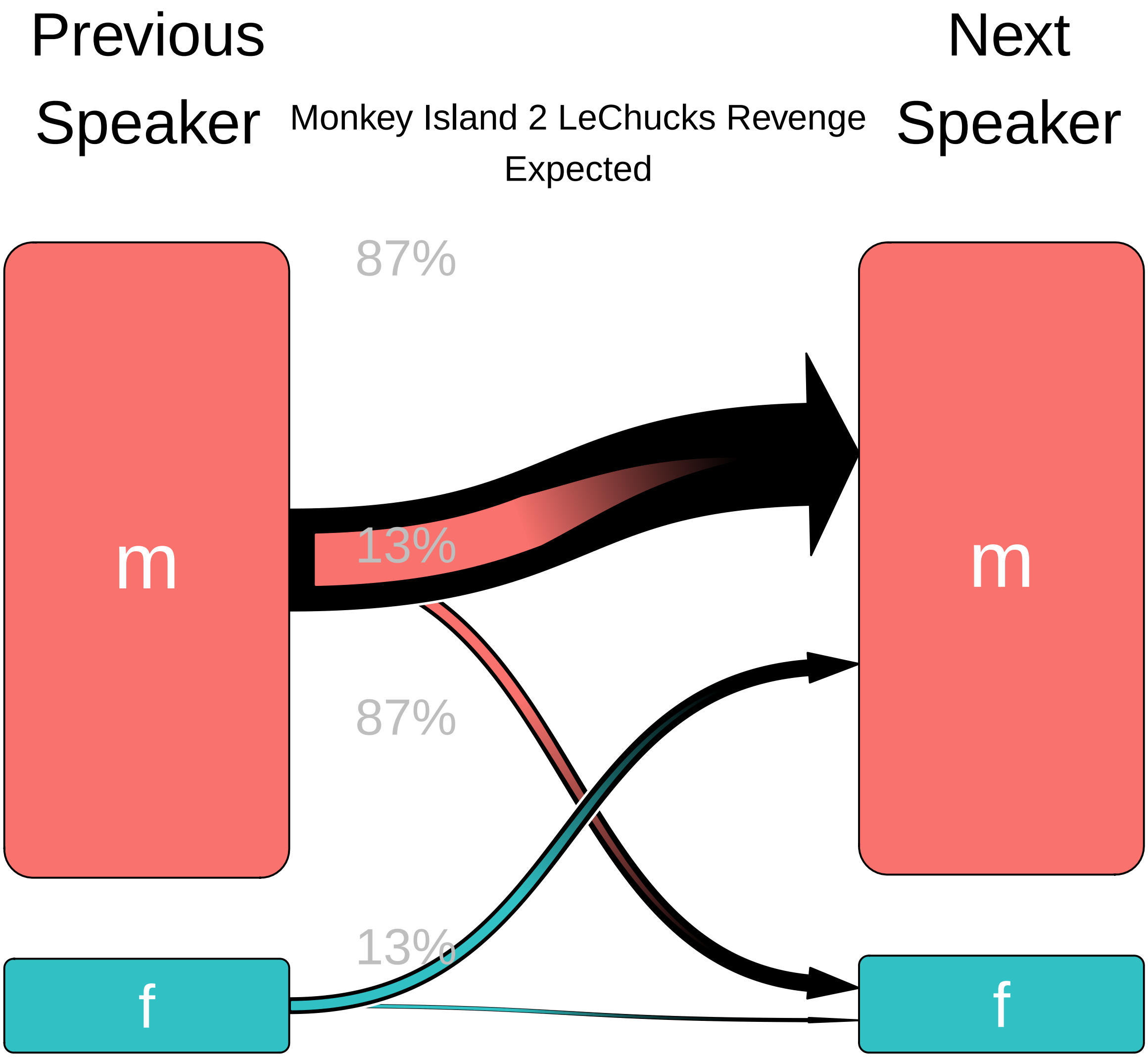


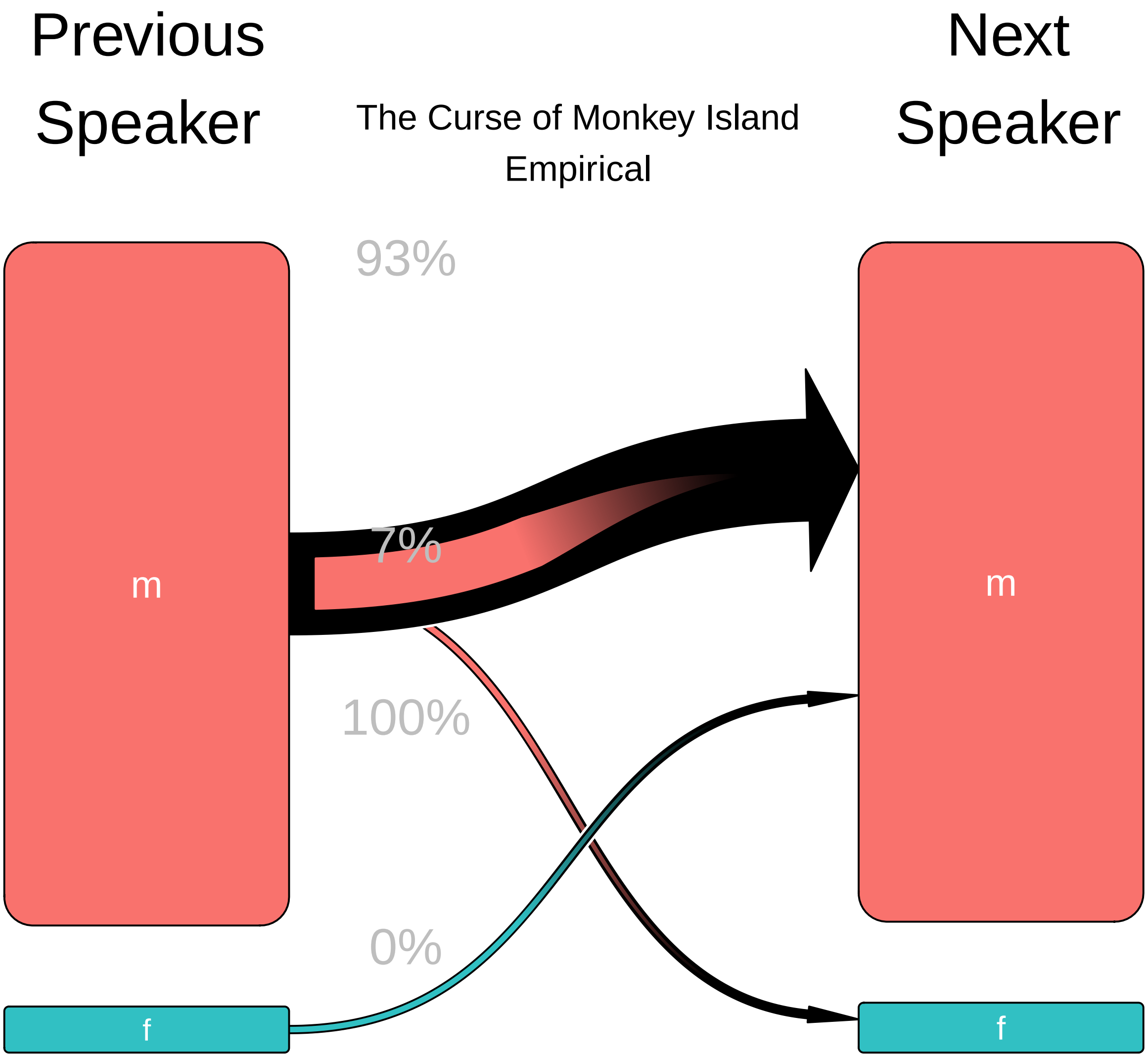
Previous
Speaker
Kings Quest VII The Princeless Bride
Expected
Next
Speaker
42%
58%
42%
58%
m
m
f
f

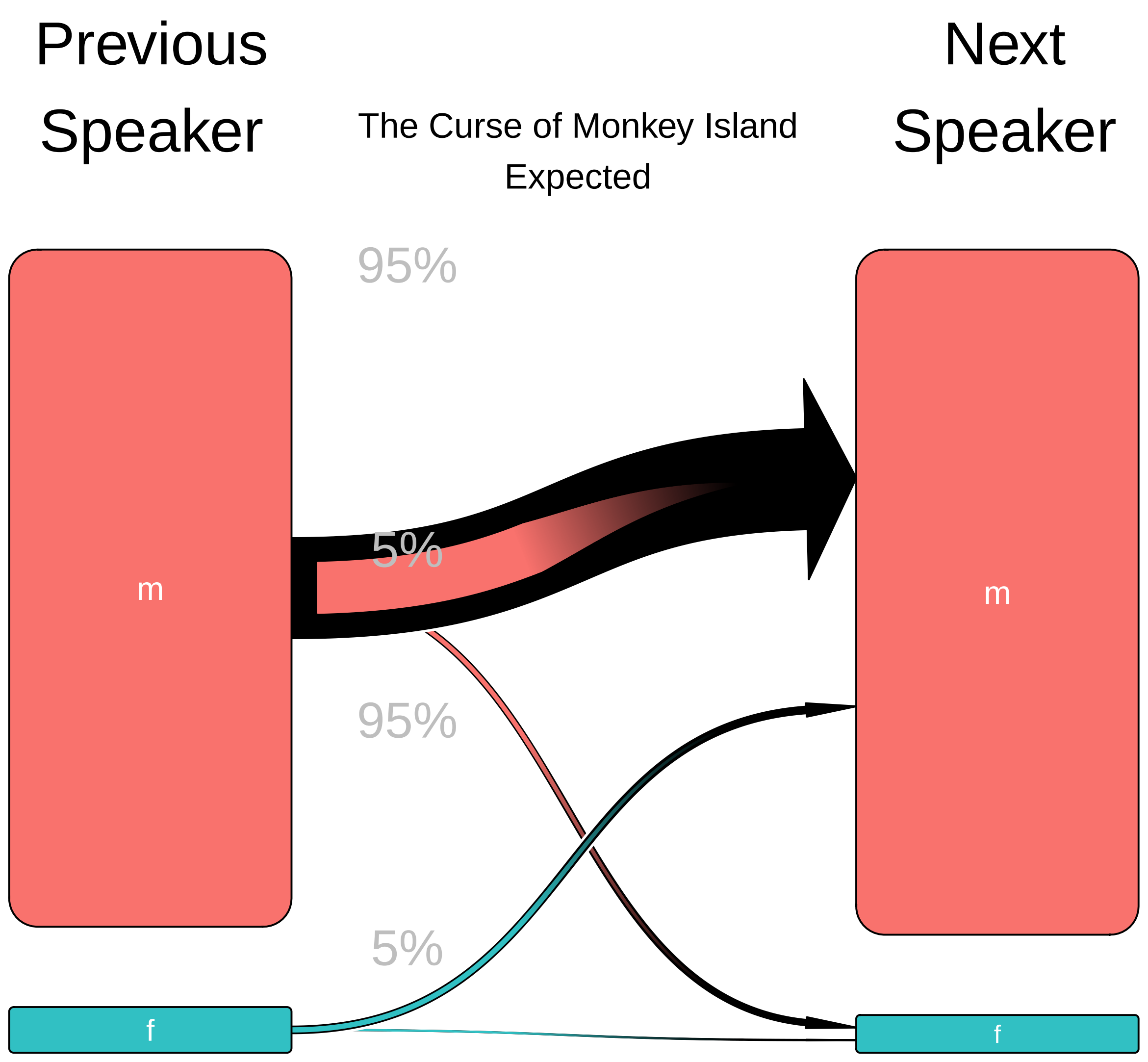

Previous
Speaker
Mass Effect 2
Empirical
Next
Speaker
m
f
m
f
0%
100%
100%
0%

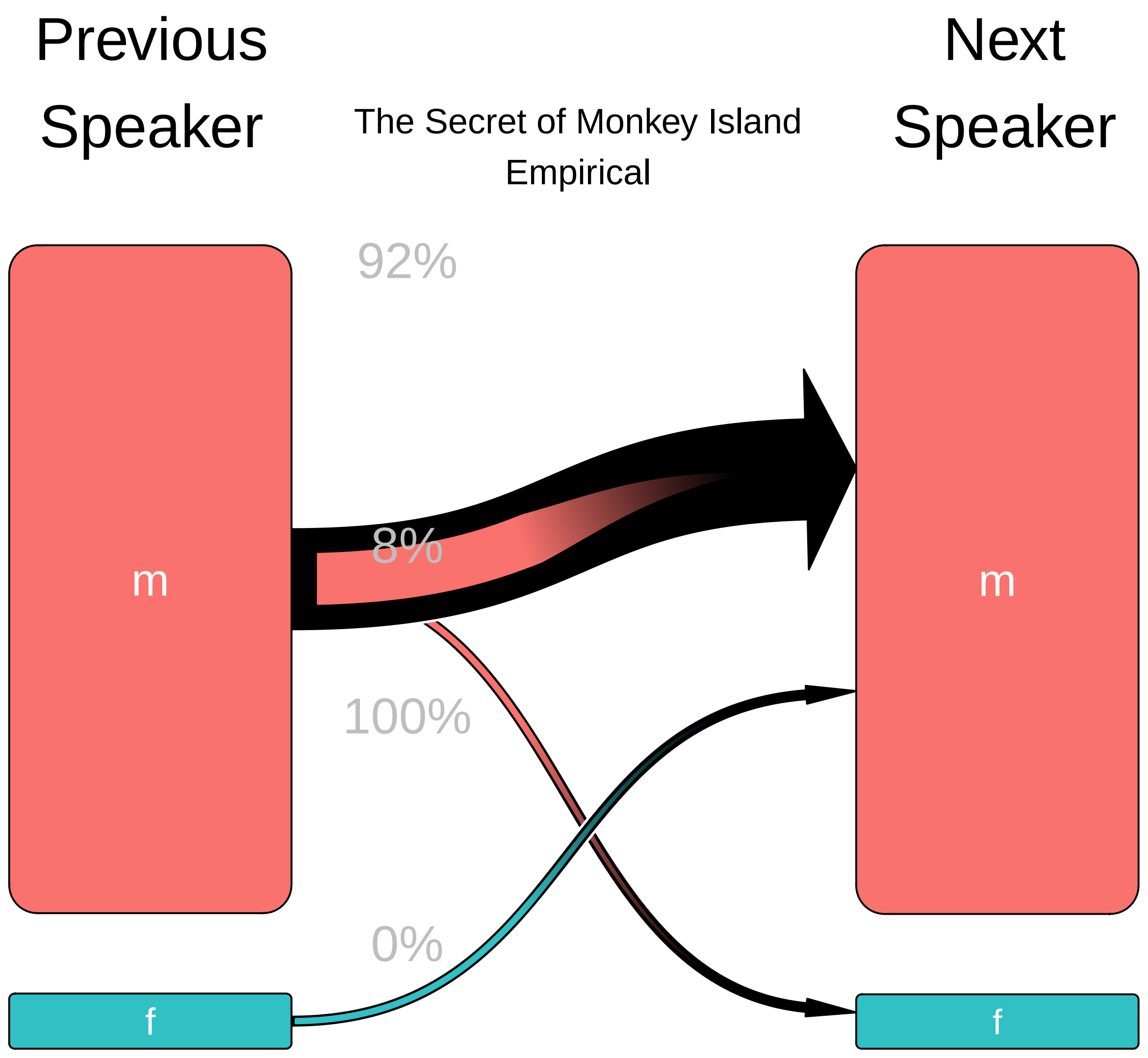

Previous
Speaker
Mass Effect 2
Expected
Next
Speaker
59%
m
m
41%
58%
f
f
42%

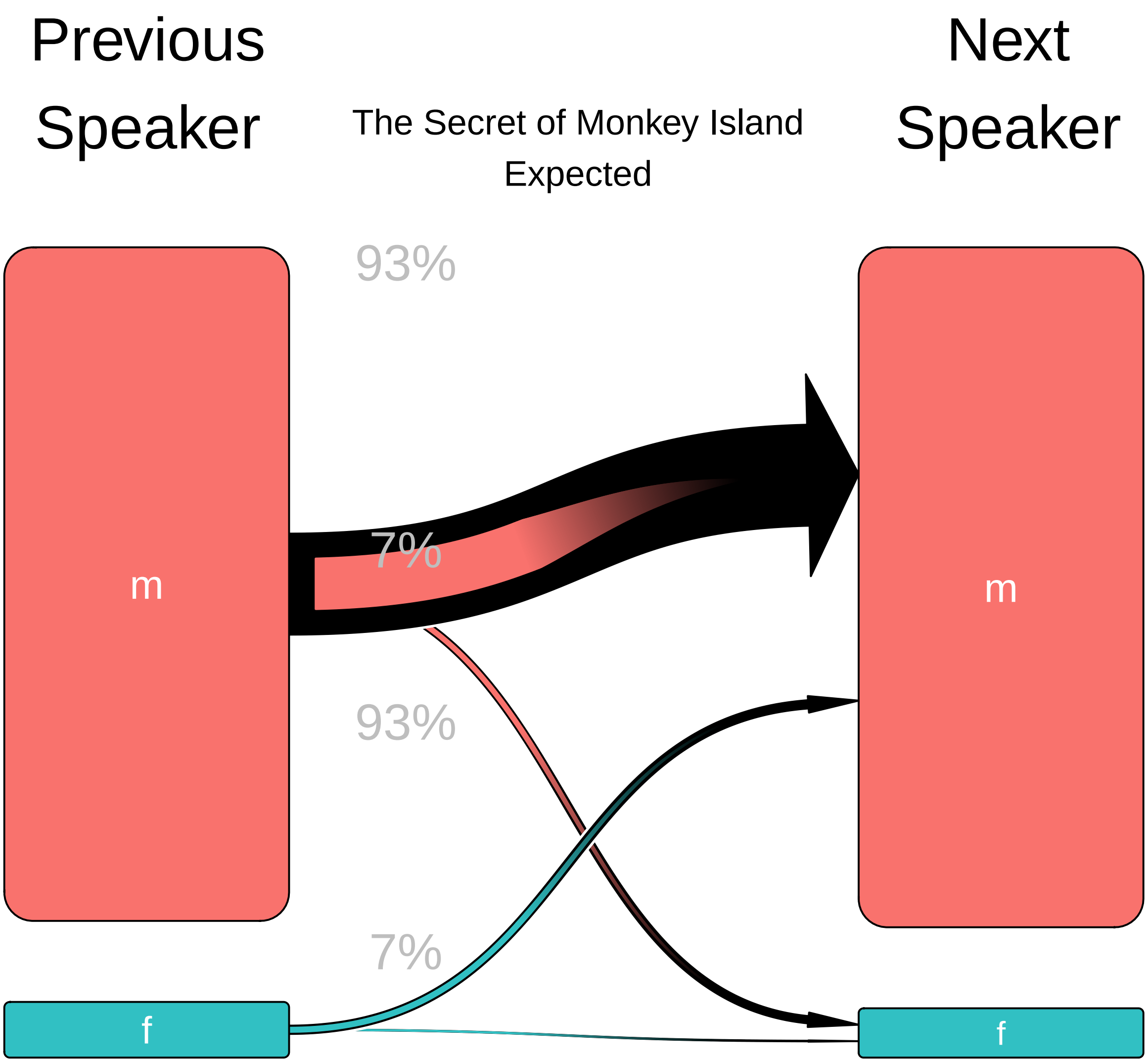

Previous Speaker
Next Speaker
Monkey Island 2 LeChucks Revenge
Empirical
85%
15%
99%
1%
m
m
f
f

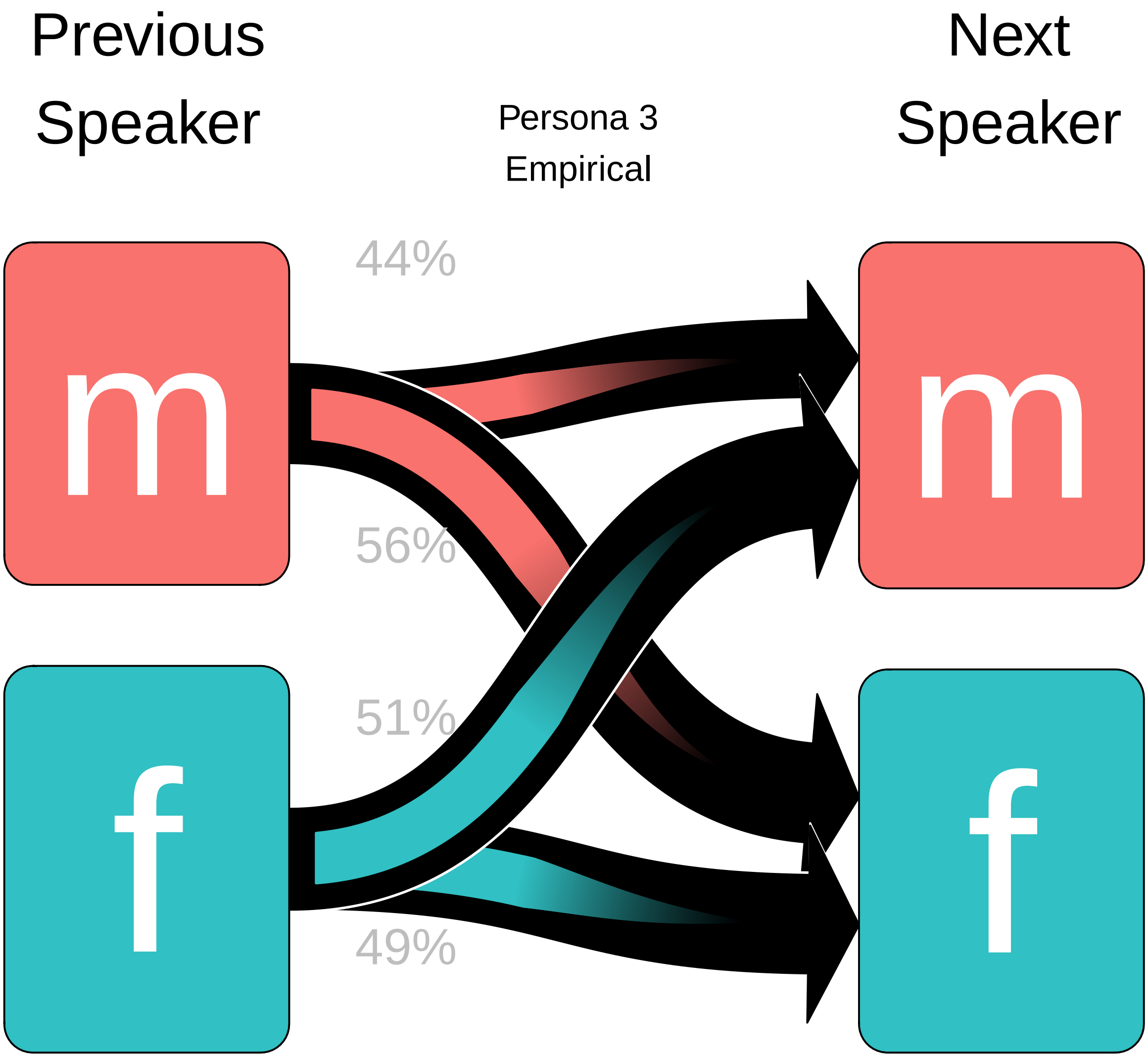

Previous
Speaker
Monkey Island 2 LeChucks Revenge
Expected
Next
Speaker
87%
13%
87%
13%
m
m
f
f

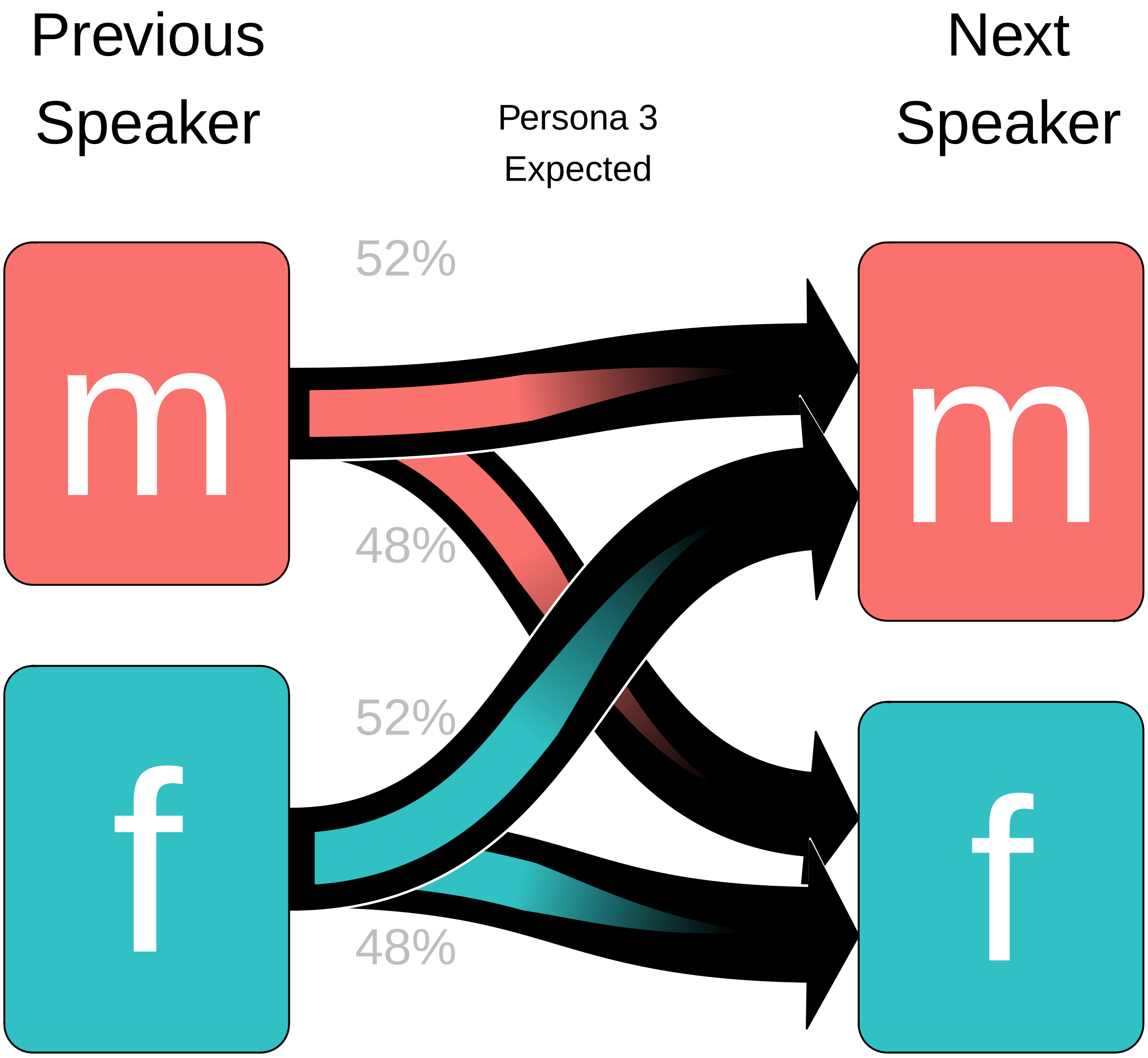

Previous Speaker
The Curse of Monkey Island
Empirical
Next Speaker
93%
7%
100%
0%
m
f
m
f

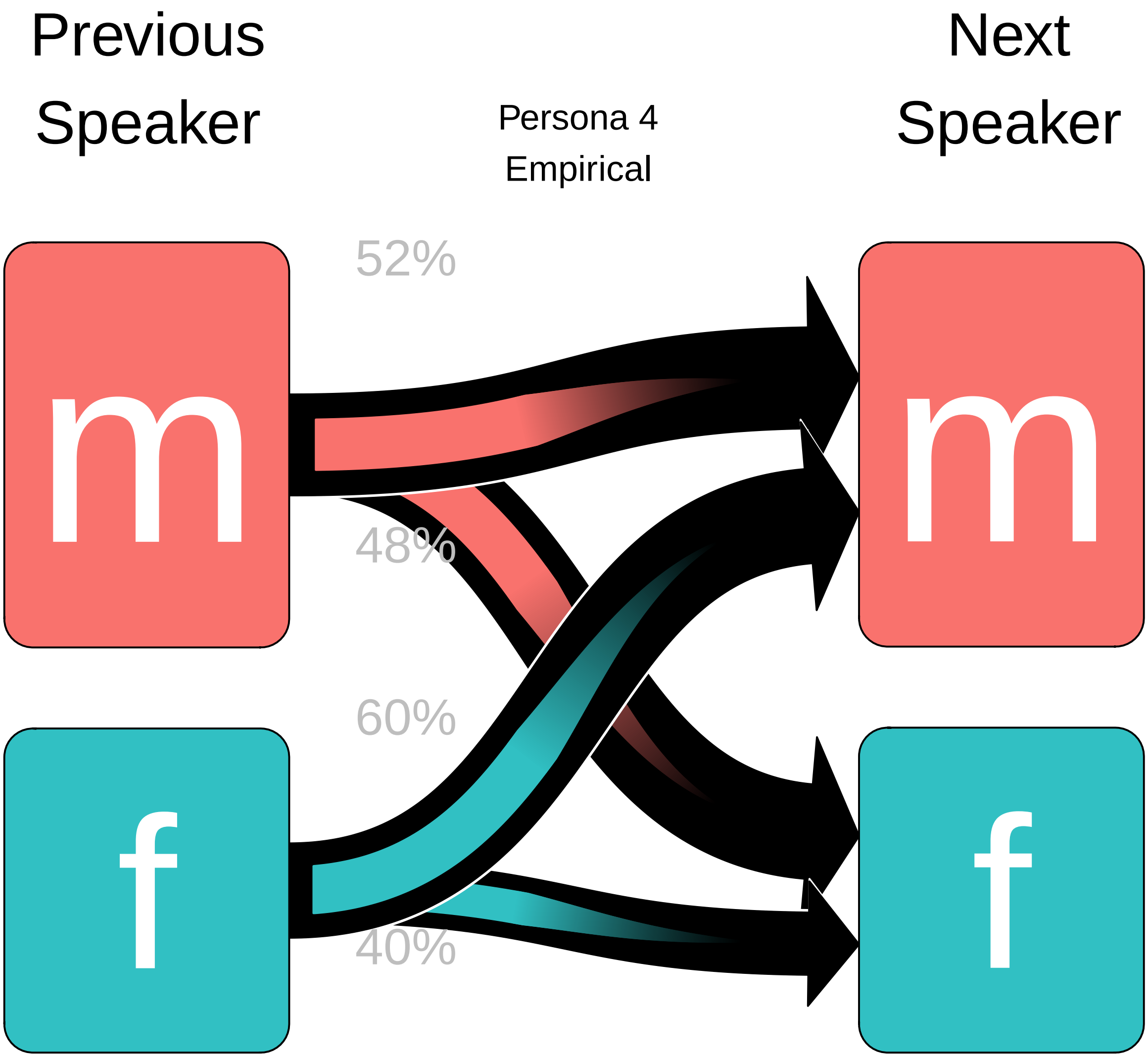

Previous Speaker
The Curse of Monkey Island
Expected
Next Speaker
95%
5%
95%
5%
m
f
m
f

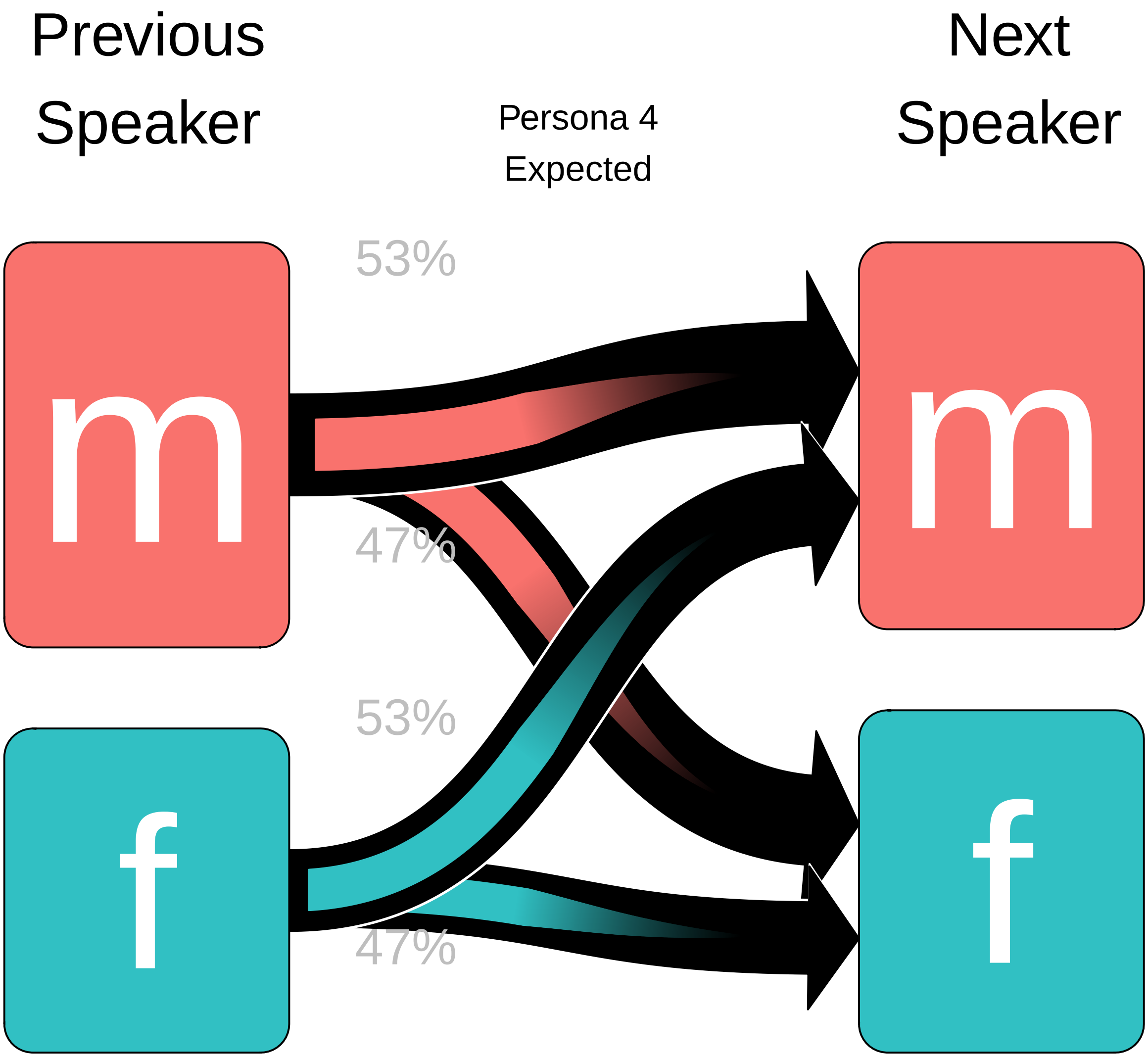

Previous Speaker
The Secret of Monkey Island
Empirical
Next Speaker
92%
m
8%
m
100%
0%
f
f


Previous Speaker
The Secret of Monkey Island
Expected
Next Speaker
93%
7%
93%
7%
m
f
m
f

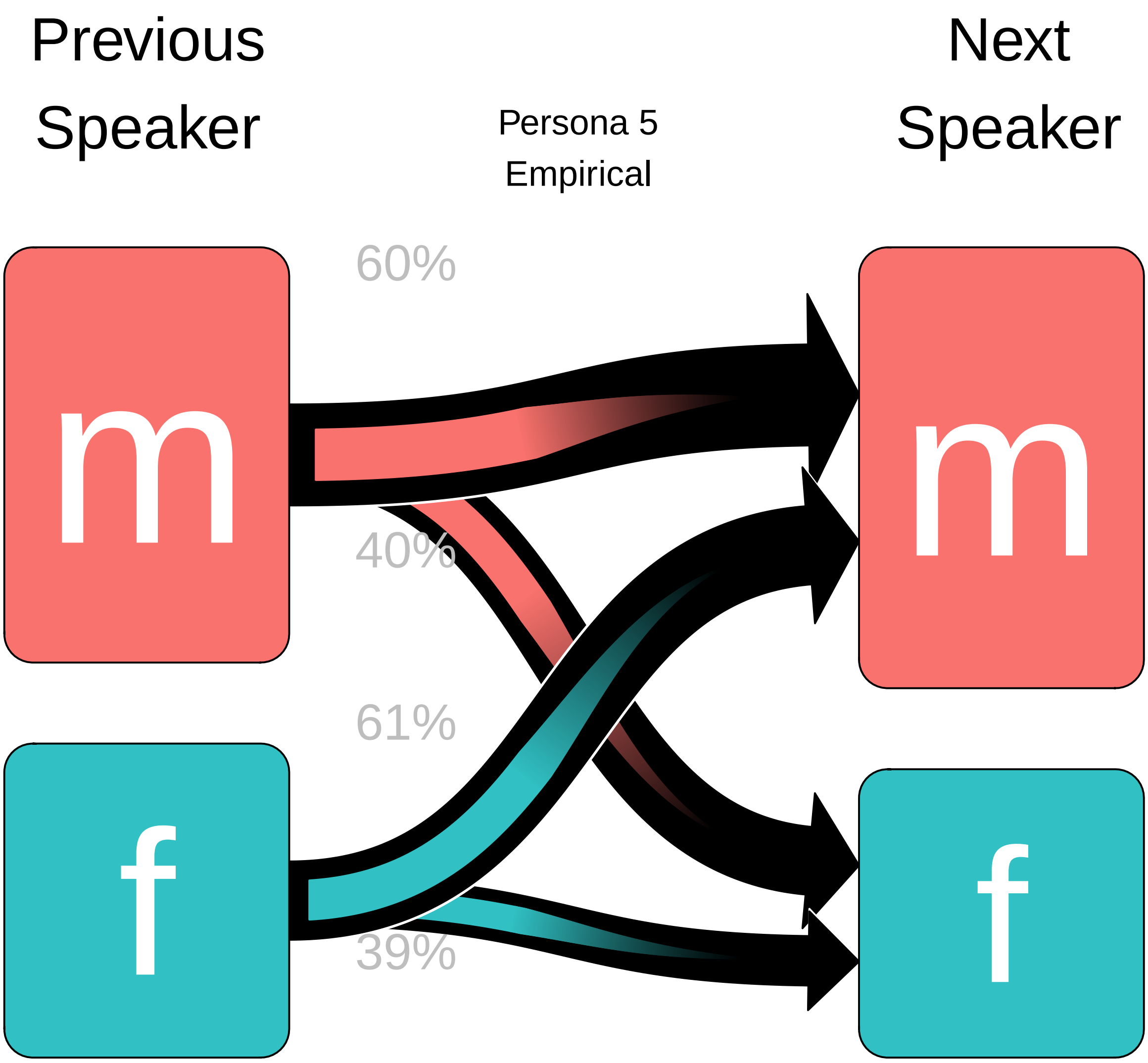

Previous Speaker
Persona 3 Empirical
Next Speaker
m
f
44%
56%
51%
49%
m
f


Previous
Speaker
Persona 3
Expected
Next
Speaker
m
52%
48%
f
52%
48%
m
f

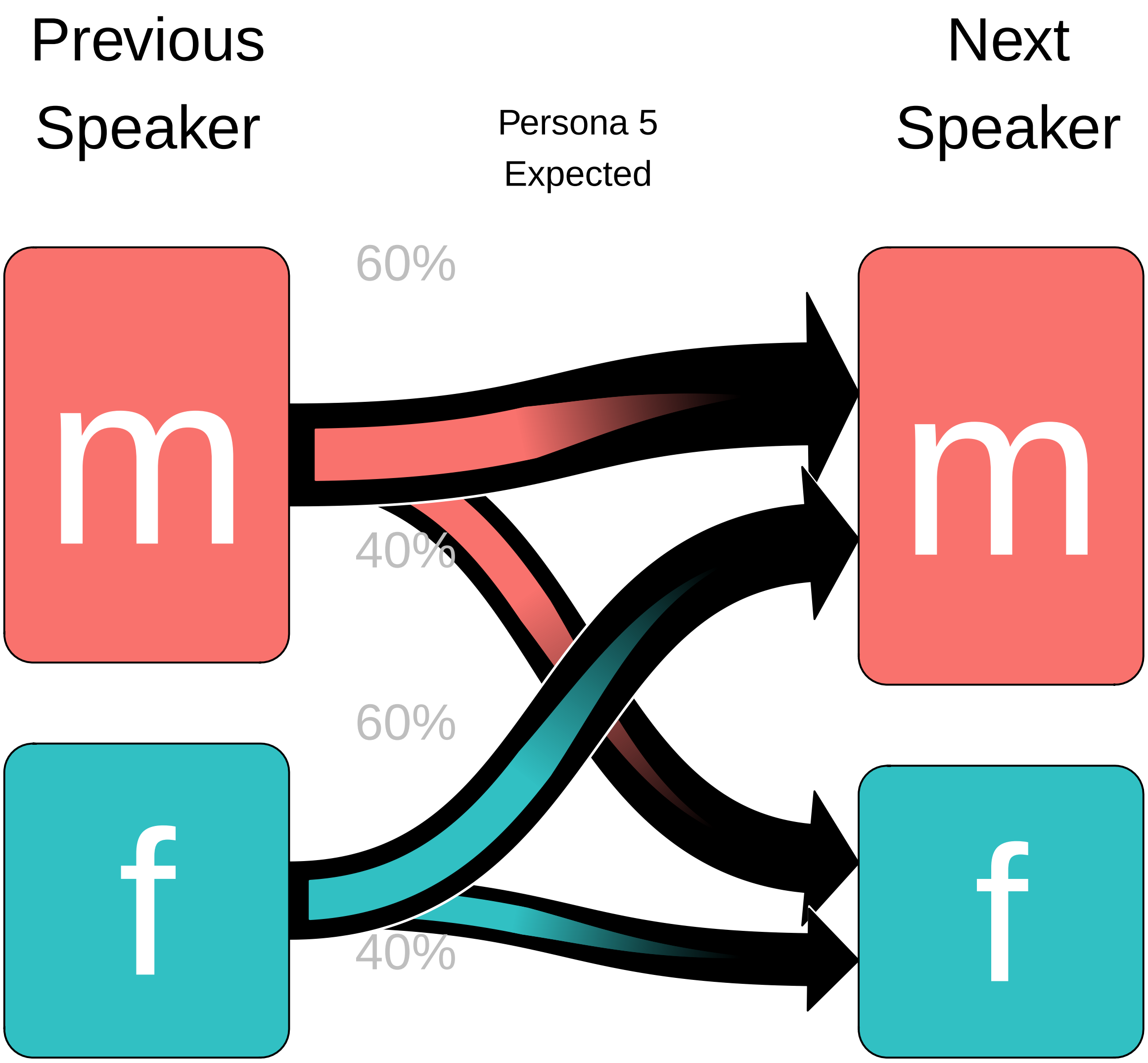

Previous Speaker
Persona 4 Empirical
Next Speaker
52%
m
48%
60%
f
40%
m
f

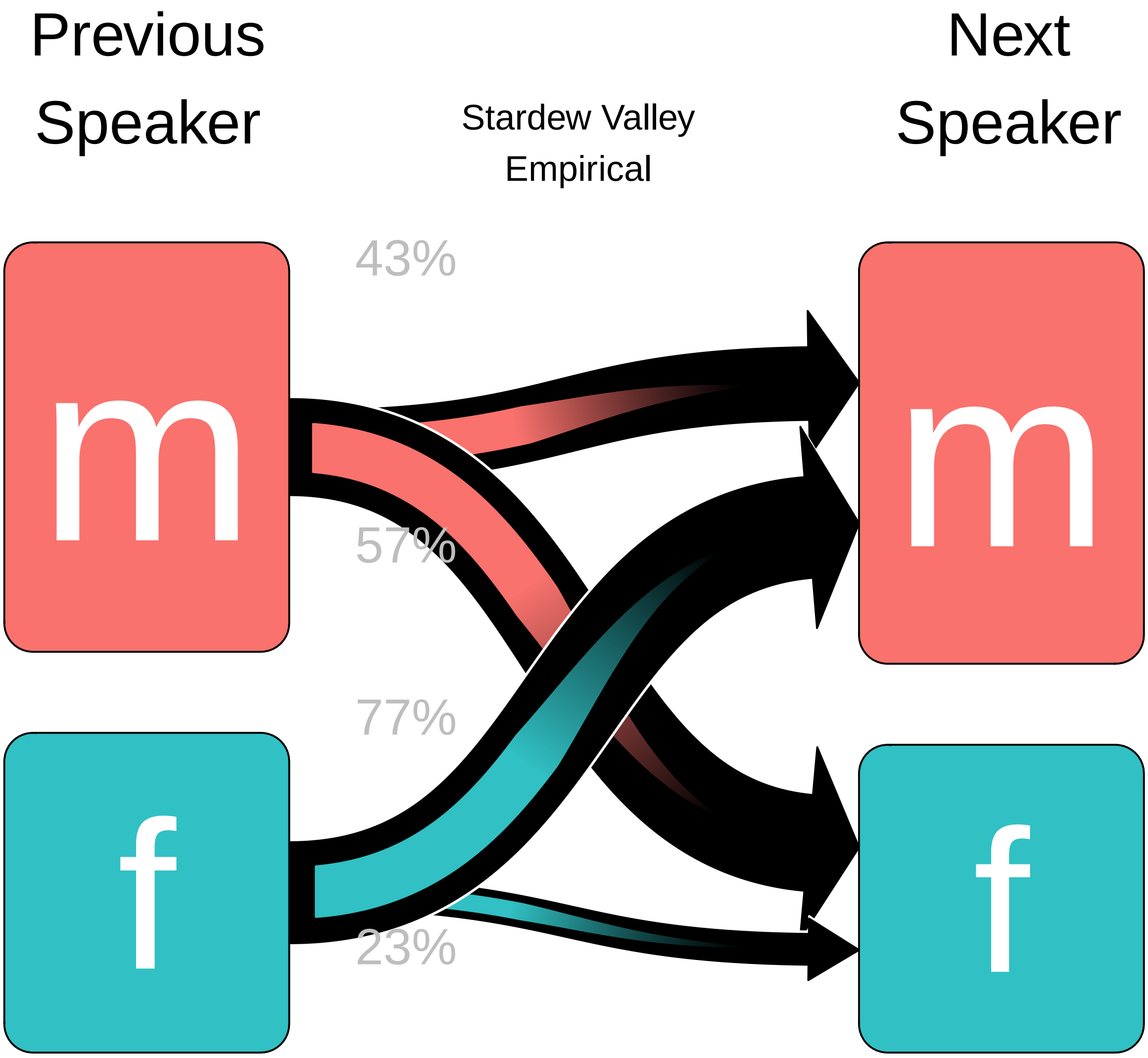


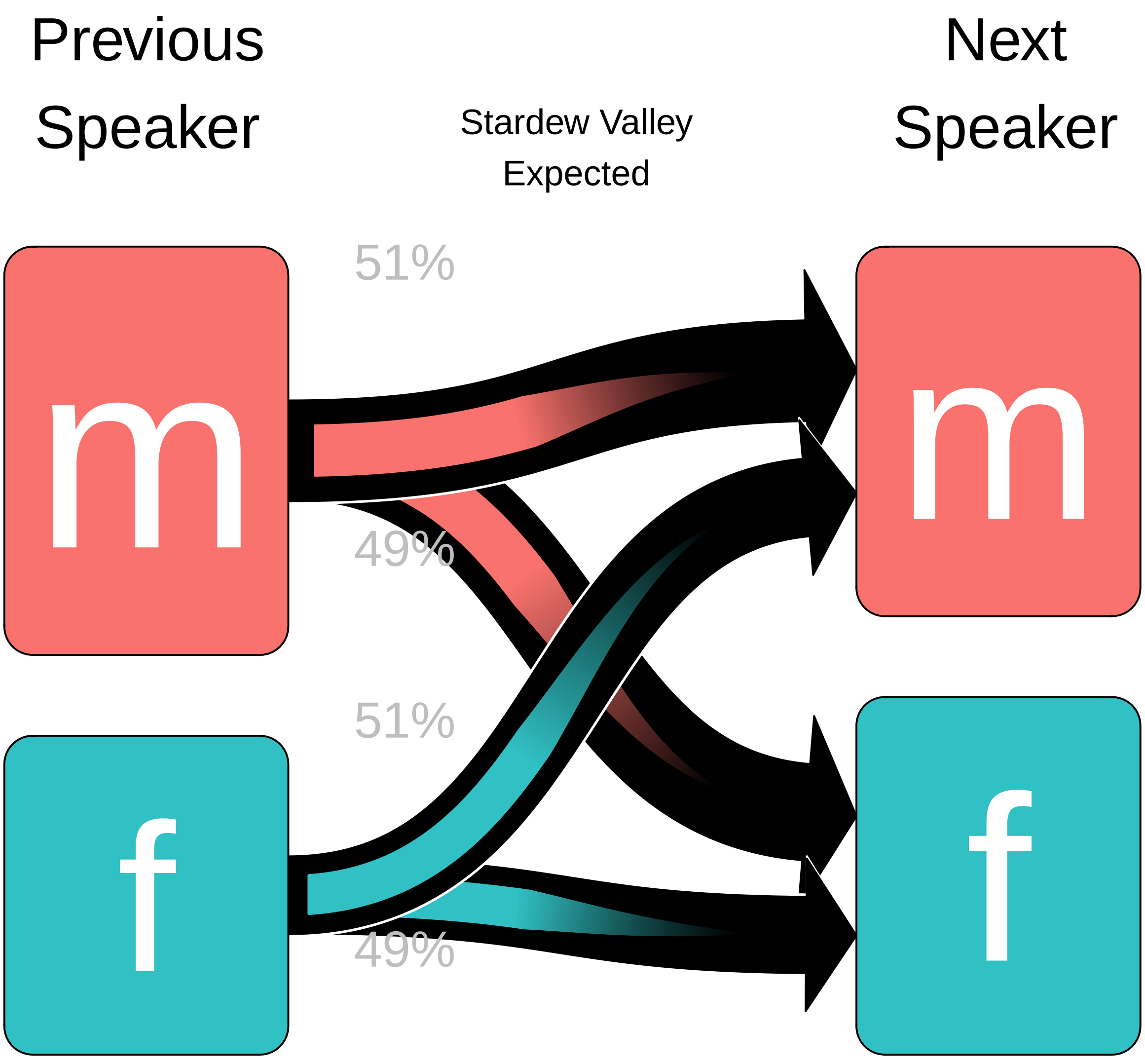
Previous
Speaker
Persona 4
Expected
Next
Speaker
m
f
53%
47%
53%
47%
m
f

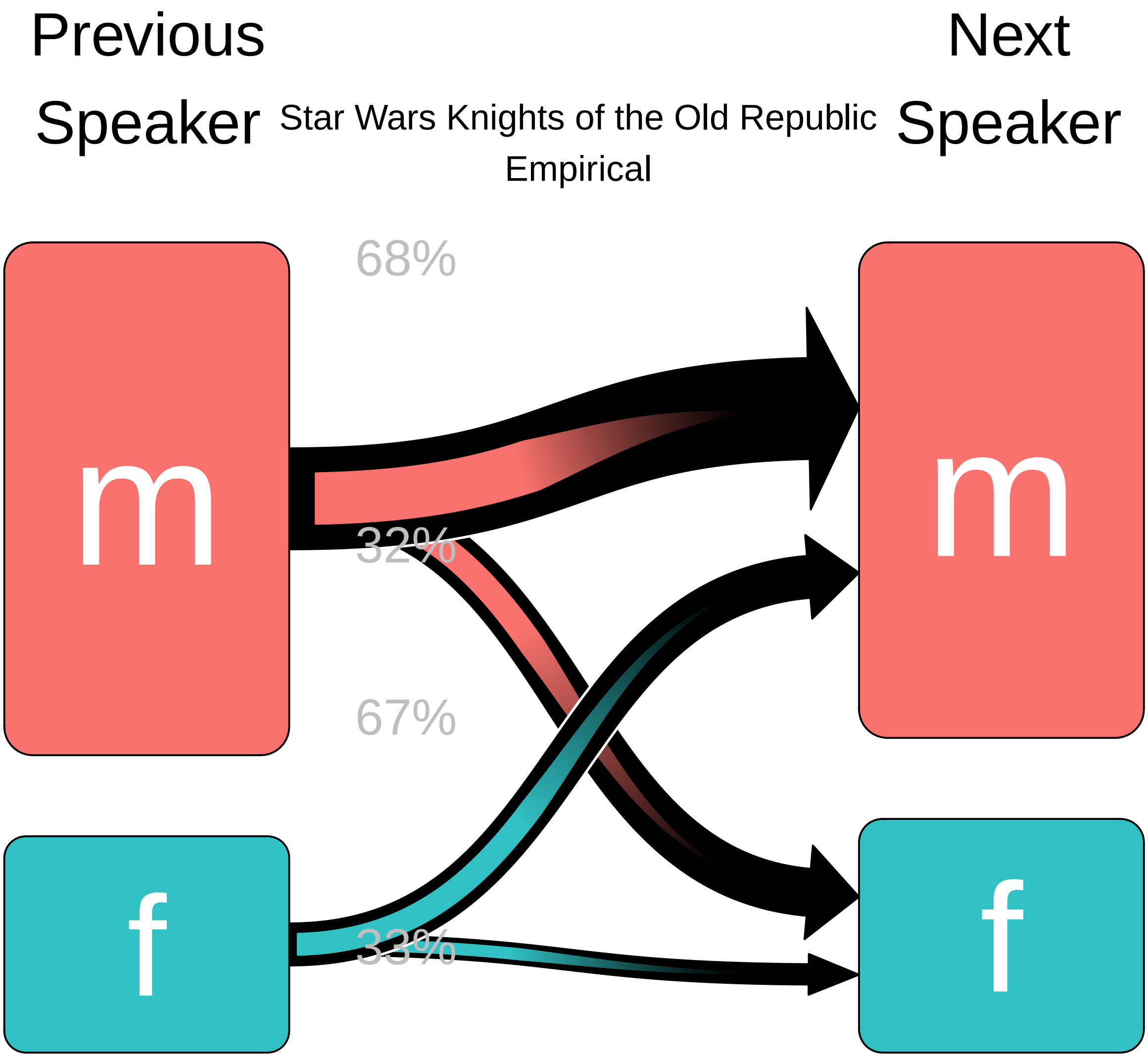


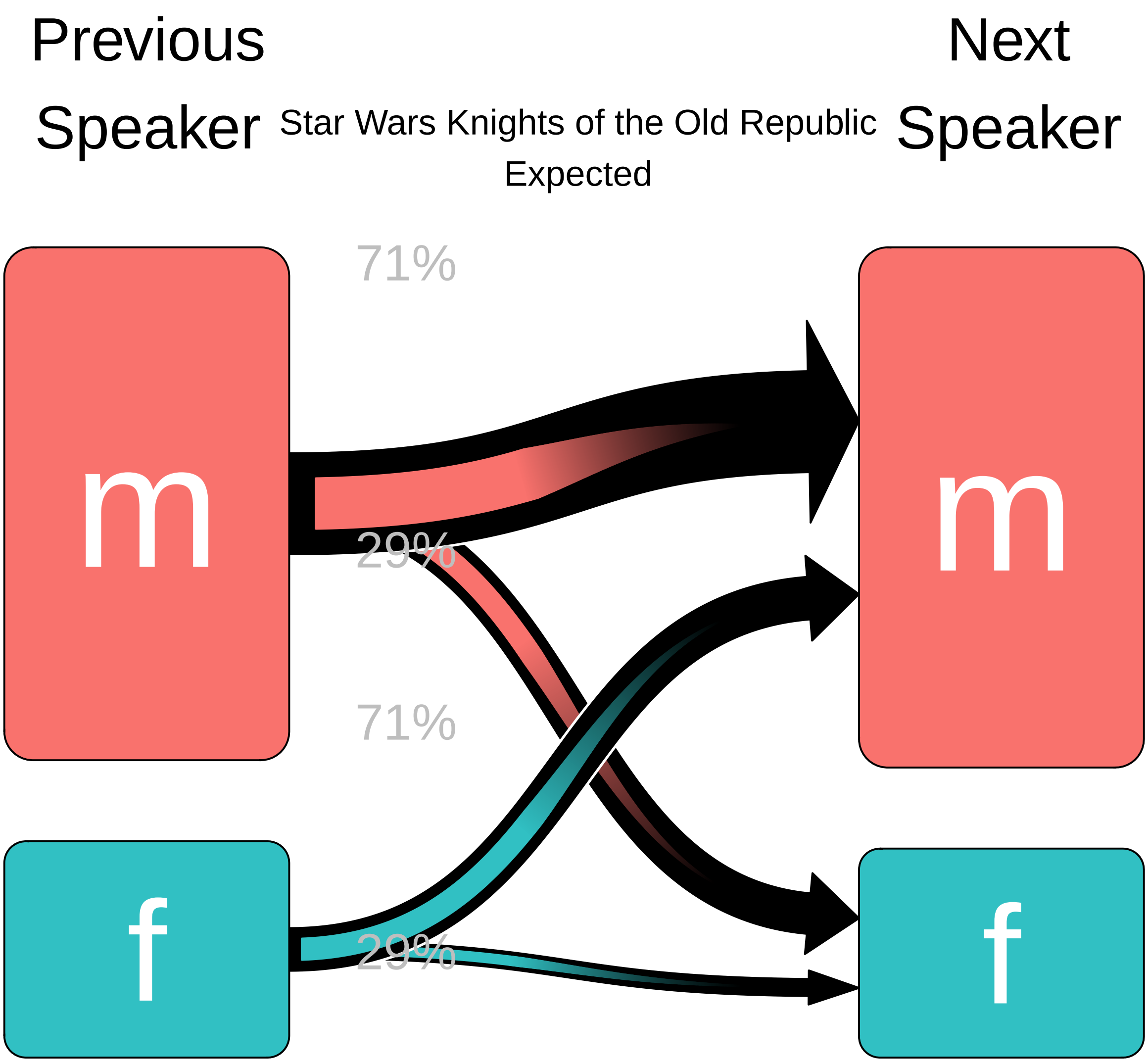
Previous
Speaker
Persona 5
Empirical
Next
Speaker
60%
m
40%
61%
f
39%
m
f

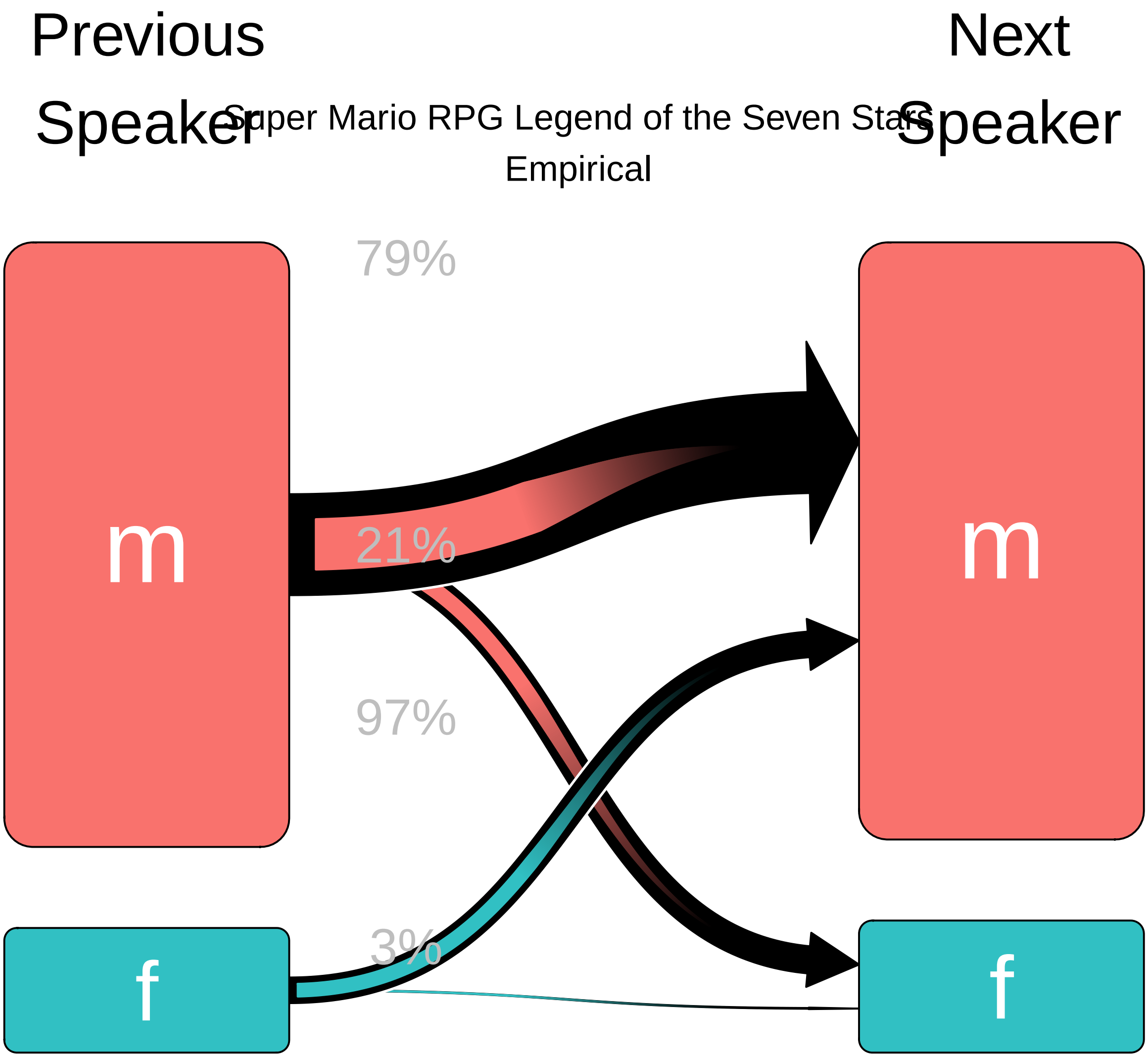

Previous
Speaker
Persona 5
Expected
Next
Speaker
60%
40%
60%
40%
m
f
m
f

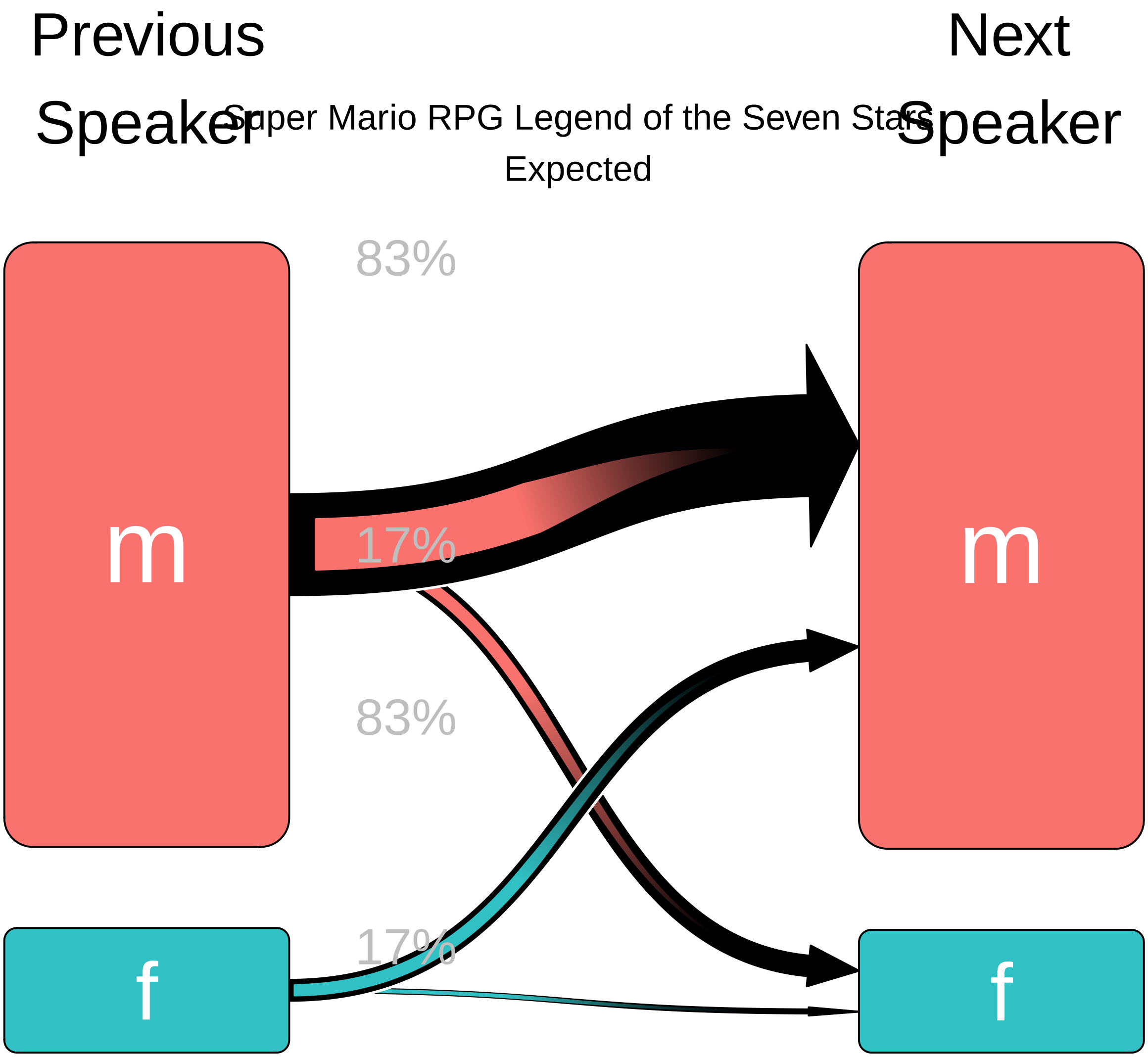

Previous
Speaker
Stardew Valley
Empirical
Next
Speaker
43%
m
57%
77%
f
23%
m
f


Previous
Speaker
Stardew Valley
Expected
Next
Speaker
51%
49%
51%
49%
m
f
m
f

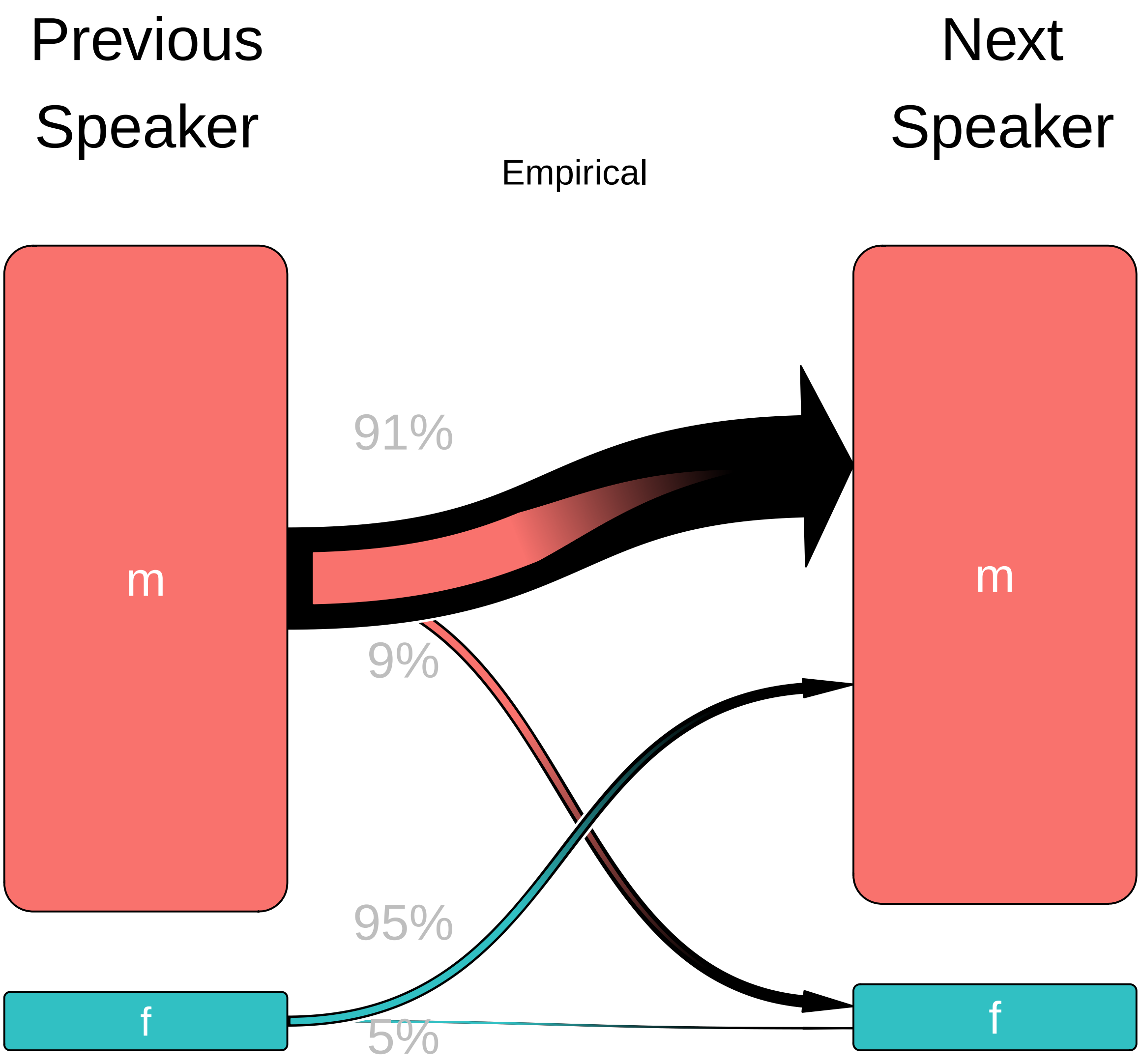

Previous
Speaker
Star Wars Knights of the Old Republic
Empirical
Next
Speaker
m
f
m
f
68%
32%
67%
33%


Previous Speaker
Star Wars Knights of the Old Republic
Expected
Next Speaker
71%
29%
71%
29%
m
f
m
f

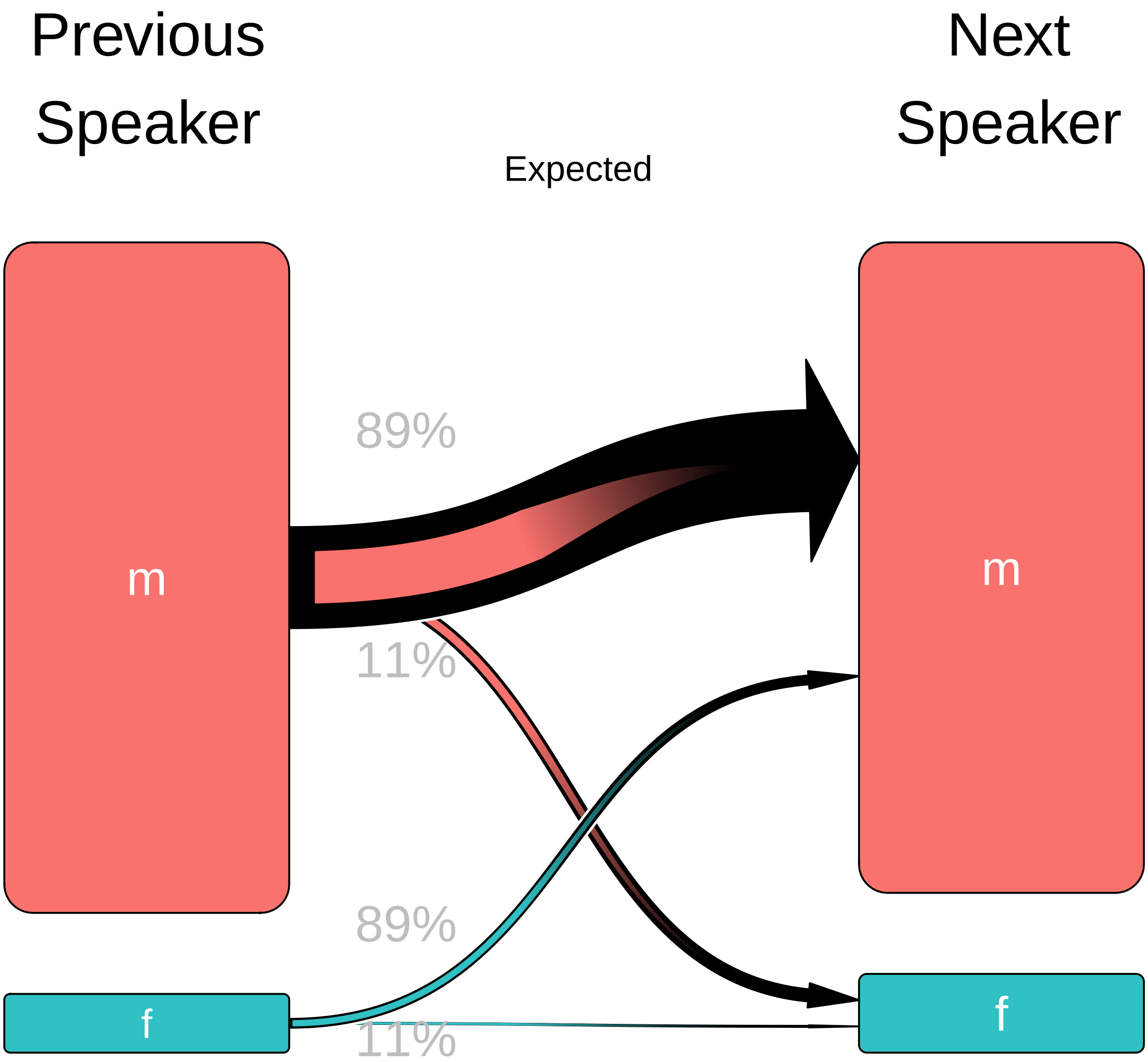

Previous Speaker
Super Mario RPG Legend of the Seven Stars
Empirical
Next Speaker
79%
21%
97%
3%
m
f
m
f




```
makeTransitionGraphForOneGame("Final Fantasy XV", "../results/graphs/transitions/Transitions_FF15.pd
```



Previous
Speaker
Empirical
Next
Speaker
m
91%
m
9%
95%
f
5%
f




```
## pdf
##   2
```

Write stats for paper:

```
numGames = length(unique(permutationResults$folder))
cat(numGames, file="../results/latexStats/transitions_NumGames.tex")

cat(round(100*trueTransitions.AllGames['m','m']),
    file= "../results/latexStats/transitions_MMPer.tex")
MM.moreThanExpected = trueTransitions.AllGames['m','m'] -
  permutedTransitionStats.AllGames['m2m.mean']
cat(round(100*MM.moreThanExpected),
    file= "../results/latexStats/transitions_MMMoreThanExpected.tex")
MMp = paste("=",permutedTransitionStats.AllGames['m2m.p'])
if(MMp <= 0.001){
  MMp = "< 0.001"
}
MMStats = paste0(
  "z = ", round(permutedTransitionStats.AllGames['m2m.z'],2),
  ", p = ", MMp)
```

```
cat(MMStats, file = "../results/latexStats/transitions_MMStats.tex")
MMNumGamesNotBiased = sum(permutationResults$m2m.p>0.05)
cat(MMNumGamesNotBiased, file = "../results/latexStats/transitions_MMNumGamesNotBiased.tex")

# For F-F transitions
cat(round(100*trueTransitions.AllGames['f','f']),
    file= "../results/latexStats/transitions_FFPer.tex")
FF.lessThanExpected = permutedTransitionStats.AllGames['f2f.mean'] -
  trueTransitions.AllGames['f','f']
cat(round(100*FF.lessThanExpected),
    file= "../results/latexStats/transitions_FFLessThanExpected.tex")
FFp = paste("=",permutedTransitionStats.AllGames['f2f.p'])
if(FFp <= 0.001){
  FFp = "< 0.001"
}
FFStats = paste0(
  "z = ", round(permutedTransitionStats.AllGames['f2f.z'],2),
  ", p ", FFp)
cat(FFStats, file = "../results/latexStats/transitions_FFStats.tex")
FFNumGamesNotBiased = sum(permutationResults$f2f.p>0.05)
cat(FFNumGamesNotBiased,
    file = "../results/latexStats/transitions_FFNumGamesNotBiased.tex")

FFp = paste("=",permutationResults[permutationResults$game=="Final Fantasy X-2",]$f2f.p)
if(FFp <= 0.001){
  FFp = "< 0.001"
}
FF.FFX2Stats = paste0(
  "z = ", round(permutedTransitionStats.AllGames['f2f.z'],2),
  ", p ", FFp)
cat(FF.FFX2Stats,
    file = "../results/latexStats/transitions_FF_FFX2Stats.tex")

FFp = paste("=",permutationResults[permutationResults$game=="Final Fantasy X-2",]$m2m.p)
if(FFp <= 0.001){
  FFp = "< 0.001"
}
MM.FFX2Stats = paste0(
  "z = ", round(permutedTransitionStats.AllGames['m2m.z'],2),
  ", p ", FFp)
cat(MM.FFX2Stats,
    file = "../results/latexStats/transitions_MM_FFX2Stats.tex")
```

References

Bechdel, Allison. Dykes to Watch Out For. Firebrand Books (October 1, 1986).

Agarwal, A., Zheng, J., Kamath, S., Balasubramanian, S., & Dey, S. A. (2015). Key female characters in film have more to talk about besides men: Automating the bechdel test. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 830-840).

Weng, C. Y., Chu, W. T., & Wu, J. L. (2009). Rolenet: Movie analysis from the perspective of social networks. IEEE Transactions on Multimedia, 11(2), 256-271.

10 Comparing Jessie's dialogue between Final Fantasy VII and Final Fantasy VII Remake

Comparing Jessie's dialogue between 'Final Fantasy VII' and 'Final Fantasy VII Remake'

The main analysis of this project shows that the proportion of female dialogue is slowly increasing over time. However, this does not necessarily mean that the qualitative portrayal of female characters is improving. To investigate this, this section uses the corpus to conduct a qualitative analysis of how a female character is portrayed through dialogue at two different points in time. This is a difficult in general, since the portrayal of characters differs by the story's genre, the character's role in the plot, the world of the game, game mechanics and many other factors. However, an opportunity for direct comparison is available since '*Final Fantasy VII'* (Square, 1997) was re-made more recently in '*Final Fantasy: VII Remake'* (Square Enix, 2020). The two games have characters, plot and settings in common. This analysis focusses on the character of Jessie, and how she is portrayed through her dialogue in the two games.

Background

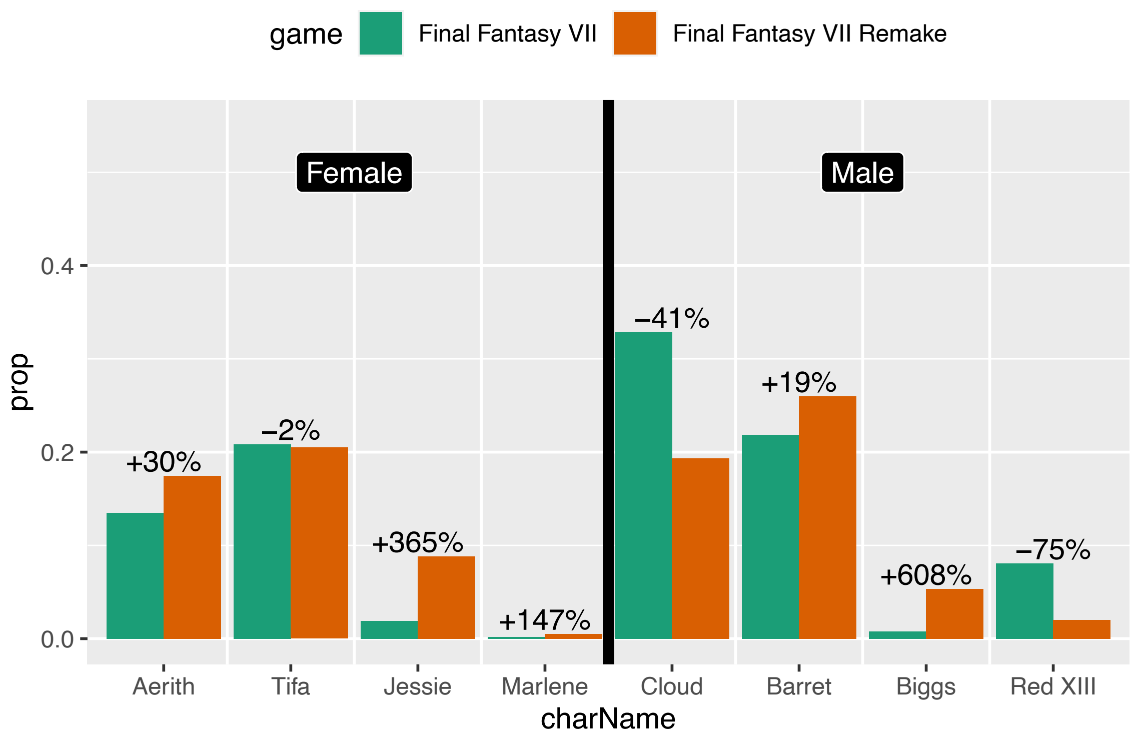
Final Fantasy VII is an action RPG in the 'Final Fantasy' series developed and produced by Squaresoft between 1994 and 1997, and published in 1997. It follows the adventures of Cloud, a soldier who joins an eco-terrorist group ('Avalanche') to fight against a mega-corporation attempting to drain the planet's lifeblood as an energy source (Final Fantasy Wiki, 2021). In 2020, Square Enix released *Final Fantasy VII Remake,* a new game made with modern high-definition graphics and game systems. It adapted the setting, characters and plot of the original, and re-tells the first part of the story of the original.

This analysis focusses on the character of Jessie Rasberry, a character in both games who used to work for the mega-corporation as senior engineer, but now works for Avalanche. During the first Avalanche mission to destroy a reactor, she hacks security doors and provides the explosives. In both games, she is injured during an Avalanche mission and (apparently) dies. While there are more prominent female characters, Jessie provides a more valid comparison than most, since her character arc is completed in both games.

Data

Jessie's dialogue was extracted from the corpus for both games (including lines that are only heard if the player chooses to help Jessie after the explosion), along with preceding lines for contextual cues. Notable moments were identified for particular focus, such as the explosion at the Mako Reactor and Jessie's death.

In the original, Jessie has 32 lines with 655 words. In the remake she has 301 lines with 3,334 words. This is an increase both absolutely and proportionally compared with other characters. The figure below shows the relative proportion of dialogue spoken by four female and four male characters in both games. Jessie has over three times more dialogue as a proportion of these characters.



Methods

Thematic Analysis

Thematic Analysis (TA) is a method of qualitative analysis that allows for the identification, analysis and interpretation of patterns of meaning or 'themes' (Braun and Clarke, 2017). This moves beyond counting words or phrases and focuses on identifying and describing both implicit and explicit content (Guest et al., 2012: 9). Therefore, TA allows for comparing data sets and analysing emerging themes in relation to their broader context, rather than just at a linguistic level.

Braun and Clarke's (2017) 5-step process was followed. The first step is familiarisation with the content. To aid this step, three YouTube videos were used to gain a more comprehensive overview of the dialogue and the context in which it was situated. World of Longplays (2016) provided a complete walkthrough of the original game and Gamer's Little Playground's (2020) allowed for greater understanding of Jessie's scenes and dialogue. For greater qualitative analysis, Infernix Gaming's (2020) video was used to directly compare Jessie's death scenes and identify differences in dialogue.

In the second step, “codes” are generated for each line of dialogue. In TA, “codes” are the smallest units of analysis that can capture interesting data features which help answer the research question (Braun and Clarke, 2016: 297). These small units are then related to broader themes. Finally, there is an iterative process of reviewing, defining and naming themes until they encompass the target domain. Finally, these themes were considered scene-by-scene to examine what functions Jessie's speech serves and whether her dialogue fulfils stereotypically 'feminine' roles within video games.

Results

The thematic analysis of Jessie's dialogue revealed that in Final Fantasy VII, Jessie's character represents an individual who is knowledgeable, confident in her role yet shy and sensitive in herself (see Figure 1). Her dialogue exposed different aspects of her personality and indicated that despite the Original game being produced in 1997, her character does not seemingly fall into many gender stereotypes. Her character does exhibit aspects of the 'Damsel in Distress' trope (Cloud rescues Jessie when trapped after the explosion). However, Jessie gains confidence when she can use her knowledge as a technical expert, dispensing information to the player to help them progress through the game.

22% of Jessie's dialogue from the original game involves dispensing information. This includes dialogue that instructs the player about the game mechanics ("Push the [OK] button in front of a ladder to grab on to it."), and the game world ("The 8 Reactors provide Midgar with electricity."). A further 22% of her dialogue involves displaying technical ability and knowledge ("Code deciphered", and when setting the bomb "OK! Now everyone get back."). The remaining 56% of dialogue falls under 'personality revealing'. These are lines that function to help express Jessie's personality ("Oh yeah, you might regret it. I'm the type who takes things personal."), flirting with Cloud ("It's me, Jessie. How do I look, Cloud? Do I look good in a Shinra uniform?"), and expressing regret over her actions ("Because... of our actions... many... ... people died... ... this probably... ... is our punishment...").

Despite having a limited amount of dialogue in the Original game, her character is not there to play a romantic role or serve as a reward to the main PC. Rather, the character is valued for her knowledge and what she brings as a technical expert and confident team player.

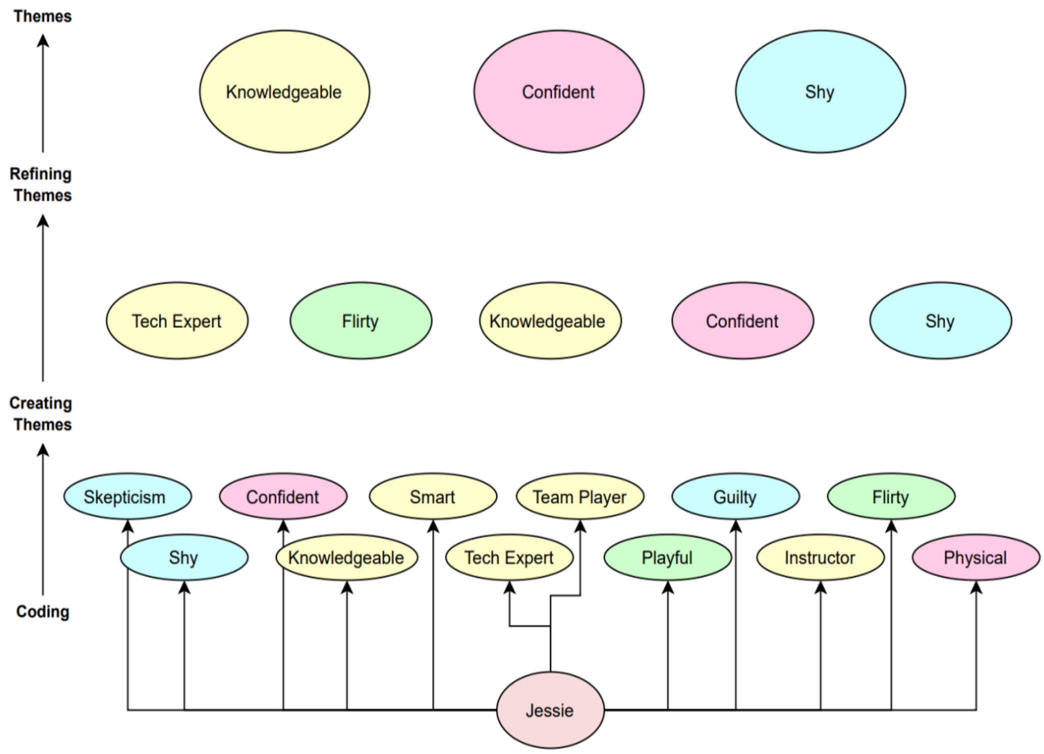


Figure 1: Codes and Themes of Jessie's Dialogue in Final Fantasy VII

In Final Fantasy VII Remake, Jessie has more dialogue. The same themes are observable (see figure 2), but the proportion of dialogue devoted to each theme has changed (see Figure 3). 10% of lines are spent on technical ability/knowledge, 8% on dispensing information, and 82% on personality revealing dialogue. Jessie's character had developed into being less introverted and insecure toward the main PC and became obnoxiously confident, energetic and flirty. While the character still maintains her knowledgeable persona, it is more subtle when she acts explicitly as a technical expert, since there is less emphasis on Jessie deciphering 'code panels'. Instead, her character serves as more of a romantic interest for the player character (Cloud), being more playful, teasing, passionate, flirtatious and charismatic.

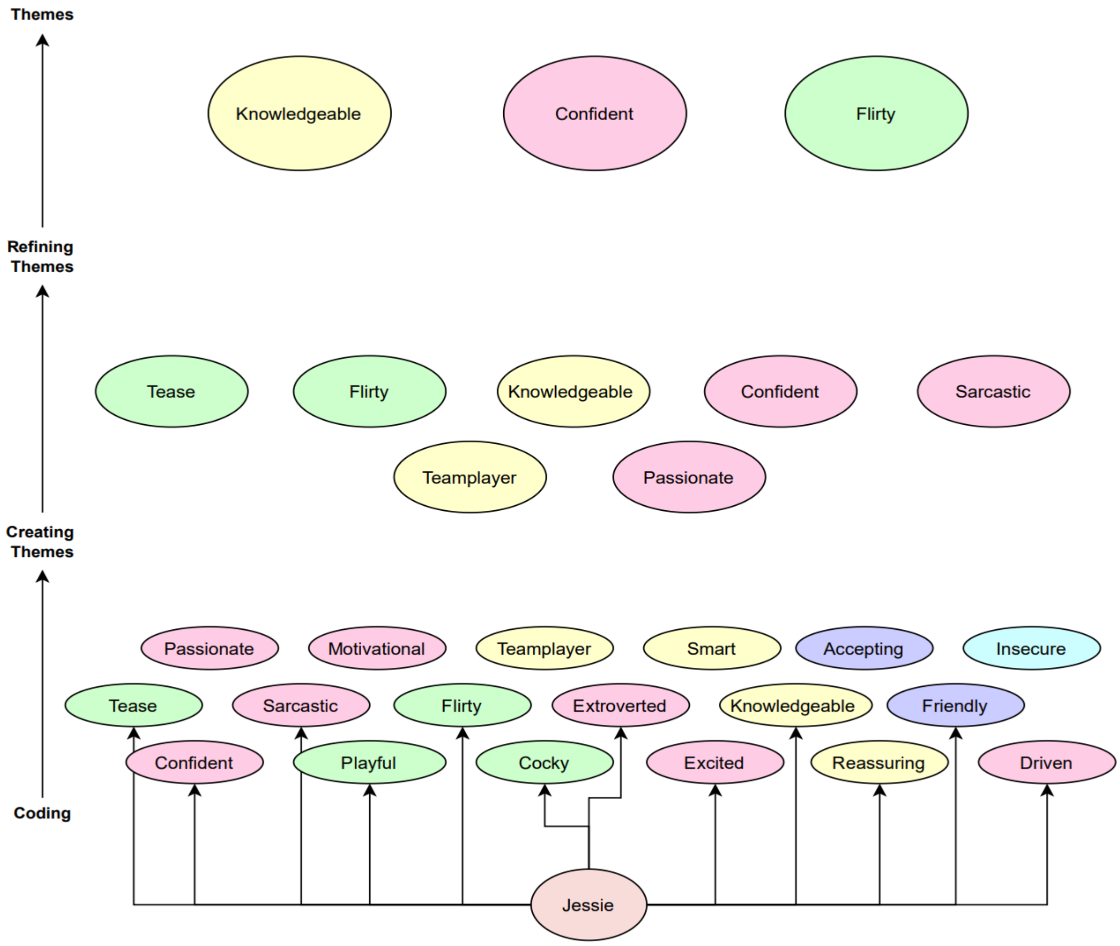


Figure 2: Codes and Themes of Jessie's Dialogue in Final Fantasy VII Remake

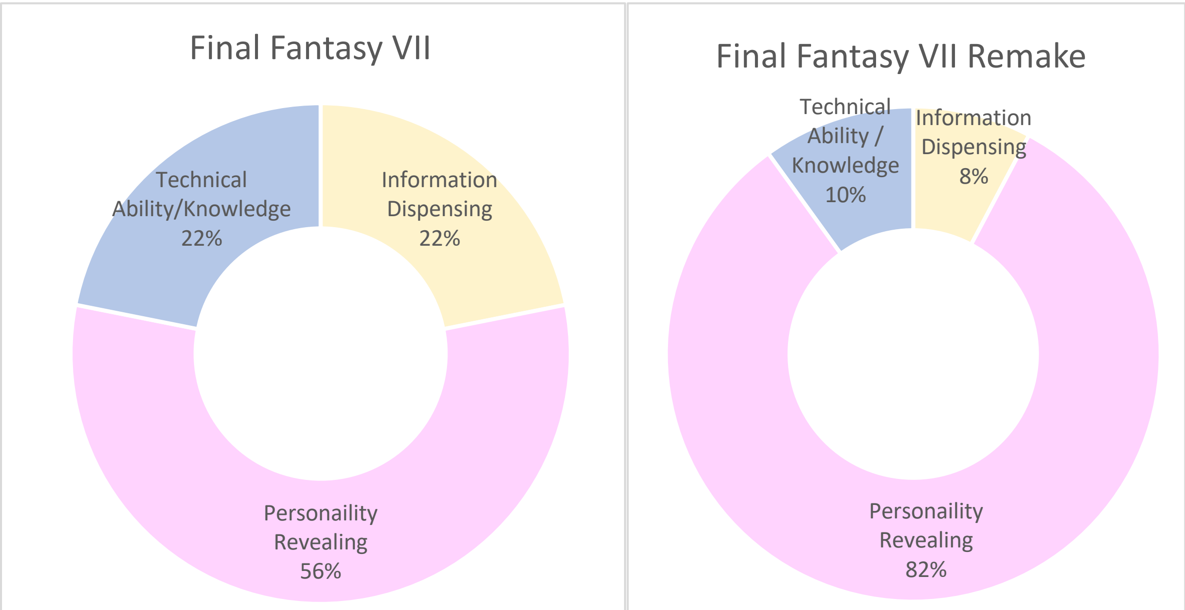


Figure 3: Functions of Jessie's Dialogue in Final Fantasy VII Remake

Direct Comparison

In the Original script, Jessie's character has 32 lines of dialogue, confining the Original Jessie to a minor supporting role. These lines are reflected in the Remake but are significantly expanded upon. This allowed for considerable character development from Jessie being a technical and bomb expert to being an unbridled, passionate and 'unforgettable' character (CBR, 2021).

For example, Table 1 shows a comparison between scenes. There is an increase in the amount of Jessie's dialogue which changes the character from being rather shy and distant towards the main PC to developing into another member of the group with much more personality. Jessie's character becomes obnoxiously flirty, confident and sarcastic, and this is evident in Lines 3 and 4. In the Remake, her attraction towards the main PC is not subtle, and it is explicit in this scene. Similarly, as this is the player's first introduction to Jessie's character, they do not see the technical and knowledgeable aspects of her character but rather are instantly introduced to Jessie's crush.

	FINAL FANTASY VII	FINAL FANTASY VII Remake
1	SOLDIER? Aren't they the enemy? What's he doing with us in AVALANCHE?	So, what's SOLDIER boy's deal? Is he one of us now? He's got balls, this, uh... Uh... what was his name again?
2		Right.
3		Real joy to look at too.
4		Looks are what people notice first.
5		I'd say you're not even reading the same book.
6		Or even the same—

***Table 1**: Jessie's introduction to the main PC*

Table 2 similarly shows two scenes where Jessie expresses regret over making a mistake with the bomb. While her original character discusses her guilt, it was short-lived and quickly accepted. As her line was short, Jessie's character still appears shy, and her guilt is evident (Line 1). However, despite this, it is brushed aside, and she quickly becomes more confident in herself and her achievement (Line 2). This confidence is mirrored in the Remake and although Jessie is less accepting at first, she confidently discusses her guilt. Phrases such as 'I can't stop thinking about it' and 'it doesn't make sense' (Line 1) show her confidence discussing her failures, and she does not ask for reassurance from the main PC. This is confirmed in Line 4 when accepting her responsibility and highlights her confidence and knowledge.

Table 2: *Jessie's guilt and acceptance over her bomb's destruction*

	FINAL FANTASY VII	FINAL FANTASY VII Remake
1	Oops... Hey, look at the news... What a blast. Think it was all because of my bomb? But all I really did was just make it like the computer told me. Oh no! I must've made a miscalculation somewhere.	I can't stop thinking about it. The bomb I made shouldn't have produced an explosion that big. It doesn't make any sense...
2	Hey, that was my bomb's debut. Makes me kinda proud.	
3		That was my first guess—but shouldn't a reactor have fail—safes to prevent that kinda thing? You mentioned "invisible enemies" back there, right?
4		Hmm... No. I'm just looking for excuses for something that was clearly my own fault. Gotta own up to it if I'm gonna learn from this and move on. Thanks, Cloud. You're a good listener.

Finally, Table 3 demonstrates Jessie's significant character development in her death sequence. Jessie's character gains 166% more dialogue in this scene in the remake. There is a central focus on her crush on the main PC. This attraction is evident in both scenes but is significantly developed upon in the Remake. Her flirtatiousness is originally subtle (Line 3) but becomes explicit in the Remake (Line 2), as well as her sarcastic nature (Line 1, 2 and 8). Jessie's character accepts that her death is a punishment resulting from her actions rather than as a consequence of Avalanche's combined efforts. This is clear from her shift in pronouns. Originally Jessie uses 'our' to represent the group as a whole and diminishes her role in Avalanche. This changes in the Remake where she explicitly states that they were 'her' victims and that 'she [I] had it coming'. This further indicates Jessie's character development from only being a minor role in Avalanche's missions to that of a character who has a defined role within the group and signifies her significance to the context of the game.

Table 3: Jessie's Death

	FINAL FANTASY VII	FINAL FANTASY VII Remake
1	... Cloud... I'm glad... ... I could talk with you... one last time.	Oh, jeez... Tifa... If you could see the look on your face...
2	That's... all right... Because... of our actions... many... ... people died... ... this probably... ... is our punishment...	So... you're the guy... who gets to hear my last words... Heh... Lucky me.
3	... Is... that so... ... ? Ha... ... cool... ... as usual... ex-... SOLDIER. ... always... I liked that... in you...	It's okay, Cloud... It's okay. They were my bombs. They were all... my victims. I had it... coming.
4		My hero. So gentle...
5		So dramatic... I... I just wish that... I could've had you over again... Everyone... With Mom's cooking... I really wanted... to believe... we could...
6		That's right. I do. But... I don't think...
7		Oh no... Tifa's crying. Did I say something wrong?
8		Don't you guys... have somewhere to be? It's not polite... to stare... you know...

Conclusion

Jessie's character is expressed differently in dialogue in the two games made more than twenty years apart. Although Jessie has more dialogue in the remake and therefore has a higher representation in the overall context of the game, the content of her dialogue has changed. She moves from a character who dispenses expert information that focuses on knowledge of coding and explosives to serving more as a flirtatious romantic interest for the player character. While sexualisation of characters is not a necessarily bad, when applied to female characters it can perpetuate existing tropes in video games that female NPCs exist to titillate the presumed straight male player (XXXX). In conjunction with Jessie's death at the hands of the antagonists, there are also aspects of the "disposable woman" trope where a female love interest's death is used to motivate the main character (XXXX).

References

Beasley, B & Standley, T. C. (2002) Shirts vs. Skins: Clothing as an Indicator of Gender Role Stereotyping in Video Games. *Mass Communication and Society*, 5:3, 279-293

Braun, V. and Clarke, V. (2017). Thematic Analysis. *The Journal of Positive Psychology*, 12(3), pp.297-298.

CBR. (2021). *Jessie Rasberry Is FINAL FANTASY VII Remake's Best Character*. [online] Available at: <https://www.cbr.com/jessie-rasberry-final-fantasy-vii-remake-best-character/> [Accessed 11 May 2021].

Dickerman, C, Christensen, J. & Kerl-McClain, S. (2008). Big Breasts and Bad Guys: Depictions of Gender and Race in Video Games. *Journal of Creativity in Mental Health*, 3(1), pp.20-29.

Final Fantasy Wiki. (2021). *FINAL FANTASY VII*. [online] Available at: <https://finalfantasy.fandom.com/wiki/Final_Fantasy_VII#Characters> [Accessed 11 May 2021].

Gamer's Little Playground. (2020). FINAL FANTASY 7 REMAKE – ALL Jessie Scenes. Available at: https://www.youtube.com/watch?v=KDVomVk9eIg. [Accessed 11 May 2021]

Guest, G., MacQueen, K.M. and Namey, E.E. (2012). Introduction to Applied Thematic Analysis. In: *Applied Thematic Analysis*, Thousand Oaks, CA: SAGE Publications, Inc. pp. 3-20.

Infernix Gaming. (2020). Change in Jessie's Death Scene – FINAL FANTASY VII Remake PS4 vs Original PS1. Available at: https://www.youtube.com/watch?v=ttDTDkhv-Wo. [Accessed 11 May 2021].

Matthews, N, Lynch, T. & Martins, . (2016). Real ideal: Investigating how ideal and hyper-ideal video game bodies affect men and women. *Computers in Human Behavior*, 59, pp.155-164.

Perry, K. (2021). Damsels and darlings: decoding gender equality in video game communities. *Feminist Media Studies*, pp.1-18.

Square (1997). Final Fantasy VII. [Video Game]

Square Enix (2020). Final Fantasy VII: Remake. [Video Game]

World of Longplays. (2016). PSX Longplay [003] FINAL FANTASY VII (part 1 of 4). Available at: https://www.youtube.com/watch?v=BOUMiEbkZUM [Accessed 11 May 2021].

11 Recruitment strategies in Daggerfall

Recruitment strategies in Daggerfall

Introduction

This report looks at what negotiation strategies female characters use in *Daggerfall* and whether they reflect real-life usage.

While *Daggerfall* belongs to a previous generation of games, it has been remade and re-released to be played on modern consoles, reaching a new audience and therefore continuing to influence players and game creators. The game puts the player in the position of a hero in a medieval fantasy world. Completing "quests" is a central part of the game.

The source for the dialogue of Daggerfall marks "quest offers": dialogue where an NPC tries to convince the player character to accept a quest. This involves a line of dialogue from the NPC that explains the details of the quest and (usually) proposes a reward. The player character can accept or decline, after which the NPC speaks another line depending on the player's decision. The game content is partly procedurally generated, with the gender of the NPC possibly randomly determined. However, this report focusses on a sub-sample of the quests which are scripted for specific NPC characters with pre-set genders.

For this report a pragmatic approach will be taken to analyse how female characters negotiate. Literature surrounding gender stereotypes in relation to female behaviour in negotiations will be reviewed along with research into what constitutes typically female features of speech.

Gender differences in negotiation strategies

Gender stereotypes are deeply ingrained within society (Fiske and Stephens 1993) with women thought to be warm, nurturing and friendly. Men's interactional styles are noted to be assertive, dominant and task-oriented, while women's are indirect, supportive and person-oriented (Holmes and Stubbe, 2003, 574; Lee et al., 2021). Kray and Thompson (2005) argue that this gender belief system dictates how men and women behave in negotiations; women are expected to behave emotionally, with concern for others, and passively. Miller (2002) points out that generally, women place more importance on relationships, preferring to engage in social talk before negotiations start, while men prefer to get straight down to business. In addition, Miller advises women to make men feel comfortable in negotiations by trying to give in on the first request in order for them to drop their guard and want to do something in return. This suggests that conforming to gender stereotypes can work to a woman's advantage.

Real life differences in speech between men and women have been identified. Lakoff (1973) noted features such as hesitations, question tags, and qualifiers and intensifiers that act as hedges, are typical of female speech. These features of language are associated with passiveness and a lack of confidence. Furthermore, it has been found that women prefer to suggest, rather than command (Goodwin 1980). However subsequent research has questioned whether these features really are favoured by women. For example, Dubois and Crouch (1975) found that men use tag questions more than women. Others question whether the features of language attributed by Lakoff to women, are actually better described as powerless language (O'Barr and Atkins 1980). In a courtroom study, O'Barr and Atkins found that witnesses holding less power, for example in terms of their social and professional standing, displayed more use of what Lakoff identified as female language features. This supports the assertion by Wodak and Meyer (2009, 10) that differences in power in hierarchical social structures can be expressed through language. The result is a clash between what society expects of women and what society expects from a person with a high status (Coates, 204: 201).

Translating this to the world of Daggerfall, NPCs are either male or female, and also have a range of social statuses from peasants to royalty. While the game only contains minimal interaction during negotiation, it may still reflect and perpetuate beliefs about how language varies by gender and power relations.

Methodology

Quest offers from Daggerfall were identified where the gender of the quest setter was specified. The data relating to female characters was analysed using a discourse analysis approach. Due to the juxtaposition of feminist issues with those of social status in relation to power it more specifically falls under the category of feminist critical discourse analysis (Lazar 2007). Lazar describes the central concern of feminist critical discourse analysts as "critiquing discourses which sustain a patriarchal social order – relations of power that systematically privilege men as a social group, and disadvantage, exclude, and disempower women as a social group." (Lazar, 2007: 145)

Firstly, linguistic features such as hesitations, question tags and hedging language were identified and highlighted. Secondly, the content of the dialogue was analysed to see if any features stereotypically attributed to one gender were present. Finally, the age and social class of the character was noted. Throughout, dialogue attributed to male characters was analysed to see if there were any differences between male and female characters.

Results

The source for Daggerfall includes 128 quest offers by male quest-givers, 13 from female quest-givers, and 68 from quest-givers whose gender is determined procedurally. Quests from pre-set named characters are a sub-set of all of these quests, including 11 quests offered by females and 3 quests offered by males. The named characters include:

- Cyndassa
- King of Worms
- Lady Brisienna
- Mynisera
- Prince Helseth
- Prince Lhotun
- Princess Elysana
- Princess Morgiah
- Queen Akorithi
- Queen Aubk-i
- Queen Barenziah

As an aside, we note that there is an imbalance in the gender distribution of procedural quests: there are only 2 procedural quests which specify a female quest-giver, but 125 procedural quests which specify a male quest-giver.

Hedging

Hedging can be used to indicate less certainty and act to soften language (see e.g. Lakoff, 1972). Studies have showed them to be a feature of women's language with Holmes (2008) stating it shows women lack confidence in talk. Qualifiers are a form of hedging which can indicate uncertainty and soften an assertion. For example, Princess Elysana says:

> 'Perhaps you can help me out of a bit of an embarrassing predicament?'.

The use of 'perhaps' and 'a bit' indicates uncertainty that the player will do what she wants, despite her royal status, and presents an indirect request that downplays the extent of the imposition on the player. This reflects a polite negotiation strategy typically percieved as more feminine (Lee et al., 2021), and has the effect of making the player feel in a position of power, thus contributing to making the player feel comfortable which is what women have been advised to do in negotiations with men (e.g. Bowles and Flynn, 2010).

Similarly, Queen Barenziah uses the hedge 'rather' to soften the impact of the request:

> "There is a rather sensitive situation brewing".

In contrast male characters get straight to the point with direct language. Prince Helseth opens his quest with

"I need a courier."

Upon acceptance of his quest he says:

"Excellent. Meet me here this time tomorrow."

No hedges are present. This is assertive and decisive language more associated with males (see Lee et al., 2021). These examples of the characters' use of hedges reflects differences found in male and female usage in real-life (Lakoff 1973).

Hesitations are another form of hedging and can be observed in spoken language by pauses of various lengths. The pauses can be represented in written text as elipses ("…"). Other hesitation markers include "er", "um", "uh" and "ah". Queen Akorithi responds with the hesitation marker 'hmmm' when the player declines her quest, before stating

"Hmmm. I shall have to rethink your standing in this court."

While this is a hesitation, the followup sentence is a veiled threat. While she is using her social status to threaten the player, the form of the threat is less assertive and direct than some male dialogue. For example, Prince Helseth responds in a more abrupt way when a player declines his quest:

"Dullard. Why did you let me waste my time with the likes of you?"

This is direct impoliteness.

Question tags

Question tags such as 'don't you?' at the end of a question, can have two functions according to Holmes (1984): modal and affective. Modal question tags request confirmation of information while affective question tags indicate concern for the addressee. Therefore, affective question tags can act as softeners or invite the listener to take part in a conversational turn. For example, Princess Morgiah uses a question tag in as her opening line to the player:

"You're trying to track down that letter of the Emperor's, yes?"

This serves the purpose of inviting the listener to take part in a conversational turn, which is part of the relationship building expected of women in negotiations.

Cyndassa, a maid, also uses an affective question tag when the player declines a quest:

"You only work for the likes of kings, don't ya?,"

Here, the alternative spelling of you as 'ya' represents informal, uneducated, speech, and she is expecting confirmation of her question from the player. The character's language as a female of a lower social class serves to boost the status of the player.

In contrast, when a quest is declined from King of the Worms, he says:

"You must have strong sentiments for Queen Akorithi. What a pity."

This could have been said with a question tag: "You have strong sentiments for the queen, don't you?" but the choice of stronger, more direct language reflects society's expectations that men are more direct.

Topics

The majority of the quests offered by female characters include details of topics typically associated with females such as such family and relationships. For example, one of Queen Aubk-i's quests involves concern for her husband's grandmother, who is in poor health and Mynisera, the mother of King Gothryd, is concerned that her son may be enraged by her quest. Similarly, Princess Elysana's quest involves returning a cloak she borrowed from her father's good friend. In other quests, the females' relationships to other male characters are mentioned: Mynisera describes Queen Aubk-i as the daughter of late King Camaron, Princess Elysana describes herself as King Eadwyre's child and Queen Akorithi talks about her son, Prince Lhotun.

Concern for others well-being is also present. For example, Queen Alkorithi hopes that the player character has "had a comfortable stay so far", and Queen Barenziah is concerned about her family's reputation. This conforms to the trope of women as caregivers, concerned for the well-being of others.

Male characters do occasionally show concern for others (e.g. Prince Lhotun's quest also involves finding out information about his brother). However, male quest givers have a broader set of topics such as catching a thief, finding an ingredient to complete a ritual, and helping with an exorcism.

Compliments

Empirical studies suggest the women use flattery and compliments more often than men (e.g. Holmes, 1988; Herbert, 1990). This is reflected in the vocabulary used by female NPCs in Daggerfall. For example, Princess Elysana's first words to the player, before revealing her quest, are

> "You are even more dashing in person than Lord Woodborne said."

As well as using flattery to build a relationship for a negotiation, the use of the adjective 'dashing', which is typically used to describe a good-looking man, suggests that the player is assumed to be a male.

When Queen Aubk-I's offer of a quest is rejected, she responds with:

> "Still, I had hoped for more gallantry from you."

Gallantry is most often associated with men and can either mean brave behaviour, especially during battle, or a form of courtesy and respect from a man towards a woman. This again suggests the characters are expecting the player to be a male.

Social status

The majority of the female characters in the sample have royal status with the exception of Cyndassa, the maid. Within the characters of royal status there are older and younger characters. These differences in status and age can be seen in the language of the characters.

The princesses in particular use phrases that are fun-loving and child-like. For example, Princess Elysana, after her initial flattery mentioned earlier, uses the word 'dashing' and exclaims 'I love men and women of action!' and "How exciting!". When the player declines the quest she says 'You're not nearly as fun as Lord Woodborne said you would be." This dialogue makes the character appear to be uninterested in the serious implications of the mission, and more interested in fun and excitement, which is more associated with youth. Indeed, Princess Elysana negotiates her quest with:

> "Would you do this teensy-weensy favour for me?".

The reduplicated form is characteristic of child-like speech, and similarly she says:

> "How frightfully nice of you to help a silly girl like myself."

The use of 'silly girl' serves to put herself in a weakened position to the player. It conforms to a trope of using infantalised speech to refer to women ("girl" instead of "woman", MacArthur et al., 2020). While this is used by a female character about themselves, possibly as a negotiation strategy, it appeals to paternalism - a type of "benevolent sexism" that promotes the idea that females need protecting (Glick & Fiske, 1996).

The characters vary in how they use status. Princess Morgiah gives the impression that she is interested in gossip, another typically female topic:

> "There are very few scandals in the Bay of which I am totally ignorant."

The older characters' language does not show the same level of powerlessness, although their power is sometimes stated in relation to a man's. Lady Brisienna says:

> "The Emperor has instructed me to gift you with an artifact of great power and renown"

This suggests her ability to reward the player comes from a man and therefore makes her appear less powerful and more dependant.

Senior female characters are given more assertive dialogue than younger ones, although this is usually preceded by a softening statement. Queen Akorithi shows more direct language, after enquiring in a typically female way about the player's comfort, stating 'Let me get right to the point.' She also says "I need you to promise. . . " but this again is preceded by a softer "I do not know even how to broach this subject."

Cyndassa, the maid, is portrayed differently from the royal characters. She does not display the same feminine qualities such as friendliness and concern for others in her negotiation opening. Instead, she gets to the point straight away:

> “Hey, if I told you I could give you a bit of information . . . ”.

Her powerlessness is displayed in her language by the use of colloquialisms associated with the less powerful classes. She refers to the Queen as ‘my lady’ which shows deference. Furthermore, it is spelled ‘milady’ which is a marked form that distances her linguistically from the more powerful characters. This style of speech is repeated in her attempts to persuade in the negotiation:

> “Now, mind you, I ain’t givin’ it away. . . ”

Softening language is used to take power away from female characters, but in Cyndassa’s case, she was not given power to begin with, and her language is indicative of her less powerful social class.

Conclusion

This report found evidence of gender differences in negotiation strategies in the dialogue of Daggerfall. Female characters negotiate by conforming to gender stereotypes (focus on family and relationships) and using many typically female linguistic features (hedges, tag questions, compliments). The language used to describe the player (‘dashing’ and ‘gallant’) suggests that the average player is assumed to be male.

The extent to which feminine language is used is related to the social status and age/seniority of the NPC. For example, younger female characters are given more feminine language than older, higher status ones. This may suggest that negotiation strategies do not rely only on gender, but are related to power and status. For example, the maid does not use some of the feminine language, but her status is undermined by the use of marked dialectal variants associated with lower classes.

Overall, the dialogue exhibits gendered stereotypes. One source of this limitation is the relatively low number of quests given by female characters. The inclusion of a greater number of female quest givers would provide scope for additional diversity in their speech and the potential to subert some stereotypes. Since negotiation is a key context for the interplay of gender and power, and a common occurance in the mechanics of video games, understanding genered tropes during negotiation could help bring about some positive changes in the representation of female characters’ speech in video games.

References

Babcock, L. and Laschever, S. (2003) Women Don’t Ask: Negotiation and the Gender Divide, Princeton University Press.

Bogost, I. (2007) Persuasive games: The expressive power of videogames. Boston Review. https://doi.org/10.7551/mitpress/5334.001.0001

Bowles, H.R., Flynn, F. (2010) Gender and persistence in negotiation: A dyadic perspective. The Academy of Management Journal, 53 (4):769-787,

Braun C.M.J, and J. Giroux. (1989) Arcade video games: Proximic cognitive and content analyses. Journal of Leisure Research 21, 92-105

Coates, J. (2004) Women, Men and Language. Harlow: Longman.

Goodwin, M. H. (1980) Directives-response speech sequences in girls’ and boys’ task activities. In: McConnell-Ginet, S, Borker, R & Fruman, N.(Eds.) Women and Language in Literature and Society. New York: Praeger.

Goorimoorthee, T., Csipo, A., Carleton, S., and Ensslin, A. (2019). Forms of sociophonetic othering in accented character speech. In: Astrid, Ensslin, & Isabel Balteiro. (2019). Approaches to Videogame Discourse : Lexis, Interaction, Textuality. Bloomsbury Academic.

Fiske, S. T., and Stevens, L. E. (1993). What’s so special about sex? Gender stereotyping and discrimination. In: S. Oskamp & M. Costanzo (Eds), Gender Issues in Contemporary Society (pp. 173–196). Newbury Park, CA: Sage.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. Journal of Personality and Social Psychology, 70, 491–902.

Herbert, R. K. (1990). Sex-based differences in compliment behavior1. Language in society, 19(2), 201-224.

Holmes, J. (1988). Paying compliments: A sex-preferential politeness strategy. Journal of pragmatics, 12(4), 445-465.

Holmes, J. (2008) An introduction to Sociolinguistics (3rd Ed.).

Holmes, J., and Stubbe, M., (2003) "Feminine" Workplaces: Stereotype and Reality. In: Holmes, J. and Meyerhoff, M. (Eds.) The handbook of language and gender. Blackwell Publishing.

Ivory, J. (2009) Still a Man's Game: Gender Representation in Online Reviews of Video Games. Mass Communication and Society. 9. 10.1207/s15327825mcs0901_6.

Jansz, J., and Martis, R.G. (2007) The Lara Phenomenon: Powerful Female Characters in Video Games. Sex Roles 56, 141–148. https://doi.org/10.1007/s11199-006-9158-0

Kray, L. J., & Thompson, L. (2005) Gender stereotypes and negotiation performance: An examination of theory and research. In B. M. Staw & R. M. Kramer (Eds.), Research in organizational behavior. Research in organizational behavior: An annual series of analytical essays and critical reviews, Vol. 26 (p. 103–182). Elsevier Science/JAI Press.

MacArthur, H. J., Cundiff, J. L., & Mehl, M. R. (2020). Estimating the prevalence of gender-biased language in undergraduates' everyday speech. Sex Roles, 82(1), 81-93.

Michelle M. Lazar (2007) Feminist Critical Discourse Analysis: Articulating a Feminist Discourse Praxis, Critical Discourse Studies, 4:2, 141-164, DOI: 10.1080/17405900701464816

Miller, J., & Miller, L. E. (2002) A woman's guide to successful negotiating: How to convince, collaborate, and create your way to agreement. New York: McGraw-Hill.

Lakoff, G. (1972) Hedges: A study of meaning criteria and the logic of fuzzy concepts. In P.Peranteau., J. Levi., & G. Phares. (Eds.). Papers from the Eighth Regional Meeting of Chicago Linguistic Society. Chicago: Chicago University Press.

Lakoff, R. (1973). Language and woman's place. Language in society, 2(1), 45-79.

Lee, A. J., Mason, M. F., & Malcomb, C. S. (2021). Beyond cheap talk accounts: A theory of politeness in negotiations. Research in Organizational Behavior, 41, 100154.

Sarkeesian, A. (2016) All the Slender Ladies: Body Diversity in Video Games, Tropes vs Women in Video Games (video) https://www.youtube.com/watch?v=qbqRtp5ZUGE (accessed 31 May 2021).

Williams, D., Martins, N., Consalvo, M. and Ivory, J. (2009) The virtual census: Representations of gender, race and age in video games. New Media & Society. 11. 815-834. 10.1177/1461444809105354.

Wodak, R. and Meyer, M. (2009) Critical Discourse Analysis: History, Agenda, Theory, and Methodology. In R. Wodak and M. Meyer (Eds.), Methods for Critical Discourse Analysis. London: Sage.

12 Perpetuated gender differences in Stardew Valley

Perpetuated gender differences in Stardew Valley

Introduction

Stardew Valley is a farming simulator, role-playing game (RPG) created in 2016. The playable character (PC) finds themselves the inheritor of their grandfather's run-down farm in Stardew Valley. Escaping their city life, they till the soil to create the farm and life of their dreams, all while interacting with the townsfolk in the local village. Stardew Valley has been praised for its inclusion of same-sex relationship options between the PCs and NPCs (Gayming 2020).

At the start of the game, the player can choose the gender of their player character. This affects some of the NPC dialogue. This report asks whether the content of these differences perpetuate (align with) or subvert gendered stereotypes.

Methods

104 lines from *Stardew Valley* were identified that differ according to the gender of the player character. Each line was coded by hand according to whether it perpetuated or subverted a gendered stereotype. If there were differences that did not obviously relate to gendered stereotypes, these were coded as "neutral". These include differences due to grammatical elements ("he"/"she"), forms of address ("lad"/"lass"), and other neutral phrases ("fine young woman"/ "good guy"). During coding, a fourth category of gender-specific flirting was identified.

Results

```r
d = read.csv("../../data/StardewValley/StardewValley/StardewValleyGenderDiff.csv",
             stringsAsFactors = F)
```

That there is no significant difference in the amount of dialogue to male and female PCs (measured in text character length):

```r
t.test(d$dialogueToMalePC.length, d$dialogueToFemalePC.length)
```

```
##
##  Welch Two Sample t-test
##
## data:  d$dialogueToMalePC.length and d$dialogueToFemalePC.length
## t = 0.027298, df = 205.73, p-value = 0.9782
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.903107  9.153107
## sample estimates:
## mean of x mean of y
##  54.26923  54.14423
```

The results of stereotype coding were as follows:

```r
cbind(Number = table(d$Verdict),
      Percentage = paste(round(prop.table(table(d$Verdict))*100,2),"%"))
```

```
##                            Number Percentage
## ?                          "1"    "0.96 %"
## Gender-specific flirting "15"   "14.42 %"
```

```
## Mixed                        "1"     "0.96 %"
## Neutral                      "62"    "59.62 %"
## Perpetuate                   "25"    "24.04 %"
```

59.62% of cases were coded as 'neutral' (see appendix). 24.04% of cases perpetuated gendered stereotypes. Below are some specific examples.

Examples 1 and 2 perpetuate stereotypes of gendered food, where certain foods are seen as "for" certain genders (see e.g. McPhail et al., 2012). This feeds into notions of diet and weight loss that are seen as essential to feminie beauty ideals (e.g. Malkin et al., 1999).

Example 1 occurs while at a bar with the character Elliot:

> Elliot with male PC: Bartender! Two of your finest ales, please!
>
> Elliot with female PC: Bartender! Fetch me your finest ale. And bring some wine for the lady!

These gendered drink choices perpetuate stereotypes associating ale with men and wine with women.

Example 2 occurs when at a restaurant. The PC does not get to choose their meal (the "@" symbol indicates that the player character's name is used)

> To male PC: I've got a linguini with mushroom cream sauce for Mr. @.
>
> To female PC: I've got a kale and walnut salad for the lady.

The salad is a "light" food associated with dieting, while the pasta with cream sauce is a high-calorie food, which perpetuate gendered food stereotypes.

Examples are not limited to food and drink. For instance, Example 3 is spoken by Abigail in the context of the player character playing a video game:

> To male PC: Well thanks, @. You seem to really know your way around a joystick, huh? I guess that makes sense.
>
> To female PC: Thanks, @. I didn't think you'd know how to work a joystick so well! But it seems you're experienced.

Here the dialogue reinforces the idea that men play video games and women do not. This does not reflect the real world, where there are roughly equal numbers of male and female video game players (ESA, 2021; ISFE, 2021; Nico Partners, 2021; Newzoo, 2019; Pandurov, 2021; Korea Creative Content Agency, 2020).

There were two cases where gender stereotypes are mildly subverted:

> To male PC: "I'd ask you to throw the ball around, but you don't really seem like the sports type."
>
> To female PC: "If you weren't a girl I'd ask you to play catch."

This is a mild subversion of a male stereotype ('men like sports'), although note that there's nothing gendered in the language targeted at the male, which treats the PC just like a person. There is, however, a perpetuation of negative female stereotype in the dialogue to the female NPC ('throw like a girl' - trope: "Gendered Insult").

The second case refers to the player character's fast drinking (of alcohol):

> To male PC: "Heh… fast drinker, huh? Man after my own heart."
>
> To female PC: "Heh… fast drinker, huh? Woman after my own heart."

The text is nearly identical for both genders. However, there is mild subversion here as the NPC treats fast drinking as equally admirable for both genders, while alcohol consumption is often seen as a masculine trope (e.g. De Visser & McDonnell, 2011).

There were no cases of more significant subversion of gendered sterotypes.

Below is a table with a comment on each case of perpetuation:

```
knitr::kable(d[d$Verdict=="Perpetuate",
               c("dialogueToMalePC","dialogueToFemalePC","Comment")],
            row.names = F,col.names = c("to Male","to Female","Comment"))
```

to Male	to Female	Comment
Hey, you must be getting pretty strong working on that farm all day.	Hey, you must be getting pretty fit working on that farm all day.	Men are strong, women are fit.
We'd better head back before the southern wind picks up.	Uh oh... The vibration from your body has caught the attention of a Crimsonfish... We'd better get out of here.	Possible trope: https://tvtropes.org/pmwiki/pmwiki.php/Main/MonsterMisogyny. Certainly fits the stereotype that women's bodies are distractions/encourage negative behaviour.
@'s a better man than you in every respect!	@'s a better person than you in every respect!	Presumably the person being compared to is also a man, but it's interesting that they've defaulted to a 'better man' when possible, rather than just using 'better person' for both. It's possible that they wrote the dialogue for a male PC first (Male as Default) and then amended for a female PC.
Hey... Maybe I'll let you ride sometime, if you want.	Hey... maybe I'll take you for a ride some day.	Men Act, Women Are. The male character is permitted to 'ride', the female character is offered a more passive role.
Don't worry, I'll make sure to show you the ropes before I let you do anything dangerous.	There's no need to worry... I'll make sure you're safe.	Need to protect the women! (trope: Stay in the Kitchen)
Bartender! Two of your finest ales, please!	Bartender! Fetch me your finest ale. And bring some wine for the lady!	Ale for boys, wine for girls. Classic stereotype.
The girls all seem to like you, @...	You're a girl, @...	All girls want the same thing. Features in both sets of dialogue.
I've got a linguini with mushroom cream sauce for Mr. @.	I've got a kale and walnut salad for the lady.	Calorific pasta for the bloke, and a nice 'healthy' salad for the lady (gendered food associations).
Well thanks, @. You seem to really know your way around a joystick, huh? I guess that makes sense.	Thanks, @. I didn't think you'd know how to work a joystick so well! But it seems you're experienced.	Men Act, Women Are. Tropey assumption that only men are gamers.
How can two men get married? It's unnatural... Hmmph. I guess I'm just old fashioned ...	You're part of the family, now. I'm glad that my grandson found such a nice wife.	More homophobia. Note that only the woman is part of the family.
I'll admit, I thought it was... strange... for two men to be together. But you're such a nice young man, and I know you two are in love... I've changed my mind.	You're part of the family, now... and I couldn't be more proud.	More homophobia. Note that only the woman is part of the family.

to Male	to Female	Comment
Nice shoes.	Nice makeup.	Women wear shoes too. But they definitely wear makeup! (Tertiary Sexual Characteristics)
Are those made out of plastic?	Or wait... Are you even wearing any?	Follow on from the above.
@, you look rugged today. I like it.	@, you look pretty today.	Gendered positive characteristics: rugged man, pretty woman.
Look at us, with our little farm. We make a cute couple.	Hey. Maybe it's the golden light, but you look beautiful today.	This is gendered-specific flirting as well, but note female = beautiful association again.
You're very handsome... have I told you that?	You're very beautiful... have I told you that?	Gendered positive characteristics: handsome man, pretty woman.
My skill with words is unmatched, yet I can't find the way to properly describe your allure.	My skill with words is unmatched, yet I can't find the way to properly describe your beauty.	Again woman = beauty.
From the brightest winter star, to the shimmer of an iridium vein... nothing can compare to my wonderful man.	From the brightest winter star, to the fragrant fairy rose... nothing can compare with your captivating beauty.	This is like a man v woman deodorant ad: METAL for men FLOWERS for women (with heavy metal music in the former, and a woman diving into a waterfall for the latter). Also, beauty = woman. Note that men can be rugged, handsome, alluring, and wonderful. But the lady's still just beautiful.
Wow, you look really handsome today! Did you shave? Your jawline is perfect.	Your feminine allure is irresistible today. I can't keep my eyes off you.	Female PC gets to be alluring now! Man = handsome. Sharp jaw.
Wow, you look really handsome today! Did you shave?	Oh, my... you look beautiful today.	Woman = beautiful, man = handsome. Shaving again!
Wow, you look full of energy today!	You look beautiful today.	Man is full of energy, woman is beautiful again.
Wow, you look really handsome today! Did you shave?	Wow, you look beautiful today. Did you do something different with your hair?	Man = handsome, woman = beautiful.
Wow, you look really handsome today! Did you do something with your hair?	Wow, you look stunning today. Did you do something new with your hair?	Upgraded to stunning! Still, it's a synonym for beautiful.
Hey buddy. You look like you're full of energy today!	@. Um... I just wanted to say... You look nice today.	Man = full of energy, woman = looks nice. In keeping with Men Act, Woman Are!
I like your new boots!	Your hair looks nice today.	Girly hair often = tertiary sex characteristic.

Gender differences in flirting and romance

14.42% of cases differed due to gender specific flirting (see appendix). Some of these examples did display gendered attitudes, for example:

> To a male PC: Hey, farmer guy. You look a little burnt.
>
> To a female PC: Hey Farmer girl. You've got a nice tan going.

The male PC's tan is treated as a possible medical condition, while the female PC's tan is treated as a beauty asset. This perpetuates gendered norms of behaviour related to tanning and beauty (e.g. Cox et al., 2009).

More generally, we note that a female PC is repeatedly described as beautiful, while the male gets praised for a bigger range of attributes (rugged, handsome, full of energy etc.).

We note that several differences relate to heteronormativity. When the player romances Alex (a male character), the dialogue differs, with dialogue to a male PC hinting at internalised homophobia:

> To male PC: I kept telling myself 'You can't have these kinds of feelings for another guy'.
>
> To female PC: Normally, when I have a crush on someone, the feeling goes away pretty quick...

Similarly, female-female attraction is called out specifically when romacing Abigail (a female NPC):

> To male PC: You don't have to say anything right now. Let's just stay here for a while...
>
> To female PC: I didn't know I felt this way about other girls... until I met you.

It also appears that a male PC romancing Elliott requires comment:

```
{"PC": "I'm happy."},
{"CHOICE": [
        [
            {"STATUS": "Player is male"},
            {"Elliott": "... I was worried you might not feel this way about another man. "}],
        [
            {"STATUS": "Player is female"},
            {"Elliott": "So am I."}]]}
```

And for Sebastian:

> To male PC: Um... I've never... felt anything like this with another guy before... But you're different.
>
> To female PC: I don't usually bring girls to this place... In fact, you're the only one.

Homophobic views are also expressed by other NPCs. For example, if the PC romances Alex, the lines below are spoken by George, Alex's grandfather:

> To male PC: How can two men get married? It's unnatural... Hmmph. I guess I'm just old fashioned ...
>
> To female PC: You're part of the family, now. I'm glad that my grandson found such a nice wife.

Including narratives of sexual discovery does not perpetuate stereotypes in itself. And we also note that George changes their mind ("I'll admit, I thought it was... strange... for two men to be together. But you're such a nice young man, and I know you two are in love... I've changed my mind."). However, heterosexuality is consistently portrayed as the non-marked norm.

Letters

The PC in Stardew Valley can receive written letters. While these are not dialogue, we note that there are differences based on gender. For example, at the beginning of the game, a male PC will receive a letter from "Mom":

> Dear @, How are you doing, sweety? I've missed you so much since you left. I hope the farming life is everything you hoped for. Love, Mom. PS, I sent your favorite cookies.

If the PC is female, they will instead receive a letter from "Dad":

> Dear @, Have you settled into your new life yet? I can't believe you're all grown up now ... Time sure flies. Now that you're gone I have all this extra money laying around, so I included a little gift. Love, Dad.

The cookies from Mom have an in-game value of 140g ('g' being short for 'gold', or the in-game currency), while Dad provides 500g in cash. The difference in value leads to many speedrunners choosing a female PC (e.g. this advice https://www.speedrun.com/stardew_valley/guide/ldpiz). We note that this difference broadly aligns with tropes that females provide love while males provide material support.

Conclusion

In many ways, Stardew Valley is a positive, inclusive game. There are no mechanical differences when choosing the gender of the PC, the dialogue is 46% female overall (relatively balanced in comparison to many others), and there are no significant differences in the amount of dialogue spoken to male and female PCs. Furthermore, the option for same-sex marriages and relationships has been in the game since its inception. However, the content of some dialogue does perpetuate gender stereotypes, particularly related to food and beauty standards. The perpetuation of the idea that video games are a masculine hobby is particularly disappointing for a video game in a genre that has more female players than male (around 69% female players according to this study https://quanticfoundry.com/2017/01/19/female-gamers-by-genre/).

The portrayal of stereotypes in fiction is not necessarily harmful. For example, some of the sexist attitudes of Alex are clearly part of his personality and not necessarily praised. However, there were no clear cases of subverting of gender stereotypes, and little push-back on the idea that gendered attitudes are normal, default attitudes. Interactive fictional media possesses the ability to be both aspirational and subversive, and could strive to be more inclusive.

References

Gayming 2020. Can you be gay in Stardew Valley? Available at: https://gaymingmag.com/2020/10/can-you-be-gay-in-stardew-valley/ [Accessed: 10 May 2021].

McPhail, D., Beagan, B., & Chapman, G. E. (2012). "I Don't Want to be Sexist But..." denying and re-inscribing gender through food. Food, Culture & Society, 15(3), 473-489.

Malkin, A. R., Wornian, K., & Chrisler, J. C. (1999). Women and weight: Gendered messages on magazine covers. Sex Roles, 40(7), 647-655.

De Visser, R. O., & McDonnell, E. J. (2011). 'That's OK. He's a guy': A mixed-methods study of gender double-standards for alcohol use. Psychology and Health, 1–22,

Cox, C. R., Cooper, D. P., Vess, M., Arndt, J., Goldenberg, J. L., & Routledge, C. (2009). Bronze is beautiful but pale can be pretty: the effects of appearance standards and mortality salience on sun-tanning outcomes. Health Psychology, 28(6), 746.

ESA (2021). 2021 essential facts about the video game industry. Entertain. Softw. Assoc.

ISFE (2021). 2021 key facts about the european video games sector. Interact. Softw. Fed. Eur..

Nico Partners (2021). The china gamers report. https://nikopartners.com/china-gamers-report/.

Newzoo (2019). Newzoo and gamma data report. https://newzoo.com/insights/articles/newzoo-and-gamma-data-report-female- mobile-gamers-in-japan-play-more-per-week-than-men.

Pandurov, M. (2021) 15 amazing video game statistics for Canada in 2021.

Korea Creative Content Agency (2020). 7th survey report of video gamers in korea. https://seoulz.com/the-korea-creative-content- agency-kocca/.

Appendix

Cases of Neutral differences:

```
knitr::kable(d[d$Verdict=="Neutral",
               c("dialogueToMalePC","dialogueToFemalePC")],
             row.names = F)
```

dialogueToMalePC	dialogueToFemalePC
You're an interesting guy, @. I'm glad you moved here.	You're an interesting lady, @. I'm glad you moved here.
Hey farm guy, I've got a question for you.	Hey Farm girl, I've got a question for you.
That's insane. You're just jealous that I'm talented and popular and you're not.	That's ridiculous. You're just jealous that I'm way more popular than you'll ever be, freak.
Hey, I want to ask you something.	Hey, can I ask you something?
Do you think the ladies like my haircut?	What do you think about my haircut?
Oh, hello my dear. I think of you as my own grandson now.	Oh, hello my hear. I think of you as my own grand-daughter now.
Oh! You brought the cave carrot! Thank you so much, Mr. @.	Oh! You brought the cave carrot! Thank you so much, Ms. @.
I'm glad.	Is that embarrassing to say?
Ahoy there, son.	Ahoy there, miss.
Oh, good morning Mr. @!	Oh, good morning Ms. @!
Good morning, Mr. @!	Good Morning, Ms. @!
My dear boy…	My dearest grand-daughter…
It's been many years since we last spoke. You were just a little boy… Do you remember?	It's been many years since we last spoke. You were just a little girl… Do you remember?
My dear boy…	My dearest grand-daughter…
Well, he'll make it… but that was pretty violent, @.	She's fine… but that was pretty violent, @.
Well, he's fine… but that was pretty violent, @.	She's fine… but that was pretty violent, @.
He has first-hand experience living in the countryside… so he knows all about the valley's 'natural resources'!	She has first-hand experience living in the countryside… so she knows all about the valley's 'natural resources'!
Mr. @, are you going to marry Miss Penny?	Ms. @… Um, do you have a boyfriend?
Vincent! Behave yourself. Mr. @ makes his living working on the farm! It's not a playground.	Vincent! Behave yourself. Ms. @ makes her living working on the farm! It's not a playground.
Heh… fast drinker, huh? Man after my own heart.	Heh… fast drinker, huh? Woman after my own heart.
Oh, it's that new farm boy.	Oh, it's that new girl from the farm.
He has a name, you know.	She has a name, you know.
Thanks, young man.	Thanks, miss.
You've been very nice to me, young man. I appreciate that.	You've been really nice to me, miss. I appreciate it.
Not so fast! I'd like to get a second opinion from this young man.	Not so fast! I'd like to get a second opinion from this young lady.
He was always nagging me to go back to school and study business or medicine… something with a lot of money in it.	She was always nagging me to go back to school and study business or medicine… something with a lot of money in it.
Oh, Hello there, buddy.	Hi, @.
Got any tips?	What advice can you give me?
…and a grilled steak for Alex.	And a grilled steak for the gentleman.
He's right… That's what happened.	She's right… that's what happened.
What!? He's lying!	What!? She's lying!
Huh? What did you say to him, Dad?	Huh? What did you say to her, Dad?
@, I'm sorry I mistrusted you. You're a good guy.	@, I'm sorry I mistrusted you. You're a fine young woman.
Wow… I didn't think he was that dense.	Wow… I didn't think she was that dense.

dialogueToMalePC	dialogueToFemalePC
But you should really be clapping for @! Without his help we'd never have decided what kind of music to make in the first place!	But you should really be clapping for @! Without her help we would've never decided what kind of music to make in the first place!
But you should really be clapping for @! Without his help we'd never have decided what kind of music to make in the first place!	But you should really be clapping for @! Without her help we would've never decided what kind of music to make in the first place!
But you should really be clapping for @! Without his help we'd never have decided what kind of music to make in the first place!	But you should really be clapping for @! Without her help we would've never decided what kind of music to make in the first place!
But you should really be clapping for @! Without his help we'd never have decided what kind of music to make in the first place!	But you should really be clapping for @! Without her help we would've never decided what kind of music to make in the first place!
He gave me the idea for this show and the courage to go through with it. I know that sounds cheesy, but it's true.	She gave me the idea for this show and the courage to go through with it. I know that sounds cheesy, but it's true.
There he is, everyone! Look sharp!	There she is, everyone! Look sharp!
Ah, you turned up at the right moment, son.	Ah, you turned up at the right moment, miss.
He's a nice young man…	She's a nice young lady…
I can't talk right now, young man.	I can't talk right now, miss.
Take care, son.	Take care, hun.
Oh… you're that new farmer boy, aren't you?	Oh… You're that new farmer girl, or whatever. Aren't you?
Hmm… If it weren't for those horrendous clothes you might actually be cute.	Hmm… If it weren't for those horrendous clothes you might actually be pretty… Actually, nevermind.
Hi, Mr. @.	Hi, Ms. @.
You're always so nice to me, Mr. @	You're always so nice to me, Ms. @.
I love animals, Mr. @. If you treat yours well I'm sure we'll become good friends!	I just love animals, Ms. @. Treat them kindly and we'll become friends, I'm sure!
Ahh… there's nothing like a good night's sleep next to my husband!	Ahh… there's nothing like a good night's sleep next to my wife!
Good evening. Did you have a productive day, @?	Good evening. Did you have a productive day, my dear?
You look a little soggy, hun. Why don't you warm yourself by the fire?	You look a little soggy, dear. Why don't you warm yourself by the fire?
For a while there I was considering building a robotic husband. But you're a lot more fun!	For a while there I was considering building a robotic wife. But you're a lot more fun!
Oh, it's Mr. @.	Oh it's Miss @.
Hey, it's Mr. @, the new farmer! I'm Pierre, owner of the local general store.	Hey, it's Ms. @, the new farmer! I'm Pierre, owner of the local general store.
Hello, er… my son?	Hello, er… daughter?
I really do appreciate your business, Mr. @. I've been having a harder and harder time turning a decent profit.	I really do appreciate your business, Ms. @. I've been having a harder and harder time turning a decent profit.
Okay, Dad…	Okay, Mom…
Can you keep a secret, mister?	Can you keep a secret, miss?
Hi there, mister!	Hi there, miss.
Hi Uncle @!	Hi Auntie @!
Ahoy there, son.	Ahoy there, lass.

Cases of Gender-specific flirting:

```
knitr::kable(d[d$Verdict=="Gender-specific flirting",
               c("dialogueToMalePC","dialogueToFemalePC")],
             row.names = F)
```

dialogueToMalePC	dialogueToFemalePC
Oh, hey. So you're the new guy, huh? Cool.	Hey, you're the new girl, huh? I think we're going to get along great. I'm Alex.
sigh... I wish there were more girls in this town, know what I mean?	Hey, do you wanna hang out with me at the beach some time? Do you have a bikini?
Hey, what's up farmer guy?	Hey, it's farm girl. Did you get new pants? You're doing something right.
Hi @. You look sporty today.	Hey @. Did you do something different with your hair? Something keeps grabbing my attention.
Hey, farmer guy. You look a little burnt.	Hey Farmer girl. You've got a nice tan going.
That's fine with me. You're a cool guy.	I'm glad.
...I'll always remember this night.	You look so beautiful tonight... I... *gasp*
... I was worried you might not feel this way about another man.	So am I.
You don't have to say anything right now. Let's just stay here for a while...	I didn't know I felt this way about other girls... until I met you.
Um... I've never... felt anything like this with another guy before... But you're different.	I don't usually bring girls to this place... In fact, you're the only one.
Well... when we first met, I was instantly drawn to you. It was confusing... I'd never felt that way about anyone.	When I first met you, I thought you were really cute.
I kept telling myself 'You can't have these kinds of feelings for another guy'.	Normally, when I have a crush on someone, the feeling goes away pretty quick...
...But my heart was telling me something else.	...But with you, it kept growing.
Hey! You look like you've been hard at work. Can I help you relax?	Have you been working hard? You look cute when you're a little exhausted.
You're a fine lookin' young man. Why're you wastin' your time talkin' to an old girl like me?	You're a pretty gal. What 're you wastin' your time talkin' to someone like me for?

Uncoded case. This is a different reaction (mock disbelief to the male? I'm not sure without context; earnest acceptance to the female), but it's not obviously tropey.

> To male PC: Get out!
>
> To female PC: That's good to know.

13 Gender biases in translations of Chrono Trigger

Chrono Trigger Translations

Introduction

This project attempts to quantify the amount of dialogue spoken by characters of different genders in video games. The main analyses are based on English language versions of the games, but several game scripts in the corpus are translations of original Japanese scripts. There is a possibility that the measures of the amount of dialogue might have been different had they been based on the original scripts. While the English translations have been the main experience for a large proportion of players, it is still interesting to ask whether the gender bias is in the original version or has been emerged during translation. In this report, we test this possibility by looking at different translations of *Chrono Trigger*.

Chrono Trigger is an appropriate game for several reasons. The first is that the source transcript for Chrono Trigger in the corpus has three versions that have been parsed to create a parallel corpus. Thes versions include the original Japanese script (1995), the original English translation (1995), and a retranslation created by a fan (completed in 2007). For future work, it also includes an official retranslation, and various translations into other languages.

The ideal game for this test of robustness would allow us to test many different causes of differences between translations. These include:

- Problems with translation including fluency or human error.
- Wider issues such as the inherent differences between languages, between orthographic systems, or between cultures
- Different aims of the developers or marketing campaigns for different audiences
- Different censorship norms in different countries.
- Diachronic change, including differences in language norms, cultural norms, and fasion, if the translations are made years apart.

The ideal game would also have a lot of qualitative differences between translations, in order to get a conservative estimate of the robustness.

Chrono Trigger is a good fit for these requirements. Firstly, some fans have noted differences between the translations. Fan retranslations exist for many games for a variety of reasons (Sánchez, 2009), but according to its authors, their aim was to create "a clearer portrayal of Chrono Trigger as intended by its Japanese creators", and "It is not the opinion of this project that Ted Woolsey's official translation was bad or insufficient in any way – only that some essence of the game was lost or altered, given Nintendo of America's censorship standards and the inability of the game to hold all the original text when translated to English." (https://www.chronocompendium.com/Term/Retranslation.html). The final point relates to text length limits due to the screen resolution (the same message may take up less space in written Japanese than in written English). In general, the retranslation is a more direct translation.

Further qualitative differences have been noted by various previous studies on the translation of *Chrono Trigger*. For example, Antonijevic (2018) suggests that the character of "Frog" (an anthropomorphic frog warrior) is portrayed in the original Japanese text as "a tough stereotypical [Japanese] warrior", but was originally translated to be a "pseudo-Shakespearean speaking knight". The example below demonstrates this, with the retranslation being a more direct translation:

> **Frog**: バ、バカヤロー！それより、あお白いツラしたマントのヤローは、いなかったか !?
>
> **Translation**: P, perish the thought, lass! By the way, whither the blue-haired one?
>
> **Retranslation**: Wh, why you! More importantly, that pale-faced caped bastard wasn't there!?

The translation changes are due to Western censorship norms (ommitting swearwords), and an attempt to convert the cultural tropes to ones familiar to a Western audience (a "poor attempt to adapt a form of Shakespearean-style of speech", Antonijevic, 2018, p.28).

Various other changes during translation of *Chrono Trigger* are noted by Müller Galhardi (2014), including "dialogue additions and omissions; the re-creation of play on words; the re-naming of characters and terminology; censored items; the deliberate use of regional expressions, and the modification of a character's speech style" (see Mangiron, O'Hagan & Orero, p.15).

Finally, Williams (2014)'s examination of the development of Chrono Trigger includes interviews with the original English translator. These document various complications with the Japanese, including communicating the game world through text.

In short, there is good reason to believe that *Chrono Trigger*'s translations have many qualitative differences. If the measures of dialogue proportions for different genders are robust here, it's likely that they will be robust for many other games.

Measuring text length

The Texstatistic package provides estimates of text length in English in terms of number of words, number of syllables and number of alphanumeric characters. This was used to calculate measures of length for the original English translation and the fan retranslation:

- **eng_char**: number of alphanumeric characters (excluding spaces and punctuation).
- **eng_syll**: number of syllables.
- **eng_word**: number of words.

There are two main challenges to estimating length in the Japanese text. Firstly, the text of *Chrono Trigger* is a mix of Kanji, Hiragana and Katakana characters. These have different relationships with phonetic word length. Secondly, the estimate for English words, which is the main measure used in the main paper, depends heavily on spaces and punctuation in the orthography. The Japanese text typically does not include spaces between orthographic words.

To address this, we used the *Pykakasi* package (version 2.2.1, https://pypi.org/project/pykakasi/) to tokenise the string of Japanese text into words. After removing punctuation and spaces from this tokenisation, the number of tokens are counted. This is used as a measure of the number of 'words'. These tokens are then used to calculate various measures of text length:

- **jap_char**: Number of text characters in the original Japanese game text (mix of writing systems), excluding punctuation and spaces. This does not relate easily to a particular spoken linguistic unit, but is not converted by the *Pykakasi* package, so does not contain any biases from conversion.
- **jap_kana**: Original is converted to pure Katakana (a syllabic system). The number of katakana syllabograms is then used as a measure of the number of syllables in the text.

- **jap_romaji**: Original is converted to Romaji (phoneticisation using latin alphabet) based on Kunrei-shiki conventions. This is used as an estimate of the number of phonemes. Kunrei-shiki is preferred here over Hepburn conventions because it more closely represents one phoneme as one orthographic character (し = 'si' in Kunrei-shiki, but 'shi' in Hepburn).
- **jap_word**: The number of romaji tokens identified in the tokenisation step (excluding punctuation and spaces). This is used as a measure of the number of words.

See the script in the corpus repository for more details: `data/ChronoTrigger/ChronoTrigger/compareTranslations.py`

As is shown below, all of these measures are highly correlated and all lead to the same conclusion regarding our goal of estimating dialogue length.

Estimates of gender bias

We can estimate the proportion of female dialogue according to the three texts: Original Japanese, original English translation, and later retranslation.

Load libraries:

```
library(ggplot2)
library(GGally)
library(lmtest)
library(sjPlot)
library(ggpubr)
```

Load the data:

```
# Read the data
d = read.csv(
  "../../data/ChronoTrigger/ChronoTrigger/compareTranslations.csv",
  stringsAsFactors = F,encoding = "UTF-8")
# Remove non-dialogue data
d = d[!d$charName %in% c("SYSTEM","ACTION", "LOCATION"),]
# Remove characters in the 'netural' gender group
d = d[d$gender %in% c("male","female"),]
```

Calculate the proportion of dialogue by female characters:

```
measures = c(
  "eng_char","eng_syll","eng_word",
  "re_char", "re_syll", "re_word",
  "jap_char","jap_romaji","jap_kana", "jap_word")
propFemaleDialogue =
  sapply(measures,function(X){
  total=tapply(d[,X],d$gender,sum)
  prop.table(total)[1]})
maxDiff = max(abs(
  outer(propFemaleDialogue,propFemaleDialogue,"-")))
ests = data.frame(Measure = measures,
  "PercentFemaleDialogue" =
    paste0(round(propFemaleDialogue *100,2),"%"))
knitr::kable(ests)
```

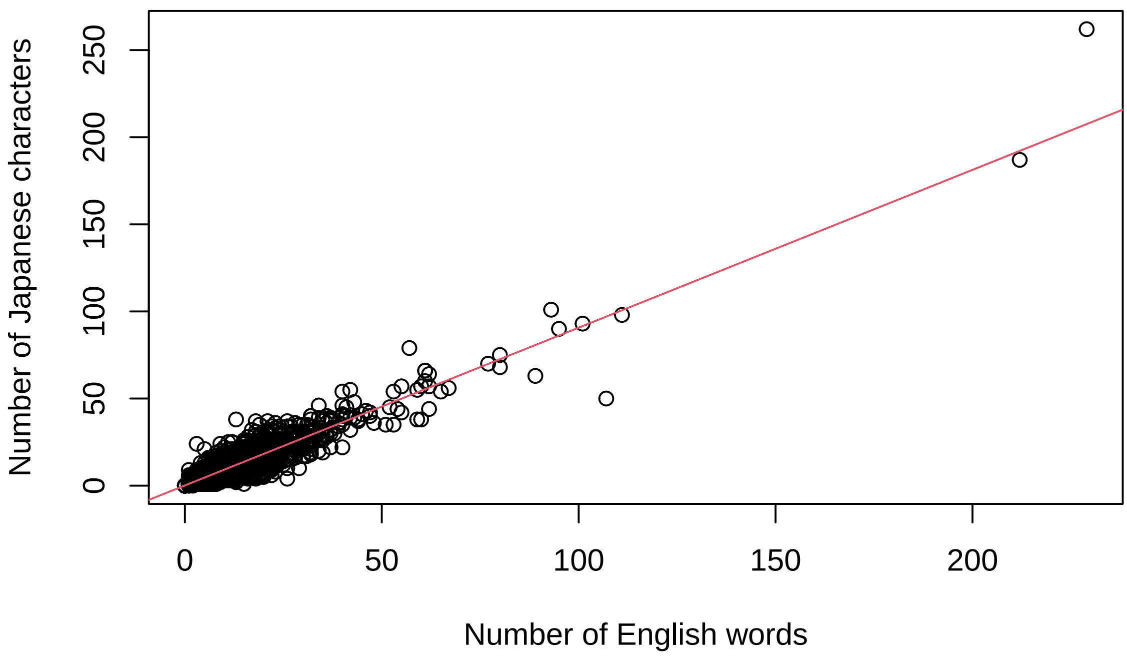
Measure	PercentFemaleDialogue
eng_char	38.16%
eng_syll	37.66%
eng_word	38.27%
re_char	37.96%
re_syll	37.64%
re_word	37.97%
jap_char	38.35%
jap_romaji	38.26%
jap_kana	38.12%
jap_word	37.99%

The estimates are very close, all within 0.7 percentage points of each other. This suggests that the estimate of the proportion of dialogue does not differ greatly according to the translation.

Correlation between different measures of length

The correlation between the lengths of individual lines was analysed. First, we plot the correlation between English words and Japanese characters. We use a simple linear model to plot a straight line through the data (though this isn't used for inference).

```
m1 = lm(d$jap_word~d$eng_word)
plot(d$eng_word,d$jap_word,
     xlab = "Number of English words",
     ylab = "Number of Japanese characters")
abline(m1,col=2)
```



The relationship seems strong and linear. A simple Pearson correlation is very high:

```
cor(d$eng_word, d$jap_word)
```

```
## [1] 0.9273341
```

While this is a suitable measure of association, establishing significance requires meeting some assumptions. The data is essentially discrete (counts of characters/ words / syllables), and has many ties. So we'll use Kendall rank correlation:

```
cor.test(d$eng_word,d$jap_word, method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  d$eng_word and d$jap_word
## z = 65.256, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.7367588
```

The correlation is highly significant.

The plot shows two extreme outliers. These are extended speeches by two characters, explaining lore or game rules. However, the correlation is almost identical without these:

```
cor.test(d$eng_word[d$eng_word<200],
         d$jap_word[d$eng_word<200], method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  d$eng_word[d$eng_word < 200] and d$jap_word[d$eng_word < 200]
## z = 65.21, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.7364721
```

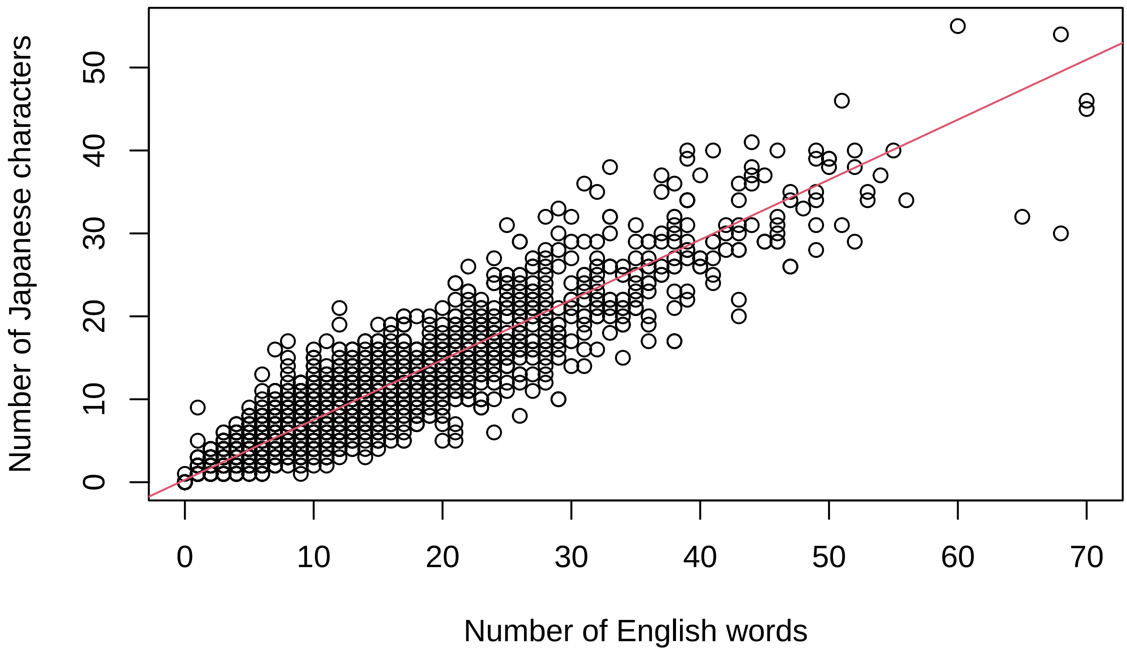
99% of the data has an English word length of 43 or less. But the correlation remains very similar when looking at just these:

```
wordThreshold = quantile(d$eng_word,c(0.99))
cor.test(d$eng_word[d$eng_word <= wordThreshold],
         d$jap_word[d$eng_word <= wordThreshold],
         method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  d$eng_word[d$eng_word <= wordThreshold] and d$jap_word[d$eng_word <= wordThreshold]
## z = 64.342, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.7310396
```

The correlation between the number of Retranslated English words and the number of Japanese characters is higher, which fits with the Retranslation being generally closer in content.

```
m2 = lm(d$jap_word~d$re_word)
plot(d$re_word[d$eng_word <= wordThreshold],
     d$jap_word[d$eng_word <= wordThreshold],
     xlab = "Number of English words",
     ylab = "Number of Japanese characters")
abline(m2,col=2)
```

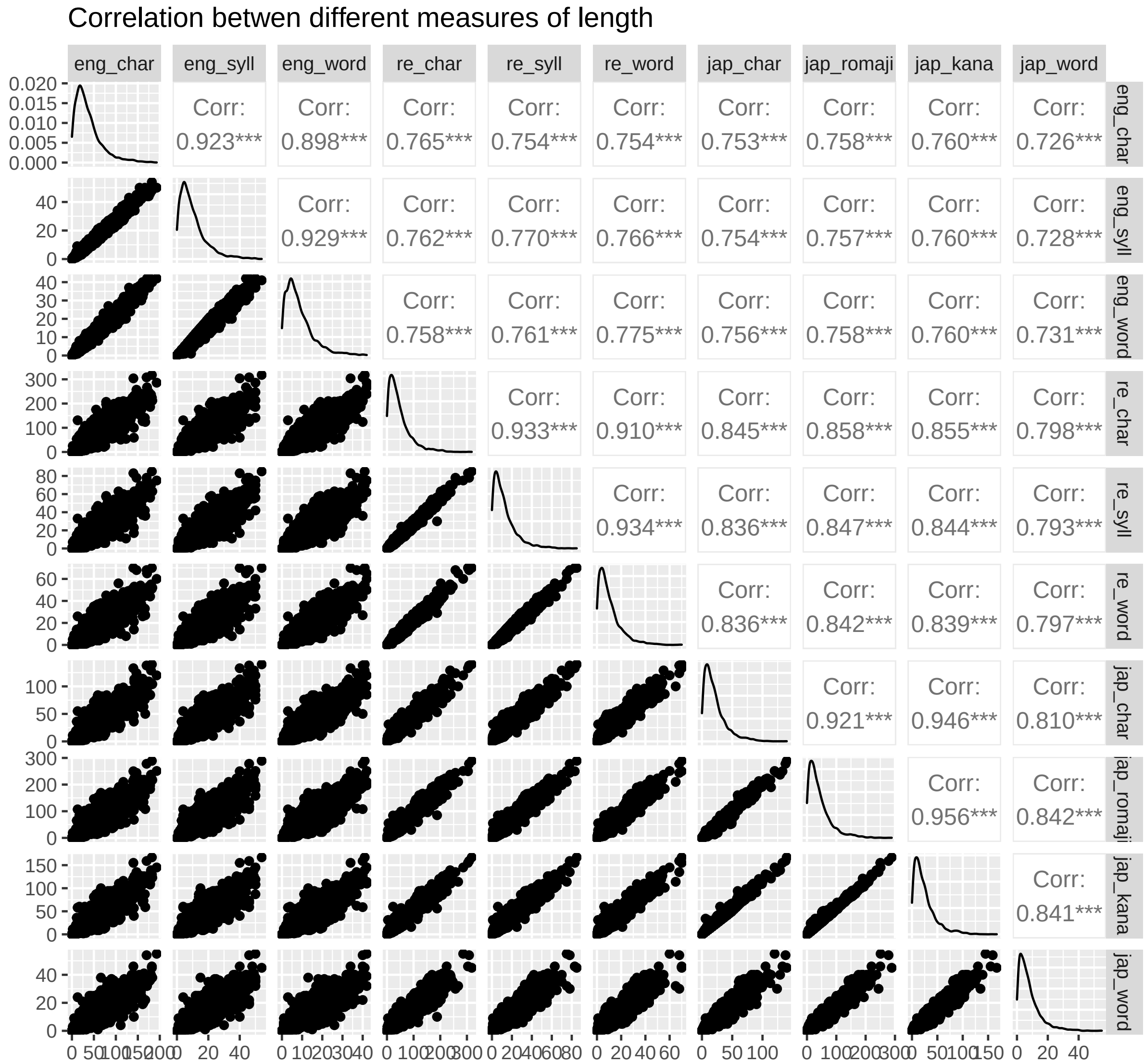



```
cor.test(d$re_word,d$jap_word, method = "kendall")
```

```
##
##  Kendall's rank correlation tau
##
## data:  d$re_word and d$jap_word
## z = 71.297, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##       tau
## 0.8012162
```

Plot correlation between all measures:

```
corK = function(data,mapping,...){
  ggally_cor(data,mapping,method="kendall")
  }
ggpairs(d[d$eng_word< wordThreshold,measures],
        upper = list(continuous = corK),
        title="Correlation betwen different measures of length")
```



The measures are very similar, though the Japanese characters are more highly correlated with the Retranslation measures than the original English measures.

Gender bias during translation

Even though the estimates are very similar, the translation may still cause a confound if it is done differently according to gender. For example, if male dialogue is given longer translations than female dialogue. Although this process might be evidence of a gender bias in itself, it would suggest that part of the bias was in translation rather than the original authorial intent.

One way to test this is to build a statistical model that tries to predict the amount of dialouge for a given line in English from two sources of information: the amount of dialogue for a given line in Japanese, and the gender of the speaker of that line. The models below use Poisson regression, to reflect the fact that the data are counts that are highly skewed.

The first model predicts the number of English words from the number of Japanese characters. Since the data are counts and highly skewed, we should use a Poisson regression:

```
mI1.poisson = glm(
  eng_word ~ jap_char,
  family = "poisson",
  data = d[d$eng_word < wordThreshold,])
summary(mI1.poisson)
```

```
##
## Call:
## glm(formula = eng_word ~ jap_char, family = "poisson", data = d[d$eng_word <
##     wordThreshold, ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -9.5988  -0.9977  -0.1808   0.7328   5.3504
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.4769235  0.0086034   171.7   <2e-16 ***
## jap_char    0.0245583  0.0001893   129.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 19502  on 3866  degrees of freedom
## Residual deviance:  6868  on 3865  degrees of freedom
## AIC: 21016
##
## Number of Fisher Scoring iterations: 4
```

```
pseudo.R2 = cor(d[d$eng_word < wordThreshold,]$eng_word,
    predict(mI1.poisson))^2
pseudo.R2
```

```
## [1] 0.8100734
```

The second model adds the gender of the character. We use sum coding for the contrasts for the gender variable. The results below suggest that, overall, female characters receive fewer words on average than male characters. This is the general bias we demonstrate in the corpus as a whole.

```
d$gender = relevel(factor(d$gender), "male")
contrasts(d$gender) = contr.sum(2)/2
mI2 = update(mI1.poisson, ~.+gender)
summary(mI2)
```

```
##
## Call:
```

```
## glm(formula = eng_word ~ jap_char + gender, family = "poisson",
##     data = d[d$eng_word < wordThreshold, ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -9.6640  -1.0158  -0.1615   0.7204   5.3105
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.4754680  0.0086397 170.778   <2e-16 ***
## jap_char    0.0245364  0.0001897 129.361   <2e-16 ***
## gender1     0.0221487  0.0112772   1.964   0.0495 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 19501.8  on 3866  degrees of freedom
## Residual deviance:  6864.1  on 3864  degrees of freedom
## AIC: 21014
##
## Number of Fisher Scoring iterations: 4
```

The third model adds an interaction between the number of Japanese characters and gender. This helps test whether there is a bias for the translations of male characters to be different in length compared to translations of female characters.

```
mI3 = update(mI2, ~.+jap_char:gender)
summary(mI3)
```

```
##
## Call:
## glm(formula = eng_word ~ jap_char + gender + jap_char:gender,
##     family = "poisson", data = d[d$eng_word < wordThreshold,
##         ])
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -9.7670  -1.0110  -0.1602   0.7219   5.3041
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      1.4762117  0.0087121 169.445   <2e-16 ***
## jap_char         0.0245082  0.0001946 125.938   <2e-16 ***
## gender1          0.0134230  0.0174241   0.770    0.441
## jap_char:gender1 0.0002555  0.0003892   0.657    0.512
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 19501.8  on 3866  degrees of freedom
## Residual deviance:  6863.7  on 3863  degrees of freedom
## AIC: 21015
##
## Number of Fisher Scoring iterations: 4
```

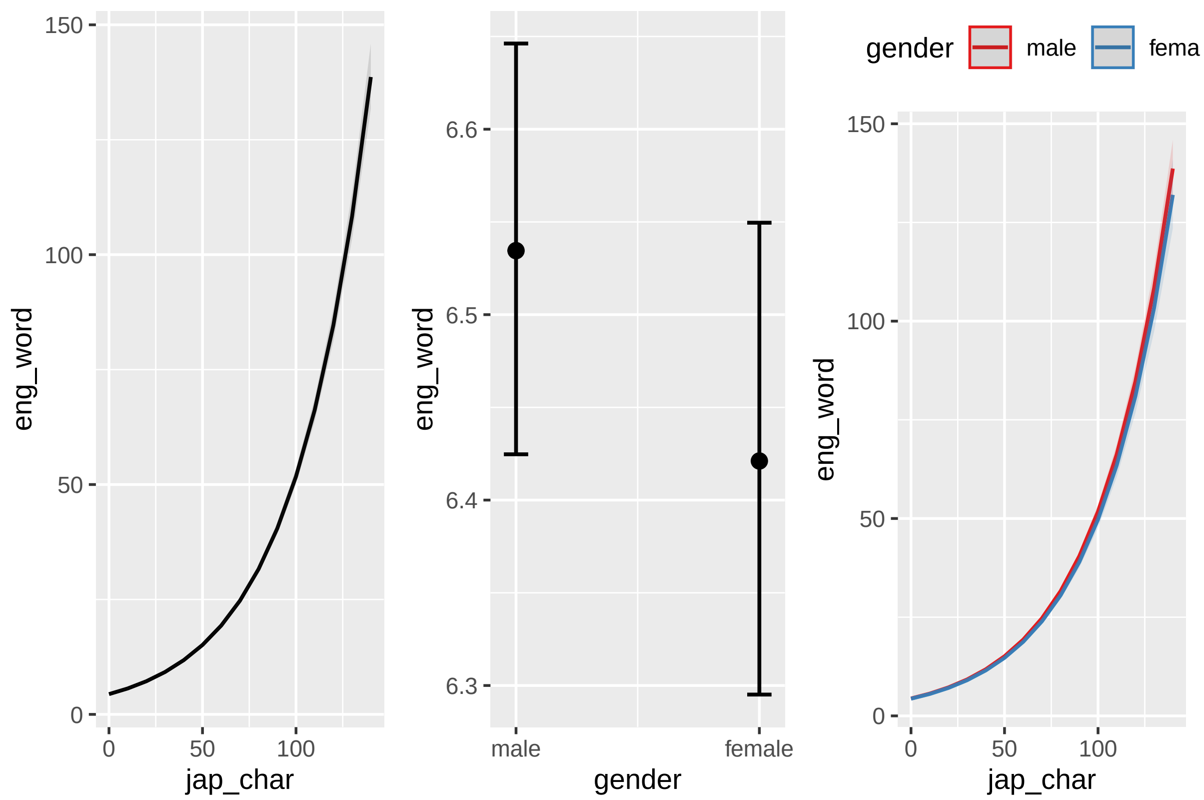
The third model significantly improves the fit compared to the previous two:

```
lrtest(mI1.poisson, mI2, mI3)
```

```
## Likelihood ratio test
##
## Model 1: eng_word ~ jap_char
## Model 2: eng_word ~ jap_char + gender
## Model 3: eng_word ~ jap_char + gender + jap_char:gender
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   2 -10506
## 2   3 -10504  1 3.8631    0.04936 *
## 3   4 -10504  1 0.4314    0.51132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Plot the estimates:

```
p1 = plot_model(mI3,'pred')
p2 = plot_model(mI3,'int') +
  theme(legend.position = "top")
ggarrange(p1[[1]] +ggtitle(element_blank()),
          p1[[2]]+ggtitle(element_blank()),
          p2+ggtitle(element_blank()), nrow=1)
```



This suggests that there is a bias in XXX

We can repeat this analysis for predicting the retranslated words. This comes to the same conclusion:

```
mI1.Re = glm(re_word ~ jap_char,
            family= "poisson",
  data = d[d$eng_word <= wordThreshold,])
mI2.Re = update(mI1.Re, ~.+gender)
mI3.Re = update(mI2.Re, ~.+jap_char:gender)
summary(mI1.Re)
```

```
##
## Call:
## glm(formula = re_word ~ jap_char, family = "poisson", data = d[d$eng_word <=
##     wordThreshold, ])
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -11.9308   -1.1368   -0.1191    0.8258    4.0204
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.5811757  0.0078498   201.4   <2e-16 ***
## jap_char    0.0272948  0.0001625   168.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 28298.0  on 3866  degrees of freedom
## Residual deviance:  7705.6  on 3865  degrees of freedom
## AIC: 22352
##
## Number of Fisher Scoring iterations: 4
```

```
summary(mI2.Re)
```

```
##
## Call:
## glm(formula = re_word ~ jap_char + gender, family = "poisson",
##     data = d[d$eng_word <= wordThreshold, ])
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -12.0725   -1.1454   -0.1247    0.8146    3.9693
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.5785909  0.0078898 200.079  < 2e-16 ***
## jap_char    0.0272618  0.0001629 167.391  < 2e-16 ***
## gender1     0.0366108  0.0102129   3.585 0.000337 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 28298.0  on 3866  degrees of freedom
## Residual deviance:  7692.7  on 3864  degrees of freedom
## AIC: 22341
##
## Number of Fisher Scoring iterations: 4
```

```
summary(mI3.Re)
```

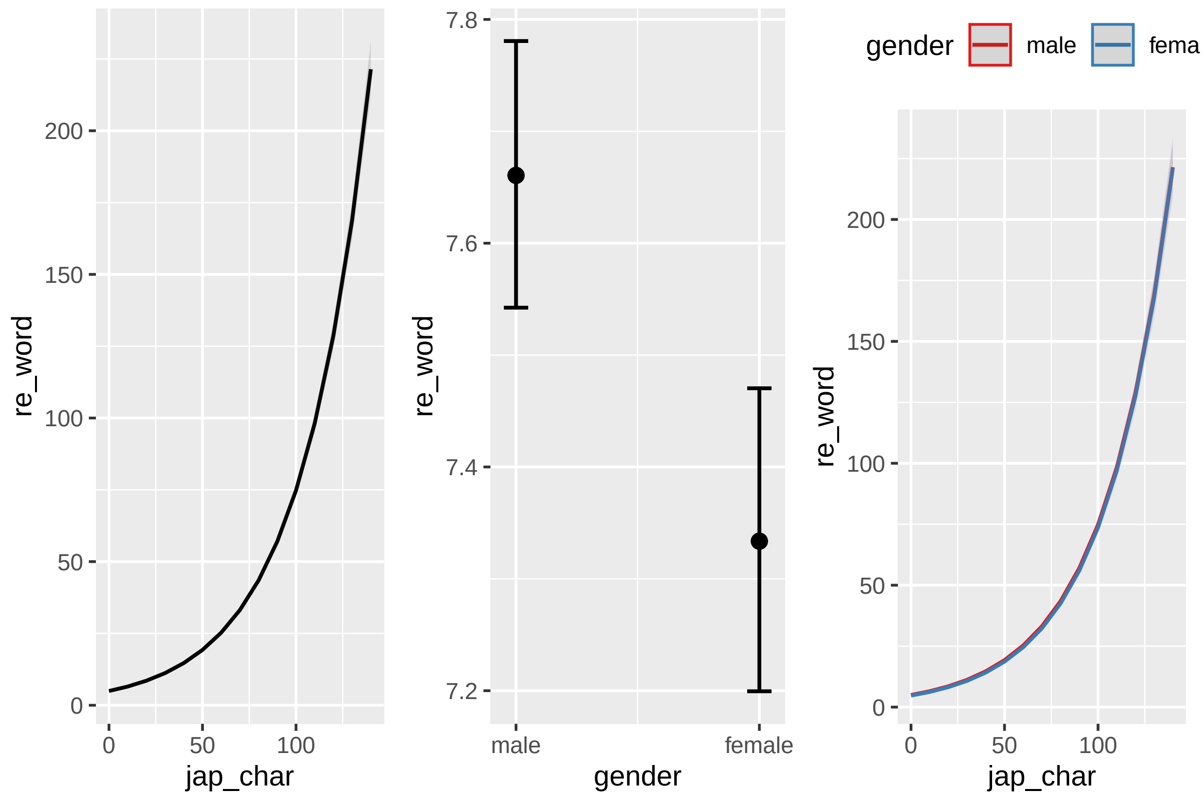
```
##
## Call:
## glm(formula = re_word ~ jap_char + gender + jap_char:gender,
##     family = "poisson", data = d[d$eng_word <= wordThreshold,
##         ])
##
## Deviance Residuals:
```

```
##      Min        1Q    Median        3Q       Max
## -11.9011   -1.1367   -0.1283    0.8221    3.9592
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        1.5774989  0.0079649 198.056  < 2e-16 ***
## jap_char           0.0272991  0.0001668 163.685  < 2e-16 ***
## gender1            0.0490532  0.0159298   3.079  0.00207 **
## jap_char:gender1 -0.0003398  0.0003336  -1.019  0.30836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 28298.0  on 3866  degrees of freedom
## Residual deviance:  7691.7  on 3863  degrees of freedom
## AIC: 22342
##
## Number of Fisher Scoring iterations: 4
```

```
lrtest(mI1.Re, mI2.Re, mI3.Re)
```

```
## Likelihood ratio test
##
## Model 1: re_word ~ jap_char
## Model 2: re_word ~ jap_char + gender
## Model 3: re_word ~ jap_char + gender + jap_char:gender
##   #Df LogLik Df   Chisq Pr(>Chisq)
## 1   2 -11174
## 2   3 -11168  1 12.8836  0.0003315 ***
## 3   4 -11167  1  1.0366  0.3086234
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
p1R = plot_model(mI3.Re,'pred')
p2R = plot_model(mI3.Re,'int') +
  theme(legend.position = "top")
ggarrange(p1R[[1]] +ggtitle(element_blank()),
          p1R[[2]]+ggtitle(element_blank()),
          p2R+ggtitle(element_blank()), nrow=1)
```


re_word
jap_char
0
50
100
150
200
re_word
gender
male
female
7.2
7.4
7.6
7.8
gender
male
fema
re_word
jap_char

Translator comments

The retranslator left comments for various lines. These may be indications that there was some issue with the initial English translation. It might be a problem for the study if there were more issues with translation for female characters than for male characters. This might suggest that the female dialogue was being distorted.

Look at the proportion of lines with translator comments by gender:

```
tProp = prop.table(
  table(d$re_comment>0,
        d$gender),margin = 2)
rownames(tProp) = c("No comment", "Comment")
round(tProp*100,1)
```

```
##
##              male female
##   No comment 92.6   94.6
##   Comment     7.4    5.4
```

```
chisq.test(table(d$re_comment>0,d$gender))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(d$re_comment > 0, d$gender)
## X-squared = 5.5361, df = 1, p-value = 0.01863
```

Male character dialogue have significantly more comments than female character dialogue.

Below we use a permutation test to see if the *length* of the comments (where there are comments, counted in English words) differs between genders.

```
perm = function(){
  diff(tapply(sample(
    d$re_comment[d$re_comment>0]),
    d$gender[d$re_comment>0], mean))
}
trueDiff = diff(tapply(
  d$re_comment[d$re_comment>0],
  d$gender[d$re_comment>0], mean))
n = 1000
permDiff = replicate(n,perm())
perm.p = sum(permDiff <= trueDiff)/n
perm.z = (trueDiff - mean(permDiff))/sd(permDiff)
```

The comments are not significantly different in length for female and male characters ($p = 0.371$, $z = -0.3559$).

In sum, it seems like there are not greater problems in translations for male or female character dialogue.

Closer analysis of lines

We can use the correlation above to identify lines that differ from the expected lengths. These lines be candidates for helping explain the differences. For the analysis below, this was done with a copy of the data that included the raw dialogue. However, the files provided in the archive do not include the raw dialogue, so the code below is only for illustration.

```
# Measure how far each point is from the
# value expected by the regression
# (the residual)
d$diff = resid(m1)

# Focus on the shorter lines
shorts = d[d$eng_word<50,]
# Look at the largest outliers:
View(head(shorts[order(shorts$diff),]))
View(tail(shorts[order(shorts$diff),]))
```

This helped find specific lines where deviation was observed to be the highest between the Japanese characters and the English text. The text was then analysed from a critical discourse perspective to account for how the dialogue has been modified in translation and whether it suggests any gender bias. Various examples are discussed below.

Emotional language

Example 1 is a line of dialogue from Lucca, a female character that joins the player's party. She is reacting to a village that has been devestated.

> ## Example 1
>
> **Lucca**: なんて事…… この時代まで……。
>
> **Translation**: This is so depressing. No era is safe, and there isn't much of a future to look forward to.
>
> **Retranslation**: It can't be…… This era too……

The line is longer in the English translation compared to the original Japanese. The original script lacks any indications of gender and its shortness conveys the unspeakable horror Lucca is faced with. The English translation has increased Lucca's dialogue, adding emotional descriptions. Compared to the minimal Japanese dialogue which conveyed no emotional attachment and both sentences were elliptical,the English version adds more overt emotionality. This conforms to the belief that women express emotions more than men (Kring & Gordon, 1998; Jansz, 2000; Timmers et al., 2003).

Example 2 comes from Marle, a female princess who joins the player's party.

> ## Example 2
>
> **Marle**: こわかった…… いしきがないのに、冷たい所にいるのがわかるの。
>
> **Translation**: It was awful… I can't recall it all… I was somewhere cold, dark…and lonely. Is that what it's like to…die?
>
> **Retranslation**: I was scared…… Even though I'm not conscious, knowing I'm in a cold place…..

Something similar occurs in example 2 where Marle's dialogue is also expanded upon in English, with additional adjectives and a question at the end. The English switches to the third person when describing the trauma behind the experience whereas the Japanese dialogue has Marle admit she was scared, unable to elaborate further on her fear of death.

In example 3, the English translation includes wordplay ("permanent pit stop") and a comment on the character's appearance. This is a more humorous and direct line than the original Japanese, which is closer to the retranslation, which appears more dismissive.

Example 3

Marle: はやいだけってのもね……。

Translation: Someone ought to tell him to take a permanent pit stop. Look at that hair!

Retranslation: All he is is fast, too……

In example 4, there are two differences. The original Japanese begins with a hesitation or stutter ("そ、そうだよ", retranslated as "That, that's"), while the English translation does not. Secondly, the original Japanese uses the equivalent of "let's" with the present tense (like the Retranslation). In contrast, the original English translation does not include a stutter and is an immediate call to action, with the modal "must" denoting necessity, which is a stronger demand on interlocutor's negative face.

Example 4

Marle: そ、そうだよ！変えちゃおう！　クロノが私を助けてくれたみたいに！

Translation: There's only one thing we can do! We must change history! Just like Crono did when he saved me!

Retranslation: That, that's it! Let's go ahead and change it! Like when Crono saved me!

These kind of changes may have affected the player's interpretation of the character. Anecdotally, in a Japanese language wiki (written presumably by fans who played in Japanese), she is described as "好奇心旺盛な性格で" (having a curious personality, see here, while in the English language *Fandom* Chrono Trigger wiki, Marle is described as "vivacious, optimistic and strong-willed" (see here: https://chrono.fandom.com/wiki/Marle).

Example 5 is a case of the opposite effect of translation. At the end of the game, the developers appear as creatures in the debug room as an Easter egg for players who finished the game. The characters share jokes or information from the game's development. Fumi Nakashima was a member of the development team and the only member we could identify as female.

Example 5

Fumi Nakashima: どもども、中プウだケロ。ワールドマップのちびキャラいたケロ。ルッカは本を読んでるケロ。クロノは急がせてるケロ。わかるケロ？

Translation: Someone kiss me!

Retranslation: Thank you, thank you, (?), ribbit. There's chibi characters on the world map, ribbit. Lucca reads a book, ribbit. Crono urges you on, ribbit. Know that, ribbit?

Fumi appears as a frog and, in Japanese, drops a few hints regarding character models and animations, as she was the one who handled character graphics. However, in the original English translation, her dialogue is reduced to "Somebody kiss me!", a reference to the folk tale of a frog turning into a prince with a kiss, and reminiscent of the "smooch of victory" trope which s part of the trope that portrays women as rewards. This omits the reference to her actual work on the game, and brings in a sexual or amorous dimension where there was none in the original.

Information giving

There are a few cases of minor female NPCs who give more information in the original Japanese than in the translation. In example 6, the appearance of Marle at the Millenial Fair is foreshadowed:

Example 6

Young woman: ねえ、知ってる？　ここガルディア王国ができて、今年で 1000 年。現ガルディア王つまり、今の王様だけど 33 代なのよ。で、その王様のなやみのタネは一人娘の王女様。なんでもすっごい、おてんばなんですって。たぶん今ごろは、お城のなかで『あたしも、お祭りいきたーい !』なんて、おおさわぎしてるんじゃないかしらね。

Translation:Hard to believe Guardia is now 1000 years old, and our King is the XXXIII descendant to the throne! But how can he rule a kingdom when he can't even control his own daughter?!

Retranslation: Hey, did you know? It's 1000 years this year since Guardia Kingdom here was formed. The current King Guardia, our king now in other words, is the 33rd. And the root of that king's worries is the princess, his only daughter. They say she's an incredible tomboy. I wonder if she isn't making a big fuss in the castle about now, something like "I wanna go to the festival too!"",

In the Japanese, Marle is depicted as a source of concern for the King and depicted as an "incredible tomboy". However, this is cut from the original translation. Instead, we get a character who questions the leadership of a man based on his inability to control a woman.

Examples 7-9 are cases of female characters giving the player more information in the Japanese version than the original English translation:

Example 7

Woman: 南の地下水道をぬけた大陸には行かない方がいいわ。大災害の源……『死の山』があるから。もっとも地下水道に巣くっている強力なミュータントは、さすがにあなた達でも……

Translation: You can reach the continent to the south through the Sewer Access, but stay off of «Death Peak.

Retranslation": "You'd better not go to the continent past the underground waterway to the south. The great disaster's source……"Death Mountain" is there. But then, I'm sure even for you guys the powerful mutants lurking in the underground waterway are……"

Example 8

young woman: ハイパーほしにく？ああ、遠い祖先が作ったらしいけどあたいはレシピがわかんないねえ。今じゃパレポリの名物になってるよ。

Translation: Jerky? Seems one of his ancestors first made it, but I don't know the recipe.

Retranslation: Hyper Dried Meat? Oh right, I hear a distant ancestor made it, but I don't know the recipe. It's Parepoley's specialty now.

The English translation in example 9 conveys roughly the same narrative information as the Japanese, though in a shorter space. However, it omits a comment that reveals the young woman's opinion of the events.

Example 9

young woman: サイラス様には、グレンという名前の親友がいたんですって！そのグレンという人は、サイラス様が魔王に殺された後、伝説の剣を手に魔王軍と戦ったそうよ。男の友情よねェ、うるうる。

Translation: Cyrus's best friend, Glenn, used a legendary sword to beat the Magus's troops.

Retranslation: Cyrus-sama had a best friend named Glenn! They say that Glenn took up the legendary sword and fought Magus's army after Cyrus-sama was killed by Magus. That's male camaraderie, how moving.

There are some cases where more information is added during the original translation, increasing the amount of female dialogue. However, this does not necessarily work against gendered tropes. In example 10, the translation adds the line "I forgot" to Lucca's dialogue. This makes her seem more absent-minded, rather than portraying her as suggesting a clever solution.

Example 10

Lucca: たーいむ　マシーーン [heart] があるじゃない！オーッホッホッホ！

Translation: I forgot! We have a Time Machine! Nya ha ha!

Retranslation: We've got a Time Machiiinnne{heart}! Oh ho ho ho!

There are some lines that have been changed due to references to the female body. In example 11, Marle has received a letter that mentions the possibility of having children in the future. Ayla, a woman from prehistory whose speech is rendered as very "primitive", tries to encourage her and give her advice. The original Japanese refers to Marle's breasts, but this is removed in the Original English. The retranslator explains: "It's clear from the concept art that Marle does have breasts, but of a modest size (unlike Ayla and most other anime females), and her clothing is fairly baggy as well. It seems that Ayla is concerned that Marle may have trouble nursing her future kids, and doesn't realize that saying this could be considered embarrasing or insulting."

Example 11

Ayla: すだつ！　ねねする！子供うむ！　おっぱいやる！そしてまた　子がすだつ！オマエだいじょぶか？おっぱいないな……。

Translation: Leave nest! Have baby! Baby grow big! Leave nest too! Sure you ready leave nest? Not too big yet.

Retranslation: Leave nest! Sleep! Bear kids! Give boobs! And then kids leave nest again! You be okay? Not have boobs……"

Sexuality

Although not directly related to gender bias, we note that there are a few cases where references to a character's sexuality have been changed in the original English translation. In example 12 and 13, the suggestion that Ayla and Lucca might be sexually attracted to women is ommitted in the translation. Just before example 12, Alya has just said that she "likes" Crono, which was interpreted by Marle and Lucca as saying that she is attracted to Chrono. When Alya says that she likes men and women, Lucca denies that she feels the same way, with the implication that Lucca has interpreted Alya as being bisexual.

Example 12

Ayla: お前達も　強い。エイラ　強い者　好き。男でも　女でも。

Marle: な〜んだ、そういう事か。

Lucca: わ、私は、そのケはないわよ！

Translation:

Ayla: You strong too. Ayla respect strong people. Men and women.

Marle: Oh, brother…

Lucca: Where have they been keeping her?

Retranslation:

Ayla: Yous strong too. Ayla like strong people. Even if man, even if woman.

Marle: What's THAT supposed to mean?

Lucca: I, I'm not interested in that sort of thing!

In example 13, Lucca and Marle are reviewing characters they met during the game (a rare case of females objectifing male characters). Lucca once again denies that she is sexually attracted to women.

Example 13

Marl: 「へへっ、カッコいいのをそろえたわ！

Lucca: 「さすがね。私もうドキドキしちゃう！

Marl: 「彼だけには、どっかんピストル使わなかったでしょ。彼の名前はピーター。でも女なの。

Lucca: 「ぐらっ！私、その気はないわよ！!

Translation:

Marl: Goodness! VERY nice scenery!

Lucca: But of course, my dear!

Marl: I guess you never took a shot at him, right? Say, didn't he just wink at you?

Lucca: RELAX, Marle!!

Retranslation:

Marl: Heh, heh, he's got his coolness all in order!

Lucca: Fitting for a soldier. My heart's already racing!

Marl: He's the only one you didn't use your blast pistol on, right? His name's Peter. But he's a woman.

Lucca: Ghrah! I'm NOT into girls!!

Player interpretations may vary, but the emphasis on "NOT" suggests that Lucca is annoyed that she is having to repeat this again. This is omitted in the original English translation. Also ommitted is Marl's suggestion that Peter is a woman, directly changing the player's interpretation of that character's gender.

In example 14, in Japanese, Lucca speculates that another character is a gay man. This is changed in the original English translation to a comment that does not refer to sexuality.

Example 14

Marle: トマちゃん。あたいのお気に入り！

Lucca: 「ただの酒のみよ。

Marle: 「それに、女好きだったりして。

Lucca: 「ひょっとしたら男好きだったりして！

Translation:

Marle: Hi, Toma! Now HE'S definitely my type!

Lucca: Aw, he's just a flake.

Marle: Probably has a dozen girlfriends.

Lucca: Actually, I see him as more of an intellectual!

Retranslation:

Marle: Toma-chan. My fave!

Lucca: He's just another heavy drinker.

Marle: Besides, what if he's a woman-chaser?

Lucca: What if maybe he's a man-chaser!

The examples above related to sexuality show differences between the original Japanese and the original English translations. However, both versions have weaknesses: the English translation erases mention of same-sex attraction/bisexuality/homosexuality, whereas the re-translation acknowledges the possibility but doesn't do so particularly positively (it portrays non-straight sexuality as undesirable, and naming someone as gay is used for humor).

Conclusion

This report showed that the measures of dialogue length were highly similar between original, translation and retranslation texts. We showed some individual examples of translations that changed the representation of female characters compared to the original, including where gender tropes were reinforced. However, despite these differences, the main measure of dialogue length differed very little between the original Japanese and either of the Engilsh translations. For example, all of the estimates of the percentage of female vs. male dialogue were all within 1 percentage point of each other. This suggests that issues of translation are unlikely to affect the main quantitative results related to gender bias in the corpus as a whole.

References

Antonijevic, F. (2018). Translation of time: A translation analysis of Chrono Trigger. Student essay. Göteborgs Universitet. https://gupea.ub.gu.se/handle/2077/57305

Kring, A. M.; Gordon, A. H. (1998). "Sex differences in emotion: expression, experience, and physiology". Journal of Personality and Social Psychology. 74 (3): 686–703.

Jansz, J (2000). "Masculine identity and restrictive emotionality". Gender and Emotion. Gender and Emotion: Social Psychological Perspectives. pp. 166–186.

Mangiron, C., O'Hagan, M., & Orero, P. (2014). Fun for all: translation and accessibility practices in video games. Bern: Peter Lang.

Müller Galhardi, R. (2014). Video game and Fan translation: A case study of Chrono Trigger. Fun for All: Translation and Accessibility Practices in Video Games, 175-195.

Sánchez, P. M. (2009). Video game localisation for fans by fans: The case of romhacking. The Journal of Internationalization and Localization, 1(1), 168-185.

Timmers, M., Fischer, A. H., & Manstead, A. S. R. (2003). Ability versus vulnerability: Beliefs about men's and women's emotional behavior. Cognition and Emotion, 17, 41–63.

Williams, M. P. (2014) Chrono Trigger. Boss Fight Books.

14 Analyses of word frequency

Frequency Analyses

Introduction

This report identifies markers of politeness in video game dialogue and tests gender differences in the extent and type of politeness strategies.

During conversation, speakers use various linguistic strategies to avoid "face-threatening" acts: utterances that threaten an individual's independence (like making a demand) or their desire to be liked (like insulting them, Brown & Levinson, 1987). One strategy is 'hedging', the use of linguistic markers that affect the epistemic certainty of the speaker's claims. For example, "Maybe we should go to the shops" is less face-threatening than "We should go to the shops", because it hedges the direct demand on a person's time and provides the interlocutor the ability to suggest a different course of action without directly rejecting the speaker.

Greater use of politeness markers such as hedging has been associated with female speech (Lakoff, 1973; Fishman, 1983; Coates 2003; Holmes, 2013; Mirzapour, 2016). The classic divide is between theories that see female hedging as a form of submissiveness (e.g. Lakoff, 1973), and theories that see it as expressing affiliation (Holmes, 1990; Dixon & Foster, 1997). Various studies also suggest that power relations can trump gender relations (e.g. Mullany, 2004). However, in general, both theories predict that females use more hedging. Empirical studies have supported this in the dialogue of female characters in fiction (Karlsson Nordqvist, 2013; Jan & Rahman, 2020; Weisi, & Asakereh, 2021) though there may be differences in the distribution of particular hedges (Holmes, 1990), and there are also studies that find no significant difference between genders (Nemati & Bayer, 2007; Vold, 2006; Holtgraves & Lasky, 1999). Similar strategies for avoiding face-threatening acts include showing gratitude, polite requests (e.g. use of "please"), and apologising (see Danescu-Niculescu-Mizil, 2013). For video games, we would predict greater use of politeness strategies by female characters compared to male characters.

Polite speakers may also aim to avoid the use of negative words and swearing (Jay & Janschewitz, 2008). There are folk beliefs in Western society that men swear more than women (Coates, 2004), and judgements of the acceptability of swearing vary by the gender of the speaker, the interlocutors, and the context (Mills, 2004; DeFrank, & Kahlbaugh, 2019). However, empirical studies of real conversation show mixed relationships between gender and swearing. Some show that men swear more than women (McEnery & Xiao, 2004), some find no overall difference (Baker, 2014; McEnery, 2006), and others find differences are more based on context, age, and specific swear words (Allan & Burridge, 2006; Gauthier & Guille, 2017).

This makes predictions for the video game dialogue difficult. There are surprisingly few studies of swearing by gender in fiction. Cressman (2009) find that men swear more than women in film, and Coyne (2012) found that male characters used more profanity than female characters in adolescent literature, but only for adult characters. There are also no openly-available corpora that have dialogue from fiction which is tagged for gender at the utterance level. However, based on findings in other parts of the study (female characters have more limited roles, are more likely to have neutral emotions, and less likely to be angry), we would predict that female dialogue in video games includes fewer swear words than male dialogue.

Methods

Politeness strategies are identified automatically using 'Convokit' (Chang et al., 2020, see http://convokit.cornell.edu/). This uses machine learning methods trained on a tagged dataset to count the number of cases of various types of politeness strategy. See the python script `analysis/Analyse_Politeness.py`.

An alternative method was used to detect hedging, from Knight, Adolphs & Carter (2013). They obtain frequencies of a list of key phrases. Here we replicate their method using the R package 'Quanteda' (Benoit et al., 2018).

We compare frequencies using the log likelihood measure (G2, see Dunning et al., 1993; Rayson et al., 2004), as used by e.g. the Lancaster Log-likelihood and effect size calculator, https://ucrel.lancs.ac.uk/llwizard.html.

Load libraries

```
library(quanteda)
library(quanteda.textstats)
library(stringr)
library(rjson)
```

Functions to run log likelihood tests according to the G2 measure.

```
logLikelihood.G2 = function(a,b,c,d){
  c = as.double(c)
  d = as.double(d)
  E1 = c*(a+b) / (c+d)
  E2 = d*(a+b) / (c+d)
  G2 = 2*((a*log(a/E1)) + (b*log(b/E2)))
  return(G2)
}

logLikelihood.test = function(freqInCorpus1, freqInCorpus2, sizeOfCorpus1, sizeOfCorpus2){
  # A single test is done like this:
  # logLikelihood.test(2554, 3468, 110000, 140000)
  G2 = logLikelihood.G2(freqInCorpus1,freqInCorpus2,sizeOfCorpus1,sizeOfCorpus2)
  p.value = pchisq(G2, df=2, lower.tail=FALSE)
  #print(paste("Log Likelihood =",G2, ", p = ",p.value))
  return(data.frame(G2 = G2, p = p.value))
}
```

Load data

Load all texts, split into male and female dialouge, tokenise, and count the total number of words.

```
# Number of lines from mini-sample
miniSampleSize = 1000

textF = c()
textM = c()
textFMini = c()
textMMini = c()
stats = read.csv("../results/generalStats.csv",stringsAsFactors = F)
# Remove alternative measures
stats = stats[stats$alternativeMeasure!="True",]
stats = stats[!is.na(stats$words),]
folders = unique(stats$folder)
for(folder in folders){
  dx = fromJSON(file = paste0(folder,"data.json"))["text"]
  dx = unlist(dx)
  names(dx) = gsub("CHOICE\\.","",names(dx))
  names(dx) = gsub("text\\.","",names(dx))
  js = fromJSON(file = paste0(folder,"meta.json"))

  flines = dx[names(dx) %in% js$characterGroups[["female"]]]
  mlines = dx[names(dx) %in% js$characterGroups[["male"]]]

  textF = c(textF, flines)
  textM = c(textM, mlines)
```

```
  firstLines = dx[names(dx) %in% c(
    js$characterGroups[["male"]],
    js$characterGroups[["female"]])]
  if(length(firstLines)>=miniSampleSize){
    firstLines = firstLines[1:miniSampleSize]
      flinesMini = firstLines[names(firstLines) %in% js$characterGroups[["female"]]]
      mlinesMini = firstLines[names(firstLines) %in% js$characterGroups[["male"]]]

      textFMini = c(textFMini, flinesMini)
      textMMini = c(textMMini, mlinesMini)
  }
}
```

Create data frames, convert to the Quanteda corpus format, tokenise and count the number of words in each sub-corpus:

```
dM = data.frame(
  text = textM,
  group="male",stringsAsFactors = F
)
corpM = corpus(dM)
tokensM = tokens(corpM, remove_punct = TRUE)
maleTotal = sum(ntoken(tokensM))

dF = data.frame(
  text = textF,
  group="female",stringsAsFactors = F
)
corpF = corpus(dF)
tokensF = tokens(corpF, remove_punct = TRUE)
femaleTotal = sum(ntoken(tokensF))
```

Combine male and female corpora into one corpus, tagged with their group:

```
d = rbind(dM,dF)
corpAll = corpus(d)
tokensAll = tokens(corpAll, remove_punct = TRUE)
```

Keyness

The target group is the male corpus and the reference group is the female corpus.

```
dfmat <- dfm(tokensAll)
k = textstat_keyness(dfmat,
        target = dfmat$group=="male",
        measure = "lr",sort = F)
k$maleFreqPerMillion = (k$n_target/maleTotal) * 1000000
k$femaleFreqPerMillion = (k$n_reference/femaleTotal) * 1000000

top = k[order(k$G2,decreasing = T),][1:30,]
bottom = k[order(k$G2,decreasing = F),][1:30,]
```

Top words used more by men than women:

```
knitr::kable(top[,c("feature",
        "maleFreqPerMillion","femaleFreqPerMillion","G2")],
         digits = 0,row.names = F)
```

feature	maleFreqPerMillion	femaleFreqPerMillion	G2
alexander	403	38	839
kupo	377	42	735
ya	314	94	311
yeah	897	530	242
ain't	262	84	239
ye	220	68	209
the	43339	40796	208
em	300	122	194
got	1879	1388	189
eh	230	87	166
ah	589	349	157
gotta	313	147	153
hey	894	601	148
uh	326	161	144
dude	56	3	141
shit	126	36	130
cassima	50	3	123
yer	99	24	122
thou	113	32	120
gonna	566	358	119
yo	44	2	117
hell	280	143	115
sora	176	72	114
alexander's	43	2	108
aye	229	113	102
king	444	277	100
riku	68	13	99
noct	53	7	95
no	4587	4029	93
goin	82	22	89

Top words used more by women than by men:

```
knitr::kable(bottom[,c("feature",
        "maleFreqPerMillion","femaleFreqPerMillion","G2")],
         digits = 0,row.names = F)
```

feature	maleFreqPerMillion	femaleFreqPerMillion	G2
i	27144	30322	-472
husband	29	161	-286
he	3866	4696	-212
thank	727	1104	-207
geth	265	504	-203
mother	224	446	-201
flemeth	2	58	-200
crono	26	118	-179
cloud	130	295	-178
she	1915	2457	-178
skipper	1	51	-175
ajira	0	44	-174
oh	1416	1863	-161
father	282	485	-145
please	801	1118	-139
um	115	245	-130
giggle	1	39	-130
narrating	10	65	-126
think	2153	2621	-121
noel	16	76	-119
inmate	1	36	-116
benezia	7	51	-109
niket	0	28	-108
so	4155	4745	-103
cerberus	197	337	-100
griff	1	30	-99
child	111	220	-98
jeff	1	27	-97
yunie	0	23	-96
habasi	0	25	-96

Write out:

```
write.csv(rbind(top,bottom),"../results/keyness.csv")
```

Politeness

Load the politeness measures, calcualted in `analysis/Analyse_Politeness.py`:

```
pol = read.csv("../results/politeness.csv",stringsAsFactors = F)

politeness = NULL
for(feature in unique(pol$feature)){
  nF = pol[pol$group=="female" & pol$feature==feature,]$count
  nM = pol[pol$group=="male" & pol$feature==feature,]$count
  propF = 1000000 * (nF/femaleTotal)
  propM = 1000000 * (nM/maleTotal)
  ll = logLikelihood.test(nM,nF,maleTotal, femaleTotal)
  politeness = rbind(politeness, data.frame(
    feature=feature, nFemale = nF,nMale = nM,
    nFemalePerMillionWords = propF,
    nMalePerMillionWords = propM,
    G2 = ll[1],
    p = ll[2]
  ))
}
```

Adjust p-value for multiple comparisons:

```
politeness$p = p.adjust(politeness$p,method = "bonferroni")
```

Results:

```
knitr::kable(politeness, digits = 2)
```

feature	nFemale	nMale	nFemalePerMillionWords	nMalePerMillionWords	G2	p
Please	763	1042	381.57	282.24	39.30	0.00
Please_start	1320	1738	660.12	470.75	84.06	0.00
HASHEDGE	31160	52489	15582.88	14217.12	163.18	0.00
Indirect_(btw)	77	150	38.51	40.63	0.15	1.00
Hedges	11636	18263	5819.07	4946.70	185.08	0.00
Factuality	4182	6461	2091.39	1750.02	79.50	0.00
Deference	1751	3492	875.66	945.84	6.99	0.64
Gratitude	3218	4487	1609.30	1215.34	145.03	0.00
Apologizing	2157	3077	1078.70	833.43	82.90	0.00
1st_person_pl.	30883	59572	15444.35	16135.62	39.17	0.00
1st_person	44732	78761	22370.13	21333.13	63.99	0.00
1st_person_start	42191	67291	21099.39	18226.38	548.98	0.00
2nd_person	61570	113053	30790.68	30621.43	1.21	1.00
2nd_person_start	14054	27488	7028.30	7445.37	31.09	0.00
Indirect_(greeting)	1707	3773	853.66	1021.95	38.88	0.00
Direct_question	10733	20667	5367.49	5597.84	12.53	0.04
Direct_start	13949	25928	6975.79	7022.83	0.41	1.00
HASPOSITIVE	77780	142014	38897.18	38465.78	6.24	0.93
HASNEGATIVE	61993	116566	31002.22	31572.96	13.49	0.02
SUBJUNCTIVE	600	1009	300.06	273.30	3.26	1.00
INDICATIVE	720	1202	360.07	325.57	4.53	1.00

Summarise results and write to stats:

```
getStatText = function(feature,femaleDiff,G2,pval,w=F){
  diffx = "+"
  if(femaleDiff<0){
    diffx = ""
```

```
  }
  diffx = paste0(diffx,
                 round(100*femaleDiff),
                 "\\%")
  p = pval
  if(p < 0.001){
    p = "< 0.001"
  } else{
    p = paste("=",round(p,3))
  }
  statText = paste0(diffx,", G2 = ",round(G2,2),", p ",p)
  if(w){
    cat(statText, file=paste0(
      "../results/latexStats/Freq_",feature,
      ".tex"))
  }
  return(statText)
}

getStat = function(X,feature,w=F){
  px = X[X$feature==feature,]

  femaleDiff = (px$nFemalePerMillionWords -
                 px$nMalePerMillionWords) /
                px$nMalePerMillionWords
  statText = getStatText(feature,femaleDiff,px$G2,px$p,w)
  return(statText)
}
```

Many of the results are compatible with females exhibiting more frequent politeness strategies than males. For example, compared to male characters, female characters use:

- More hedging (+18%, G2 = 185.08, $p < 0.001$)
- More gratitude (+32%, G2 = 145.03, $p < 0.001$)
- More apologies (+29%, G2 = 82.9, $p < 0.001$)
- More 'please' (+35%, G2 = 39.3, $p < 0.001$)

No significant differences for:

- Direct questions (-4%, G2 = 12.53, $p = 0.04$)
- Negative words (-2%, G2 = 13.49, $p = 0.025$)

Alternative measure of hedging

Knight, Adolphs & Carter (2013) use a different approach to quantifying hedging. They obtain frequencies of a list of key phrases. Here we replicate their method using Quanteda. Two different methods are used to obtain frequency, one for single-word key phrases, and one for multi-word key phrases.

```
hedges = c(
  "Actually", "Generally", "Likely",
  "Only",  "Really", "Surely",
  "Apparently", "Guess", "Maybe",
  "Partially", "Relatively", "Thing",
  "Arguably", "Necessarily", "Possibility",
  "Possibly", "Roughly", "Typically",
  "Broadly", "Just", "Normally",
  "Probably", "Seemingly", "Usually",
  "Frequently", "Quite"
)

hedgePhrases = c(
  "I think",
  "Kind of",
  "Of course",
  "Sort of",
  "You know")

dfmat <- dfm(tokensAll)
dfmat <- dfm_select(dfmat,pattern=hedges)
freq.hedges = textstat_keyness(
  dfmat,target = dfmat$group=="male", measure = "lr",
     sort = F, correction = "none")
ll = logLikelihood.test(freq.hedges$n_target,freq.hedges$n_reference,maleTotal,femaleTotal)
freq.hedges$G2 = ll$G2
freq.hedges$p = ll$p
freq.hedges$maleFreqPerMillion = (freq.hedges$n_target/maleTotal) * 1000000
freq.hedges$femaleFreqPerMillion = (freq.hedges$n_reference/femaleTotal) * 1000000

getPhraseFrequency = function(w,group){
  corp = corpus(d[d$group==group,])
  # We don't want punctuation between phrase parts
  toks = tokens(corp, remove_punct = FALSE)
  k = kwic(toks, pattern = phrase(c(w)))
  length(k$post)
}

for(w in hedgePhrases){
  freqF = getPhraseFrequency(w,"female")
  freqM = getPhraseFrequency(w,"male")
  ll = logLikelihood.test(freqM, freqF, maleTotal, femaleTotal)
  freq.hedges = rbind(freq.hedges, data.frame(
    feature = w,
    G2 = ll[1],
    p = ll[2],
    n_target = freqM,
    n_reference = freqF,
    maleFreqPerMillion = (freqM/maleTotal) * 1000000,
    femaleFreqPerMillion = (freqF/femaleTotal) * 1000000
  ))
}
```

```r
freq.hedges$sig = freq.hedges$p<0.05
freq.hedges[freq.hedges$sig,]
```

```
##      feature        G2            p n_target n_reference maleFreqPerMillion
## 1      guess 12.413760 2.015516e-03     2346        1119         635.435353
## 5      quite  6.074512 4.796634e-02     1755        1047         475.357649
## 7     really 64.708051 8.888434e-15     4142        2739        1121.898224
## 8   actually 11.392939 3.357799e-03     1032         662         279.526549
## NA      <NA>        NA           NA       NA          NA                 NA
## 21 seemingly  6.113577 4.703852e-02       17           2           4.604604
## 25 partially  6.211820 4.478375e-02        8          13           2.166872
## 26   I think 91.206243 1.566087e-20     2584        1875         699.899809
## 27   Kind of 10.606339 4.975799e-03     1098         697         297.403247
##    femaleFreqPerMillion  sig
## 1            559.603247 TRUE
## 5            523.596604 TRUE
## 7           1369.752719 TRUE
## 8            331.061081 TRUE
## NA                   NA   NA
## 21             1.000185 TRUE
## 25             6.501199 TRUE
## 26           937.673001 TRUE
## 27           348.564310 TRUE
```

```r
names(freq.hedges)[names(freq.hedges)=="n_target"] = "freqMale"
names(freq.hedges)[names(freq.hedges)=="n_reference"] = "freqFemale"

hedgeMPerMillion = sum(freq.hedges$freqMale)/maleTotal * 1000000
hedgeFPerMillion = sum(freq.hedges$freqFemale)/femaleTotal * 1000000

ll = logLikelihood.test(sum(freq.hedges$freqMale),
                   sum(freq.hedges$freqFemale),
                   maleTotal,femaleTotal)

femaleDiffHedge = (hedgeFPerMillion - hedgeMPerMillion) /
                hedgeMPerMillion
```

As with the main method, females are more likely to use hedging (male frequency per million = 1.316×10^4, female frequency per million = 1.397×10^4, +6%, G2 = 62.83, $p < 0.001$)

Swearing

Using keywords identified in TV show dialogue from Bednarek (2019).

```
swears = read.csv("https://gist.githubusercontent.com/tjrobinson/2366772/raw/97329ead3d5ab06160c3c7a
                  stringsAsFactors = F,header = F)
swears = swears[,1]
swears = swears[!swears %in% c("snatch")]
nx = c("hell","dago","ass")
swears[!swears %in% nx]= paste0(swears[!swears %in% nx],"*")

# Specific to games in the corpus
swears = c(swears,"vermin","scum")

# Bedarek
bednarekSwears = c("god",
  "hell",  "damn", "crap", "screw",
  "fuck", "fucktard", "fuckwad", "fucks", "fucking", "butt-fuck",
  "butt-fucking", "fuck-up", "fuckable", "fucked", "pencil-fucked",
  "fucked-up", "ass-fucked", "fucker",
  "motherfucker", "motherfuckers", "motherfucking",
  "bullshit", "dipshit", "shit", "shit-faced", "shit-ass",
  "shitheads", "shits", "shittiest", "shitting", "shitty",
  "damned", "fricking","freaking","frigging",
  "gosh", "heck", "jeez", "shucks")

swears = c(swears,bednarekSwears)
swears = dictionary(list(swears=swears))

dfmat_swears = dfm(tokensAll)
dfmat_swears = dfm_select(dfmat_swears, pattern=swears)
tstat_freq_swears <- textstat_frequency(dfmat_swears, groups = d$group)

swearFreq = tapply(tstat_freq_swears$frequency,tstat_freq_swears$group,sum)
swearFreqPerMillion = (swearFreq / c(femaleTotal,maleTotal)) * 1000000

femaleDiff = (swearFreqPerMillion["female"] -
                 swearFreqPerMillion["male"]) /
                swearFreqPerMillion["male"]

llSwear = logLikelihood.test(swearFreq["male"],
                   swearFreq["female"],
                   maleTotal,femaleTotal)
```

Female characters swear less than male characters (-37%, G2 = 414.06, $p < 0.001$)

Hesitations

```
hesitations = c("um","umm","ummm","ummmm","ummmm","ummmmmm",
                'er','err','errr','errrr',
                "uh","uhh","uhhh", "uhhhh","uhhhhhh",
                "uhhhhhhhhhhhh","uuh","uuuh",
                "uuuuh", "uuuuhh", "uuuuuuuh",
                "ur","ur","urrr")

dfmat_hes = dfm(tokensAll)
dfmat_hes = dfm_select(dfmat_hes, pattern=hesitations)
tstat_freq_hes <- textstat_frequency(dfmat_hes, groups = d$group)

hesFreq = tapply(tstat_freq_hes$frequency,tstat_freq_hes$group,sum)
hesFreqPerMillion = (hesFreq / c(femaleTotal,maleTotal)) * 1000000

femaleDiff = (hesFreqPerMillion["female"] -
                 hesFreqPerMillion["male"]) /
                hesFreqPerMillion["male"]

llHes = logLikelihood.test(hesFreq["male"],
                   hesFreq["female"],
                   maleTotal,femaleTotal)
```

No significant difference in hesitation (-12%, G2 = 12.35, p = 0.002).

Balanced corpus

Re-create corpus with a balanced amount of lines from each game.

```
dM = data.frame(
  text = textMMini,
  group="male",stringsAsFactors = F
)
corpM = corpus(dM)
tokensM = tokens(corpM, remove_punct = TRUE)
maleTotal = sum(ntoken(tokensM))

dF = data.frame(
  text = textFMini,
  group="female",stringsAsFactors = F
)
corpF = corpus(dF)
tokensF = tokens(corpF, remove_punct = TRUE)
femaleTotal = sum(ntoken(tokensF))

d = rbind(dM,dF)
corpAll = corpus(d)
tokensAll = tokens(corpAll, remove_punct = TRUE)
```

Politeness

```
polMini = read.csv("../results/politenessMini.csv",stringsAsFactors = F)

politenessMini = NULL
for(feature in unique(polMini$feature)){
  nF = polMini[polMini$group=="female" & polMini$feature==feature,]$count
  nM = polMini[polMini$group=="male" & polMini$feature==feature,]$count
  propF = 1000000 * (nF/femaleTotal)
  propM = 1000000 * (nM/maleTotal)
  ll = logLikelihood.test(nM,nF,maleTotal, femaleTotal)
  politenessMini = rbind(politenessMini, data.frame(
    feature=feature, nFemale = nF,nMale = nM,
    nFemalePerMillionWords = propF,
    nMalePerMillionWords = propM,
    G2 = ll[1],
    p = ll[2]
  ))
}
politenessMini$p = p.adjust(politenessMini$p,method = "bonferroni")
```

Results:

```
knitr::kable(politenessMini, digits = 2)
```

feature	nFemale	nMale	nFemalePerMillionWords	nMalePerMillionWords	G2	p
Please	60	119	412.97	315.29	2.81	1.00
Please_start	126	154	867.24	408.02	37.59	0.00
HASHEDGE	2663	4746	18329.11	12574.58	232.68	0.00
Indirect_(btw)	11	22	75.71	58.29	0.49	1.00
Hedges	970	1584	6676.39	4196.83	123.98	0.00
Factuality	366	672	2519.13	1780.47	27.47	0.00
Deference	149	341	1025.55	903.48	1.64	1.00
Gratitude	319	480	2195.64	1271.77	54.49	0.00

feature	nFemale	nMale	nFemalePerMillionWords	nMalePerMillionWords	G2	p
Apologizing	299	443	2057.98	1173.73	53.66	0.00
1st_person_pl.	2504	5553	17234.73	14712.74	42.31	0.00
1st_person	3941	7774	27125.43	20597.31	191.89	0.00
1st_person_start	3669	6466	25253.29	17131.74	338.32	0.00
2nd_person	6091	12434	41923.63	32944.03	230.65	0.00
2nd_person_start	1465	3066	10083.42	8123.40	45.08	0.00
Indirect_(greeting)	199	655	1369.69	1735.43	8.91	0.24
Direct_question	1170	2723	8052.97	7214.62	9.74	0.16
Direct_start	1530	3043	10530.81	8062.46	70.35	0.00
HASPOSITIVE	6850	14241	47147.73	37731.70	223.32	0.00
HASNEGATIVE	5249	11280	36128.24	29886.49	125.78	0.00
SUBJUNCTIVE	60	132	412.97	349.74	1.12	1.00
INDICATIVE	85	143	585.04	378.88	9.64	0.17

The effects replicate:

- More hedging (+59%, G2 = 123.98, p < 0.001)
- More gratitude (+73%, G2 = 54.49, p < 0.001)
- More apologies (+75%, G2 = 53.66, p < 0.001)

Except for ‘please’:

- No significant difference for ‘please’ (+31%, G2 = 2.81, p = 1)

Swearing

```
dfmat_swears = dfm(tokensAll)
dfmat_swears = dfm_select(dfmat_swears, pattern=swears)
tstat_freq_swears <- textstat_frequency(dfmat_swears, groups = d$group)

swearFreq = tapply(tstat_freq_swears$frequency,tstat_freq_swears$group,sum)
swearFreqPerMillion = (swearFreq / c(femaleTotal,maleTotal)) * 1000000

femaleDiff = (swearFreqPerMillion["female"] -
                 swearFreqPerMillion["male"]) /
                swearFreqPerMillion["male"]

llSwear = logLikelihood.test(swearFreq["male"],
                   swearFreq["female"],
                   maleTotal,femaleTotal)
```

```
getStatText("swearing (mini)",femaleDiff, llSwear[1], llSwear[2])
```

```
## [1] "-46\\%, G2 = 46.77, p < 0.001"
```

Hesitation

```
dfmat_hes = dfm(tokensAll)
dfmat_hes = dfm_select(dfmat_hes, pattern=hesitations)
tstat_freq_hes <- textstat_frequency(dfmat_hes, groups = d$group)

hesFreq = tapply(tstat_freq_hes$frequency,tstat_freq_hes$group,sum)
hesFreqPerMillion = (hesFreq / c(femaleTotal,maleTotal)) * 1000000

femaleDiff = (hesFreqPerMillion["female"] -
                 hesFreqPerMillion["male"]) /
                hesFreqPerMillion["male"]
```

```r
llHes = logLikelihood.test(hesFreq["male"],
                   hesFreq["female"],
                   maleTotal,femaleTotal)
```

```r
getStatText("hesitations (mini)",femaleDiff, llHes[1], llHes[2])
```

```
## [1] "-12\\%, G2 = 1.6, p = 0.448"
```

References

Allan, K. & Burridge, K. 2006. Forbidden words: Taboo and the censoring of language. Cambridge: Cambridge University Press.

Baker, P 2014. Using Corpora to Analyze Gender. London: Bloomsbury Publishing

Bednarek, M., 2019. 'Don't say crap. Don't use swear words.'–Negotiating the use of swear/taboo words in the narrative mass media. Discourse, Context & Media, 29, p.100293.

Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, Matsuo A (2018). "quanteda: An R package for the quantitative analysis of textual data." *Journal of Open Source Software*, *3*(30), 774. doi: 10.21105/joss.00774

Brown, P., Levinson, S.C. and Levinson, S.C., 1987. Politeness: Some universals in language usage (Vol. 4). Cambridge university press.

Coates, J. 2003. Men Talk. United Kingdom: Blackwell Publishing Ltd.

Coates, Jennifer. 2004. Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language. Edinburgh: Pearson.

Coyne, S.M., Callister, M., Stockdale, L.A., Nelson, D.A. and Wells, B.M., 2012. "A helluva read": profanity in adolescent literature. Mass Communication and Society, 15(3), pp.360-383.

Cressman, D.L., Callister, M., Robinson, T. and Near, C., 2009. Swearing in the cinema: An analysis of profanity in US teen-oriented movies, 1980–2006. Journal of Children and Media, 3(2), pp.117-135.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J. and Potts, C., 2013. A computational approach to politeness with application to social factors. arXiv preprint arXiv:1306.6078.

DeFrank, M. and Kahlbaugh, P., 2019. Language choice matters: When profanity affects how people are judged. Journal of Language and Social Psychology, 38(1), pp.126-141.

Dixon, J.A. and Foster, D.H., 1997. Gender and hedging: From sex differences to situated practice. Journal of psycholinguistic research, 26(1), pp.89-107.

Dunning, Ted. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. Computational Linguistics, Volume 19, number 1, pp. 61-74.

Gauthier, M. and Guille, A., 2017. Gender and age differences in swearing. Advances in swearing research: New languages and new contexts, pp.137-156.

Holmes, J., 1990. Hedges and boosters in women's and men's speech. Language & Communication, 10(3), pp.185-205.

Holmes, J., 2013. Women, men and politeness. Routledge.

Holtgraves, T. and Lasky, B. 1999. "Linguistic power and persuasion." Journal of Language and Social Psychology 18:2 (196-205).

Jan, S. and Rahman, M., 2020. Gender Determines Linguistic Features of One's Speech: Hedging and Interruptions in Male/Female Dialogues in One-Act Plays by male and female playwrights. Academic Journal of Social Sciences (AJSS), 4(4), pp.873-888.

Jay, T. and Janschewitz, K. (2008) The pragmatics of swearing. , Vol. 4 (Issue 2), pp. 267-288.

Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, Cristian Danescu-Niculescu-Mizil. 2020. "ConvoKit: A Toolkit for the Analysis of Conversations". Proceedings of SIGDIAL.

Karlsson Nordqvist, R., 2013. Gender Roles Via Hedging in Children's Films. Undergraduate thesis. https://www.diva-portal.org/smash/get/diva2:691798/FULLTEXT01.pdf

Knight, D., Adolphs, S. and Carter, R., 2013. Formality in digital discourse: a study of hedging in CANELC. In Yearbook of Corpus Linguistics and Pragmatics 2013 (pp. 131-152). Springer, Dordrecht.

Lakoff, R. (1973). Language and woman's place. Language in society, 2(1), 45-79.

McEnery, A. 2006. Swearing in English: Bad language, purity and power from 1586 to the present. London: Routledge

McEnery, A. and Xiao, Z., 2004. Swearing in modern British English: the case of fuck in the BNC. Language and Literature, 13(3), pp.235-268.

Mills, S., (2004) Class, gender and politeness. Multilingua 23:1-2, pp. 171-190.

Mirzapour, F., 2016. Gender differences in the use of hedges and first person pronouns in research articles of applied linguistics and chemistry. International Journal of Applied Linguistics and English Literature, 5(6), pp.166-173.

Mullany, L. (2004). Gender, politeness and institutional power roles: Humour as a tactic to gain compliance in workplace business meetings.

Nemati, A. and Bayer, J.M., 2007. Gender differences in the use of linguistic forms in the speech of men and women: A comparative study of Persian and English. Language in India, 7(9), pp.1-16.

Rayson P., Berridge D. and Francis B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora. In Volume II of Purnelle G., Fairon C., Dister A. (eds.) Le poids des mots: Proceedings of the 7th International Conference on Statistical analysis of textual data (JADT 2004), Louvain-la-Neuve, Belgium, March 10-12, 2004, Presses universitaires de Louvain, pp. 926 - 936.

Vold, E.T., 2006. Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study. International Journal of Applied Linguistics, 16(1), pp.61-87.

Weisi, H. and Asakereh, A., 2021. Hedging devices in applied linguistics research papers: Do gender and nativeness matter?. Glottotheory, 12(1), pp.71-83.

15 Reliability of gender coding

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.