

A4: Project Report

Alison Yao (yy2564) | Fatima Nadeem (fn2093) | Min Jie Kim (mjk9913) | Jun Bin Ho (jbh439)

1. Introduction and Related Work/Why is this important?	2
2. Research Question	2
3. Methodology	4
3.1 Quantitative Analysis	5
Data	5
Analysis	6
3.2 Qualitative Analysis	12
Data	12
Analysis	12
Applying Test 1 to Avengers (2012)	14
Intersection with Racial Minority Representation	15
Applying Test 2 to Crazy Rich Asians (2018)	16
4. Summary of Findings, Implications and Limitations	17
Is the Bechdel Test still useful? Yes, but we should not stop there.	17
More Tests are needed	17
Limitations of Our Tests	18
5. Reflection and Future Work	18

1. Introduction and Related Work/Why is this important?

The [Bechdel Test](#) is a simple and popular measure of feminism in works of fiction. The rules of the test were first defined back in 1985 in a [comic strip](#) featuring two queer women who could not find a movie that:

- 1) Has at least two women in it,
- 2) who talk to each other,
- 3) about something other than a man.

With US adults spending [6 hours a day](#) consuming video-based media, media representation of social groups has an increasingly profound impact on the formation of prejudices and opinions about social groups. In her influential article [Gendered Media](#), Julia T. Wood defines three ways in which media “perpetuate unrealistic, stereotypical, and limiting perceptions” about women:

- 1) Women are underrepresented, which falsely implies that men are the cultural standard and women are unimportant or invisible.
- 2) Second, men and women are portrayed in stereotypical ways that reflect and sustain socially endorsed views of gender.
- 3) Third, depictions of relationships between men and women emphasize traditional roles and normalize violence against women.

The degree to which each of these factors occurs in the media is difficult to measure, but the Bechdel Test offers us a very useful measure of how women and their agency are sidelined in fiction (Theme 1) and has been praised by critics as “[a good way to ensure inclusivity and a strong, dominant female presence rather than the typical passiveness that requires a female’s character to depend on a male character for development](#).” The test has been used in various social studies, such as in investigating the [correlation](#) between the profit of movies and whether the respective movies have passed the Bechdel Test.

However, the Bechdel Test does not measure themes 2 and 3, which describe highly nuanced and layered relations between the genders. The test has also been criticized for ignoring how characters “should be validated by the content of their character and their drive to accomplish clear, female-centric goals.” Critics have also commented that “the test’s utility has been elevated way beyond the original intention”, and it should not be viewed as a benchmark for films.

2. Research Question

Encountering such mixed reviews of the Bechdel Test, we wanted to design our research questions to acknowledge the success and ups of the Bechdel Test, while also navigating the shortcomings of the test. We hope to learn about methods or proxies to measure the human-centered issue of representation, which requires us to think about how we can *qualify* how movies depict women. We also hope to explore how we can leverage other quantitative

benchmarks as a tool to “fill the gaps” of the Bechdel test. Therefore, this report will explore the questions of: How are movies performing based on the Bechdel Test compared to other off-screen metrics, what are its shortcomings and what additional metrics or revisions we can take to better measure feminism in films and filmmaking?

3. Methodology

We approach our assessment of the Bechdel Test in two steps:

1. Quantitatively, we assess:

- a. *How do Bechdel Test Scores change over time?*

If Bechdel Test Scores are rising dramatically over time, we may take this as a sign of improving female representation in movies, but it might also indicate the diminishing relevance of the test in disambiguating good female representation from bad *within* the set of films that have passed. We want to investigate if the test has held up as a robust standard, or if the standard has become outdated as society's standards have progressed.

- b. *How do Bechdel Test Scores compare with other benchmarks of off-screen representation?*

Strong correlations between quantitative benchmarks of off-screen representation and Bechdel Test scores might support the utility of the Bechdel Test as a proxy measure for female agency in film and filmmaking. Specifically, we investigate what films that pass the Bechdel Test have in common with each other, and do the same for films that fail it. For instance, does a film passing the Bechdel Test usually mean that its off-screen representation also follows the feminist agenda?

2. Qualitatively, we build upon the quantitative analysis to

- a. Identify how to assess representation among films that have passed the Bechdel Test by reviewing criticisms by viewers and critics. In particular, we look at common contradictions of the Bechdel Test:

- i. Negative Contradictions: Films that have passed the test but have been criticised for negative female representation. From these movies, we can identify the shortcomings of the Bechdel Test in measuring negative representation.
 - ii. Positive Contradictions: Films that are considered Feminist or that have been praised for positive female presentation. From these movies, we can identify which forms of positive representation the test fails to measure.

This is particularly important in relation to Point 1 under Quantitative Analysis—how can we assess the quality of representation between films that have passed the test and hence, create a more robust and contemporary measure?

- b. Revise the Bechdel Test with new metrics to address the above, as well as bridge the gap between on and off-screen representation (since the Bechdel Test only assesses the former).

3.1 Quantitative Analysis

Data

Our data is collected from several sources. The main one is a website called [Bechdel Test Movie List](#). Any movie viewer can sign up to this website and give scores to any movie as long as they provide reasonable comments to justify their scores. Other users can re-examine a movie and overturn the score if they can specify which characters or scenes of a movie violate the Bechdel Test. On one hand, we are able to access the Bechdel Test scores via the website [API](#). The human-labeled Bechdel score ranges from 0 to 3. If the score is 3, the movie passes the Bechdel Test. Otherwise, it fails. We can use an API call to save the data as a CSV file. There are four methods to get the data: (1) get movies by IMDb id, (2) get movies by title, (3) get all movie ids, and (4) get all movies. Since we are analyzing all movies available, we will use method (4) to get the data on the release year, rating (Bechdel score), imdbid, id (unique website id), and title of all movies rated on this website.

	year	rating	imdbid	id	title
0	1874	0	3155794	9602	Passage de Venus
1	1877	0	14495706	9804	La Rosace Magique
2	1878	0	2221420	9603	Sallie Gardner at a Gallop
3	1878	0	12592084	9806	Le singe musicien
4	1881	0	7816420	9816	Athlete Swinging a Pick

Unfortunately, if we attempt to get all movies, the information on each movie is not going to be as abundant as getting each movie by IMDb id. Method (1) tells us if the movie has been approved, its title, year of release, Imdb ID, unique website ID, the date of entry into the dataset, ID of the submitter, the actual Bechdel score, and whether the submitter considered the rating is dubious or not. Therefore, it took us 7 hours to loop through all the movie IMDb id's and expand the original dataset into the current version.

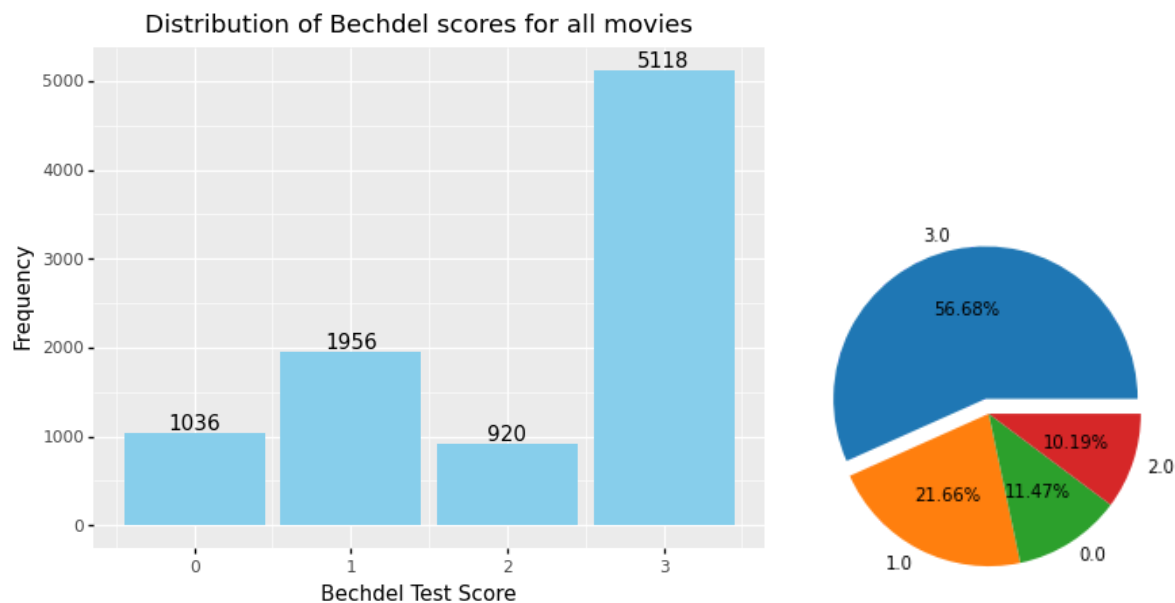
	title	year	rating	dubious	imdbid	id	submitterid	date	visible
0	Passage de Venus	1874.0	0.0	0.0	3155794.0	9602.0	18880.0	2021-04-02 20:58:09	1.0
1	La Rosace Magique	1877.0	0.0	0.0	14495706.0	9804.0	19145.0	2021-05-11 00:11:22	1.0
2	Sallie Gardner at a Gallop	1878.0	0.0	0.0	2221420.0	9603.0	18882.0	2021-04-03 02:25:27	1.0
3	Le singe musicien	1878.0	0.0	0.0	12592084.0	9806.0	19151.0	2021-05-11 23:38:54	1.0
4	Athlete Swinging a Pick	1881.0	0.0	0.0	7816420.0	9816.0	19162.0	2021-05-13 01:32:14	1.0
...
9324	Guilty, The	2021.0	1.0	0.0	9421570.0	10079.0	19599.0	2021-10-10 20:16:13	1.0
9325	Till Death	2021.0	0.0	0.0	11804152.0	10083.0	19613.0	2021-10-18 19:58:48	1.0
9326	Last Duel, The	2021.0	3.0	1.0	4244994.0	10085.0	19619.0	2021-10-20 02:30:29	1.0
9327	Venom Let there be Carnage	2021.0	2.0	0.0	7097896.0	10092.0	19627.0	2021-10-21 10:11:03	1.0
9328	Respect	2021.0	3.0	0.0	2452150.0	10106.0	19654.0	2021-10-24 21:14:06	1.0

This ready-made dataset is under the Attribution-Noncommercial 3.0 Unported (CC BY-NC 3.0) license that allows sharing and adaptation. The data collection and distribution measures have been fair since the beginning so there seem to be no ethical issues involved. But, most of the movies on record are from Hollywood, indicating a selection bias. FiveThirtyEight also points out a potential “[feminist-leaning](#)” bias that manifests as people more inclined to submit movies that pass the Bechdel Test. On the other hand, the comments available on the website are property of the commenters, so we will analyze the comments and keep in mind to give credit to these individuals.

Additionally, we obtained data on the gender information on movie cast and crew members from [Kaggle](#). Using the information, we calculated the ratio of female cast and crew members in a dataframe and joined it with the dataframe of the Bechdel Test scores. This Kaggle dataset is collected from TMDb and Movielens and is under CC0: Public Domain license, so we can use it for our analysis. The only inconvenience with this dataset is that it was uploaded 4 years ago, so we fail to possess a lot of important information regarding movies released in recent years.

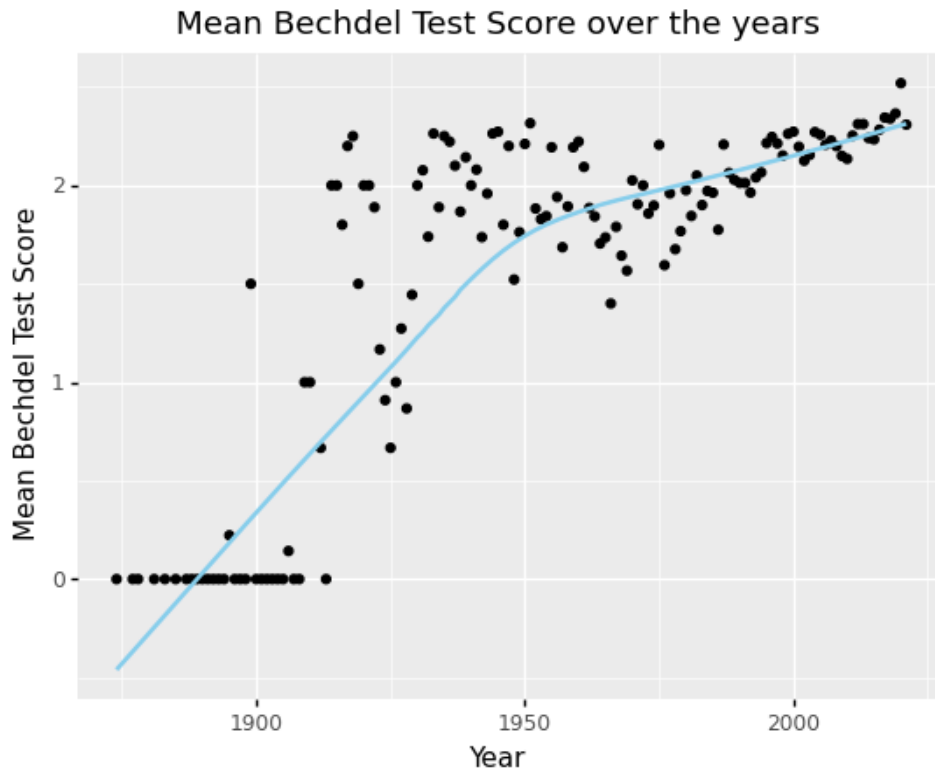
Analysis

First, we conducted exploratory data analysis on the Bechdel Test score dataset with 9329 movies. There are a couple of NA's in the column dubious, so we drop them and have 9030 movies in total. The bar graph and pie graph of Bechdel scores are:



Surprisingly, more than half of the movies in the dataset have passed the Bechdel Test. This might be related to the selection bias in the previous section.

We also plotted how the mean of Bechdel Test score changes over the years:

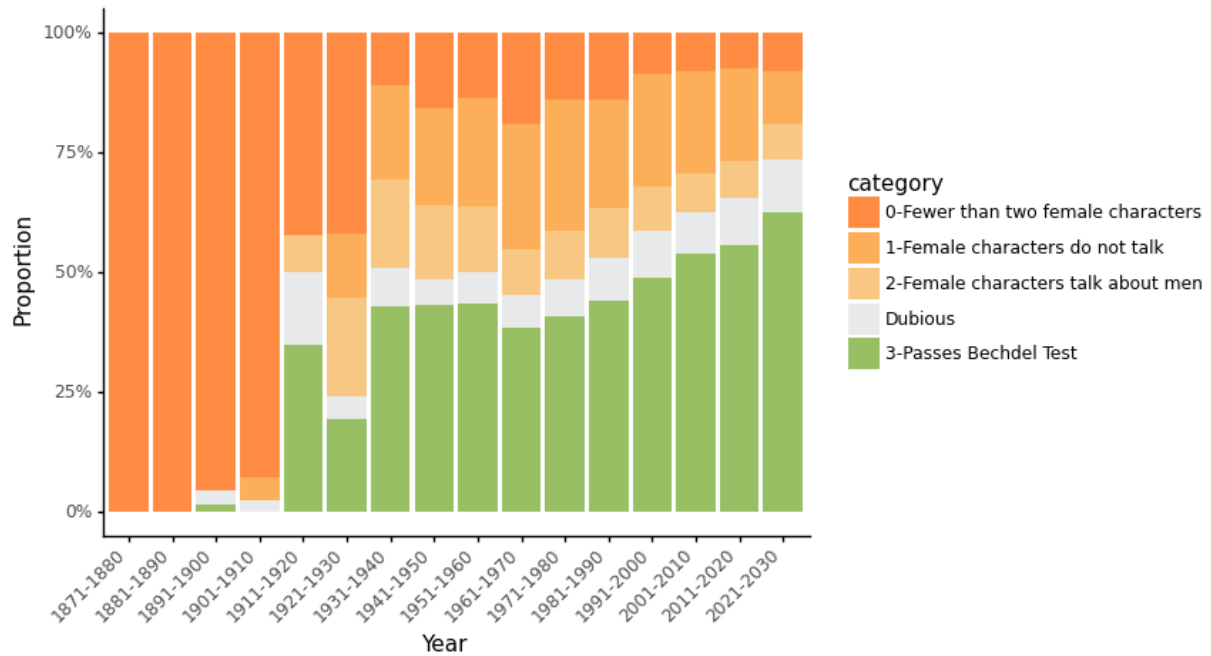


The overall trend is delightful. The mean is getting higher over time and shows a promising tendency to keep rising. However, we also need to keep in mind that 8.91% of these scores are dubious, which means these scores have not been validated and are subject to changes. Among the dubious movies, 2.73% scored 0, 9.81% scored 1, 13.54% scored 2 and 73.91% passed the test. Therefore, we temporarily treat dubious as another category and classify the 9030 movies into:

- 0 - Fewer than two female characters
- 1 - Female characters do not talk
- 2 - Female characters talk about men
- 3 - Passes Bechdel Test
- Dubious

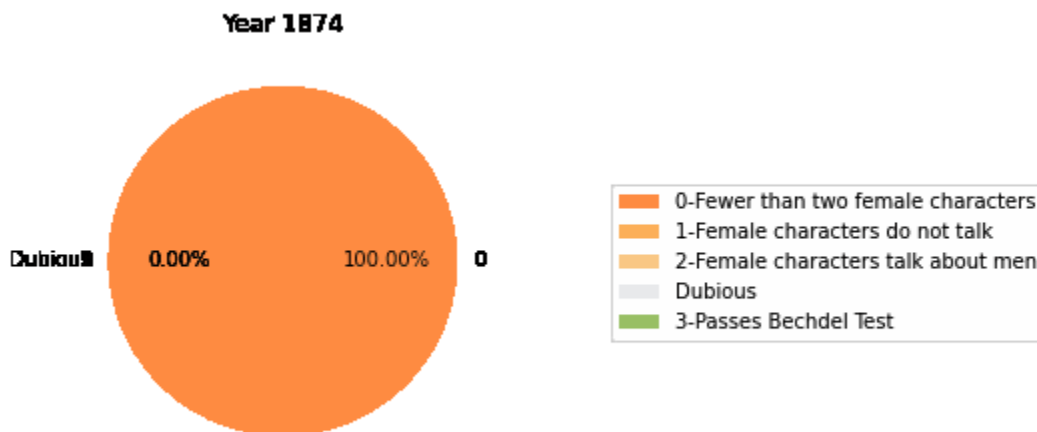
In order to further understand how the Bechdel Test score changes, we plotted a stacked percentage bar chart with an interval of 10 years:

Percentage Changes of Movies Passing Bechdel Test over the Years



We color-coded movies passing the test in green and failing the test in different shades of orange. We also put Dubious in the middle so that it is easier for the eyes to compare what is on the side. Despite the dubious scores, the general trend remains the same. There are indeed more movies passing the Bechdel Test over time.

Unfortunately, the stacked percentage bar chart cannot fit too many bars, so we made an animated pie chart to showcase the yearly changes, which shows a similar trend:



(Please see the GIF in this [GitHub repo](#) because PDF does not support animation.)

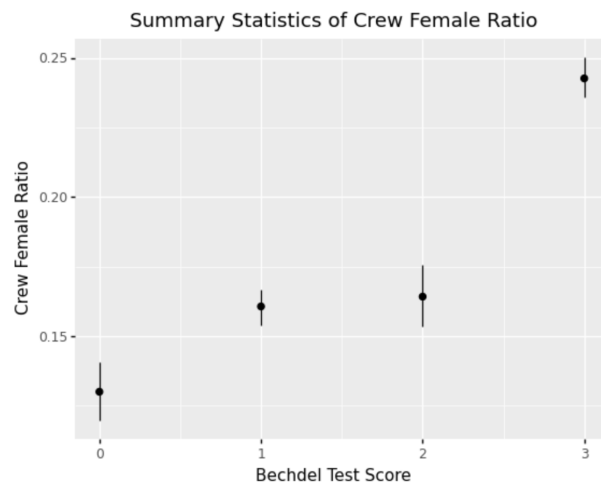
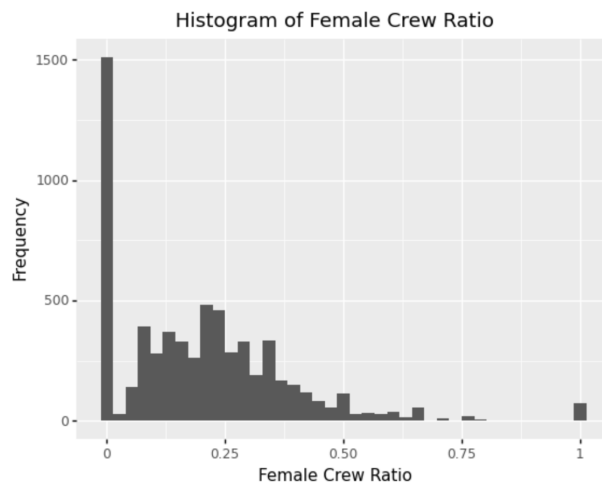
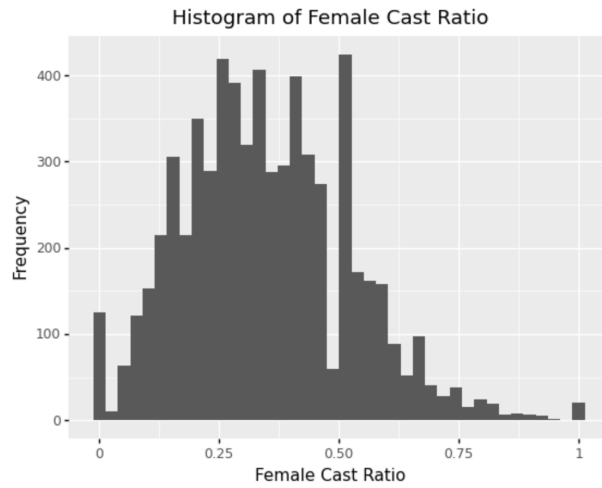
Despite the higher number of films passing every year, we see that the growth rate from 1971-2021 is relatively low, which is indicated through the small changes in proportion each year. With the growth rate of films passing the Bechdel Test still being low, we can argue that the Bechdel Test is still a test with high standards for the film industry, which indicates the test's relevance even today.

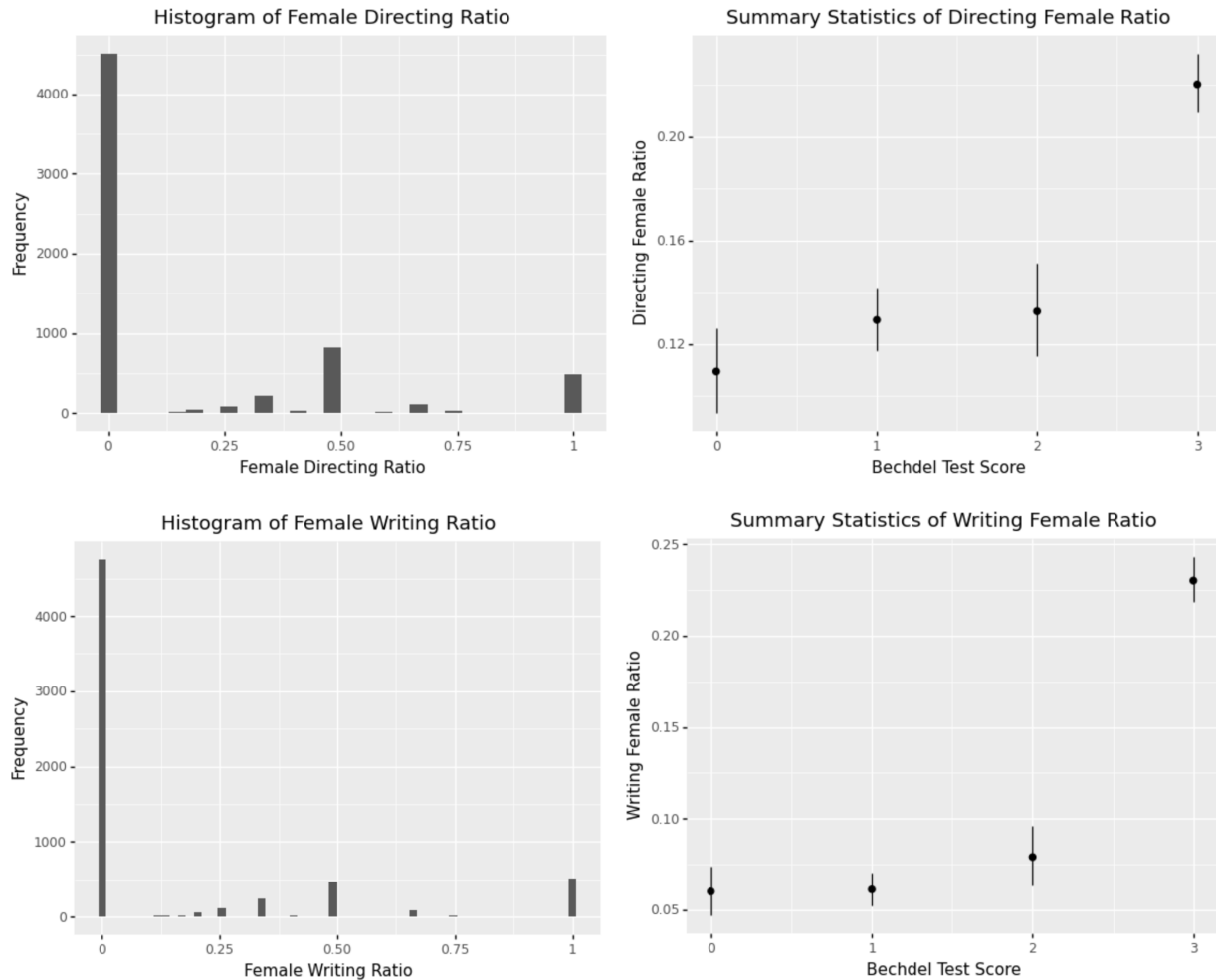
However, with more than half of the films in 2021 passing the Bechdel Test, **one can also argue that we still need an updated measurement tool to further assess movies that pass, as well as reflect the changing standards of what is considered “feminist”**. The definition of feminism has rapidly evolved since the Bechdel Test was first developed, which the Bechdel Test fails to consider and its standards are limited to how [feminism](#) mainly revolved around women having interests beyond supporting men and was privileged towards caucasian women. Today, feminism extends to being inclusive for all minority identities and revolves mainly around women having equal rights as men in leadership positions and the workplace. Therefore, while the Bechdel Test is still relevant and useful in seeing if a film is a feminist film, it requires revision in capturing the evolving definition and goals of feminism.

After having a good understanding of our main dataset and the positive changes over the years, we inner-joined the Bechdel Test dataset with the gender dataset obtained from Kaggle. After removing empty values and dubious movies, we are left with 6,376 movies. Since a lot of cast and crew members have their gender listed as unknown, we employed a gender prediction package called [gender-guesser](#) to classify gender by first name. The prediction often yields ambiguous results when encountering Asian names or names that are used by both male and female. However, there are only less than 5% of unknown gender after the prediction, which is acceptable. We then count the number of males, females and unknown gender for cast and crew, calculate the female ratio and get the following dataset:

index	title	year	bt_score	dubious	cast	crew	cast_fe...	crew_fe...	directi...	writing...
▽	▽	▽	▽	▽	▽	▽	▽	▽	▽	▽
0	House of...	1896	0	0	[{'cast_id': 5, 'c...	[{'credit_id': '52fe4ba5c...	0.5	0	0	0
1	Une nuit...	1896	0	0	[{'cast_id': 2, 'c...	[{'credit_id': '52fe4a35c...	0	0	0	0
2	Cinderel...	1899	3	0	[{'cast_id': 3, 'c...	[{'credit_id': '52fe4b45c...	0.333333...	0	0	0
3	Bluebeard	1901	0	0	[{'cast_id': 2, 'c...	[{'credit_id': '52fe478dc...	0.666666...	0	0	0
4	Le voyag...	1902	0	0	[{'cast_id': 11, '...	[{'credit_id': '52fe4275c...	0.181818...	0.083333...	0	0

We are going to focus on 4 new variables: `cast_female_ratio`, `crew_female_ratio`, `directing_female_ratio` and `writing_female_ratio`. Here are their histograms and summary statistics:





We can see from the histograms on the left that although most movies have a lot of female cast members, many movies do not have female crew or directors or writers. For the summary statistics on the right, we can see that groups failing the Bechdel test are not always different from each other, but they all have a lower female ratio compared to movies passing the test.

There is a correlation between passing and higher female off-screen representation across cast, crew, director and writing, but the ratios even for passing movies are surprisingly low—below around 25% for all but cast. With feminism today advocating for balanced and equal representation of females in the workplace, having a balanced representation of female and male members in creating the film seems necessary to advocate this feminist message in the film industry. After all, those engaged with the film production play the central role in creating and dispersing the plot and how female characters and their role on screen are portrayed. If male directors/writers own the portrayal of women in film, can we truly claim a film to be feminist even if it passes the Bechdel test? This might suggest that the off-screen benchmarks are more rigorous, and that the Bechdel test itself is limited to on-screen representation such as plot. Therefore, when assessing if a film is truly feminist today, we should go beyond the Bechdel test

to identify if other human-centered factors, such as female representation among cast, crew, director, and writing, advocate for feminism.

All code in this section can be found in this [GitHub repo](#).

3.2 Qualitative Analysis

Data

We take an ethnographic look at what viewers deem to be feminist. The Bechdel Test review site offers a rich source of empirical data in the form of viewer commentary on specific movies. We use two lists as a starting point in identifying contrarian movies that

1. [Surprisingly pass the Bechdel Test](#)
2. [Surprisingly fail the Bechdel Test](#)

From a sample of 4 “contradictory” movies, we have identified two relevant themes in on-screen representation that the Bechdel Test does not directly measure:

1. Professional Representation
2. Intersection with Racial Minority Representation

Analysis

Professional Representation

On the [Bechdel Test](#) rating site, Jurassic World, a film that passes the test, received much criticism among raters for its representation of women. In the film, the female operations manager of a dinosaur park, Claire, loses control of it when a particularly dangerous species escapes its enclosure. As an operations manager, Claire is seen multiple times in the movie talking about the dinosaur park with another female, and as a result, the film automatically passes the test. Yet, there is much criticism on the film’s representation of Claire, who, while a successful professional, was portrayed as incompetent when the situation deteriorated, and who had to depend on a male protagonist to save her. On the rating site, user Sophie writes:

“Yet she can't see what's going on in front of her nose—that the business is corrupt and executives completely reckless, all in the name of profit. The message that I got from it was that this was not really important because she had perfect hair, skin, and make-up and was hot and was probably earning a really good salary and that is success for a woman, right? Then her relationship with the dinosaur trainer was also appallingly sexist. When the shit hits the fan and the reckless behaviour of the company is exposed (i.e., monster dinosaur escapes its cage), she needs a man to save her.”

Another user Bernie describes:

“One cringe worthy scene was when they were hiding from the dinosaurs. As the manager of the park, you’d expect she’d know the park inside out. Instead she cowers in a corner, whimpering, while he struts his manly stuff, picking up guns and jumping over skulls and other stuff.”

Users seem to take issue with the way Claire was represented *as a professional*. The fact that a strong female character had a challenging occupation in itself was insufficient, as representation is relational. How is the woman represented in relation to the man? Through these criticisms of Jurassic park, we see that the Bechdel Test fails to account for factors that multiple human users were easily able to spot. Despite passing the Bechdel Test easily, Jurassic World’s portrayal of Claire strongly overshadows the Bechdel Test’s conclusion that the film is a feminist film.

Conversely, another movie, Gravity, was hailed as a feminist masterpiece despite failing the Bechdel Test. A moving story about a scientist’s fight for survival after a tragic space accident, it fails the test as it focuses very heavily on the single heroine, Dr Stone, a scientist on her first space mission, *saving herself*. Unlike Claire in Jurassic World, Dr Stone was introduced as a character that had to overcome incredible obstacles due to inexperience and extenuating circumstances, but who had found the will and strength to survive. Notably, she did this on her own without a male character rescuing her:

“There’s a part in the film when I thought, “Oh, here it comes. The ubiquitous scene where a dude comes and rescues her. As if she can’t rescue herself.” Thankfully, I was wrong. Some quibble that it’s a hallucination of Kowalski, so he’s the one who saves her. Nope, it’s all her. Sure he inspired her. But it’s her memory, it’s her imagination.”

Taking inspiration from Gravity and Jurassic World, we propose a Test 1 to assess the representation of women in professional roles:

For movies with characters that have any technical or professional occupation:

1. Women are not represented among the professionals in the film—0
2. There is at least one woman with a technical or professional career—1
3. The woman solves, or is integral to solving, a technically complex problem—2
4. The woman solves a technically complex problem independent or despite of a man—3

This is especially useful and interesting to apply to films in the sci-fi genre, where characters of technical ability and expertise feature heavily.

Applying Test 1 to Avengers (2012)

The Avengers (2012) movie failed the Bechdel Test, but featured prominent female characters—the Black Widow (Scarlett Johansson), Pepper Potts (Gwyneth Paltrow), and Agent Maria Hill (Cobie Smulders). User S wrote on the Bechdel Test Rating site:

“While I agree with the rating, I find it absolutely fascinating that the strength of the women in this movie is argued because it did not pass. Black Widow, Agent Hill, and even Pepper Potts are some of the strongest female characters I have seen portrayed and it isn't because they pass or fail a gender bias test. It is because they have story and history. They have emotions, but they use them to work forward.”

The Black Widow was the only female member of the Avengers, but her strength and expertise as a spy for S.H.I.E.L.D were showcased in her first scene where she was tasked to take down a weapons trafficker, General Luchkov. The scene showcases her combat ability and skillful diversion as a spy as she subdues three men alone while initially being chained to a chair.

From the first scene alone, the film would pass the final criteria of our test, which requires that a “woman solves a technically complex problem independent of or despite of a man”.

However, we should note that all of the female characters in Avengers - Black Widow, Agent Hill, and Pepper Potts - are caucasian. While Avengers pass our Test 1 and display women in highly professional careers and positions, “with the rise of fourth wave feminism, the [concepts of privilege and intersectionality](#) have gained widespread traction amongst younger feminists. To truly bring about change that is meaningful for all, everyone's voice needs to be at the table.” The previous quote illustrates that the current definition of feminism accentuates the importance of including all females, including those of all races. Therefore, while Avengers pass Test 1, the lack of diversity of females in Avengers suggests that it does not fit into the definition of today's feminism.

Intersection with Racial Minority Representation

While an intersectional lens is particularly useful when understanding female representation in movies, it is broad and, by definition, considers diverse social identities. For the scope of our report, we look specifically at the representation of Asian women in movies as a starting point.

Another notable film that failed the Bechdel test but that has been praised for its female representation is *Downsizing* (2017). Vietnamese-American Actress Hong Chau plays Ngoc Lan Tran, whom she [claimed](#) “was complex and exciting and unlike any character I’ve seen in a movie”. For her role, Hong Chau was nominated for the Golden Globe for Best Supporting Actress, the [third such nomination](#) for an actress of East Asian descent since 1970.

The movie follows a therapist Paul Safranek (Matt Damon) that chooses to be shrunk to a miniature size to inhabit tiny communities in order to minimize their consumption of resources. Hong Chau plays the supporting role of Lan Tran, a political dissident and refugee with a tragic backstory but who is relentlessly altruistic and who helps residents of the slum she inhabits. The character is the moral compass of the film and is an undoubtedly empowered woman. Yet, her refugee and minority identity and her disability make her a layered and complex character that does not fit neatly into one-directional stereotypes of Asian Americans.

While this might seem like a low bar to set, Asian American representation has been stubbornly dogged by stereotypes, with nearly half of all Asian roles in movies from 2010 to 2019 serving as a punchline, according to a [study](#) by the Geena Davis Institute on Gender in Media. Fetishization is just a dangerous extension of racism. Asian women are often viewed only as hypersexualized bodies and they’re the punchline of sex jokes. Over the years, Hollywood movies continued to enforce these ideas:

“South Asian women are portrayed as exotic seductresses. East Asian women are often represented as “lotus blossoms,” innocent girls who are still sexually available, or “dragon ladies,” women who are evil and use their femininity to deceive men.”

For instance, a movie that passes the Bechdel Test, *Memoirs of a Geisha* (2005), has been [criticized](#) for hypersexualizing Asian women as “feminine, shy, fragile, subservient, and sexually submissive”. The story is told through the eyes of Chiyo, who becomes a Geisha, but does not feature her agency nor everyday life and instead paints her as an exotic object of Colonel Derriks.

Leveraging on India Roby’s [op-ed](#) on the representation of Asian women in film, produced in conversation with Asian American Scholars, we may create a simple Test 2 to check for stereotyping of Asian women in films.

A film passes the test if it has a named female Asian character that:

1. Is not sexualized as the “Lotus Blossom” (innocent and sexually subservient)
2. Is not sexualized as “Dragon Ladies” (evil or deceitful, weaponizing their sexuality)
3. Is not dependent on a man for her livelihood.

Of course, a film might be feminist but fail the above test because it does not have any Asian American representation. This test is useful as a tool to assess movies with known female Asian characters, and tests for *how*, not if, they are represented.

Applying Test 2 to Crazy Rich Asians (2018)

While Crazy Rich Asians passes the Bechdel Test, it would fail Test 2 because of the character Kitty Pong (Fiona Xie), a gold-digging soap opera actress that dates wealthy heir Alistair Cheng. After she is told about a wealthier heir, Bernard Tai, she moves on to seduce him instead.

While our test does prompt a closer look at how Asian women are represented, it speaks to the limits of using a set of conditional statements to assess layered, nuanced films. Crazy Rich Asians has been [praised](#) for its representation of female Asian women, and does not at all negatively portray all of them as hypersexualized. A human critic would likely see the character of Kitty Pong as intended for comedic effect, as opposed to reinforcing negative stereotypes.

4. Summary of Findings, Implications and Limitations

Is the Bechdel Test still useful? Yes, but we should not stop there.

Quantitative evaluation reveals both the limitations and strengths of the Bechdel Test. There is still a large segment of films that do not pass or are dubious, which reflects the enduring relevance of the test since it was created, as well as the limited progress the film industry has made in on-screen representation. Since the Bechdel Test's creation in 1985, the pass rate has increased by an average of 0.41% per year. Assuming this growth rate is maintained, we would only reach a 100% pass rate in 2110.

However, more than half of the films in our dataset now pass the Bechdel Test. This calls for other benchmarks that further disentangle representation among the tests that *have* passed, including off-screen benchmarks, which still reflect severe underrepresentation of women in non-cast roles, as well as other tests for on-screen representation.

While films that pass all 3 levels of the Bechdel Test have much higher ratios of off-screen female representation compared to films that do not, women still represent less than 25% of directors, writers and crew among the films that *do* pass. This is indicative of the demanding standards of simple off-screen representation metrics for today's film industry.

Further, we have found many empirical reviews of films that “contradict” the Bechdel Test (films with good representation that surprisingly fail, or films with bad representation that surprisingly pass), we have suggested two tests that critically assess two recurrent forms of female stereotyping in film: professional under or misrepresentation of women and the sexualization of Asian women.

More Tests are needed

The tests we suggest are far less universal than the Bechdel Test, and are specific to the stated types of stereotyping while offering a more critical assessment. This is where human-centeredness is key—we cannot identify all the problems with female representation with just a single, universal Bechdel Test. Issue-specific tests informed by empirical film analysis are key to flagging harmful stereotypes in films. In fact, many issue-specific tests have been [suggested](#) by critics and film professionals to assess diverse, intersectional issues of female representation.

Limitations of Our Tests

Clearly, our tests, alongside others suggested by critics and film professionals, do not offer universal standards to aspire towards. Relating to our tests, female characters can be strong and empowered in areas outside of their professional life, and films do not need to have female Asian characters to be considered feminists. These issue-specific tests have to be applied in appropriate contexts, and not just used as a mask over all films.

A key problem with this is scalability, as a greater number of issue-specific tests with highly nuanced criteria means greater difficulty in application. The straightforward Bechdel Test has yet to be applied to every movie by a reviewer; increasing the number of tests will only grow the scope of assessment further.

Furthermore, the higher standards in tests might also be disputed, as there is no single authority on how women should be represented on film. Nuanced conditions can also be more difficult to apply, with viewers often disagreeing on even the simple Bechdel test. The language we use in our tests (e.g. “Technically complex problem”) can mean different things to different viewers.

Beyond on-screen representation, many data points in off-screen representation are difficult to obtain. Consider the off-screen representation of queer or transgender women—with an average crew size of [more than 600](#), it could be very difficult to ascertain the sexual orientations and identities of over 60,000 crew members for the 100 top-grossing movies, even if queer female representation in filmmaking might be an important sub-component of broader female representation.

5. Reflection and Future Work

In this project, we focus on human-centeredness in identifying biases in datasets, transparency and code reproducibility, qualitative research design and improving the measurement of female representation in movies from a human-centered perspective.

The problems we’ve identified with our tests and the Bechdel test make clear the need for empirical study in film, even if it can provide us with a useful abstractive measure that gives us a big picture of the entire industry.

As a baseline metric, the Bechdel Test does a good job of showing us how far Hollywood is from on-screen gender equality. However, as shown in our quantitative analysis, women are also underrepresented off-screen, and off-screen metrics remain a rigorous point-of-reference for gender equality in the film industry.

Furthermore, our qualitative analysis helped us realize the shortcomings of the Bechdel Test and reaffirm the need for more nuanced tests as well as the broader empirical study of film in order to identify and flag underrepresentation or misrepresentation of women in film. It is our hope that there will be an organized application of alternative tests at the scale that has been achieved for the Bechdel Test.

As time progresses, the definition of feminism is evolving and being redefined by humans. Therefore, by simply applying the Bechdel Test without human interaction and humans overseeing the entirety of the movie beyond just a simple scene, we run into the likelihood of lowering the standards of feminism of what it was in the past.