Online Appendix for

Bad Apples on Rotten Tomatoes: Critics, Crowds, and Gender Bias in Product Ratings

Luis Aguiar

Department of Business Administration
University of Zurich, and CESifo*

*E-mail: luis.aguiar@business.uzh.ch.

A1 Institutional Details and Rescaling of Ratings

This section provides more details on the institutional setting of Rotten Tomatoes and IMDb, in particular to understand how each platform incorporates crowd ratings. Both Rotten Tomatoes provide a very simple way for users to leave ratings on their platform. On IMDb, a user is required to create an IMDb account in order to rate a movie, and they can report their gender within their personal account at any point. Note that a user's gender does not appear publicly on the platform, and that leaving a rating on IMDb does not require writing a review.[1] Posting a rating on Rotten Tomatoes works in a similar fashion, with some small differences. After creating an account, a user can easily leave a rating, with or without a review text. In either case, the rating will count towards the average audience score displayed on the movie's main page (under "Audience Score"). If the rating is posted without a review of at least 20 characters, however, the rating will not be displayed among the reviews listed under the "All Audience" page.[2] This means that the list of crowd ratings that I obtain from Rotten Tomatoes is a subset (i.e. the subset of ratings that were left with a review of at least 20 characters) of all the ratings left by users on Rotten Tomatoes. The list of ratings that appear in the IMDb crowd sample, however, consists of all the ratings left by IMDb users (without any review text).

Rating scores on IMDb and Rotten Tomatoes – coming from both critics and crowds – differ in their scales. Crowd ratings on IMDb are based on a discrete rating scale ranging from 1 to 10, while Rotten Tomatoes' crowd ratings range from 0.5 to 5. Given that Rotten Tomatoes' scale is simply half that of IMDb, I simply multiply Rotten Tomatoes ratings by 2 to make these two groups of ratings comparable. The ratings left by critics on Rotten Tomatoes differ more significantly, however. In particular, different critics rely on different rating scales that are sometimes non-numerical. To allow for a comparison of rating scores across critics and crowd reviewers on Rotten Tomatoes and IMDb, I proceed as follows to rescale all critics ratings to a common 1-10 scale. First, I use the conversion scale from Metacritic to convert

[1] IMDb separately reports user reviews as opposed to ratings. However the gender of the reviewer is not reported in this case, and some reviews only contain review text but no score. I therefore solely focus on the user ratings on IMDb.

[2] For more detailed information, see https://www.rottentomatoes.com/faq.

non-numerical rating scales into a numerical 0-100 scale.[3] For each critic in the sample, I then convert their numerical ratings to a rating based on a 1-10 scale. More specifically, for each critic whose original scores S^O rely on a scale ranging from values S^O_{min} to S^O_{max}, I apply the following transformation to convert them into a scale ranging from 1 to 10:

$$S^{Scaled} = \frac{9 \times (S^O - S^O_{min})}{S^O_{max} - S^O_{min}} + 1$$

For instance, for a critic whose scores range from $S^O_{min} = 1$ to $S^O_{max} = 5$, an original score of $S^O = 3$ would be converted to a score equal to $S^{Scaled} = \frac{9\times(3-1)}{5-1} + 1 = 5.5$.

To test whether such rescaling has any impact on the main results of the analysis, I re-estimate the specifications presented in the main text for the subset of critics who rely on similar ratings scales as the crowd ratings. For instance, I restrict the sample to critics whose original rating scale ranges from 1 to 10, which implies that no rating transformation is needed to compare their ratings with those of the IMDb crowd. Performing such exercise results in quantitatively similar results despite relying on a very specific set of critics. Once the ratings from both critics and crowds on both platforms have been re-scaled to a common 1-10 range, I further scale them to a 10-100 range to facilitate the reading of the results in the main analysis.

A2 Detailed Descriptives

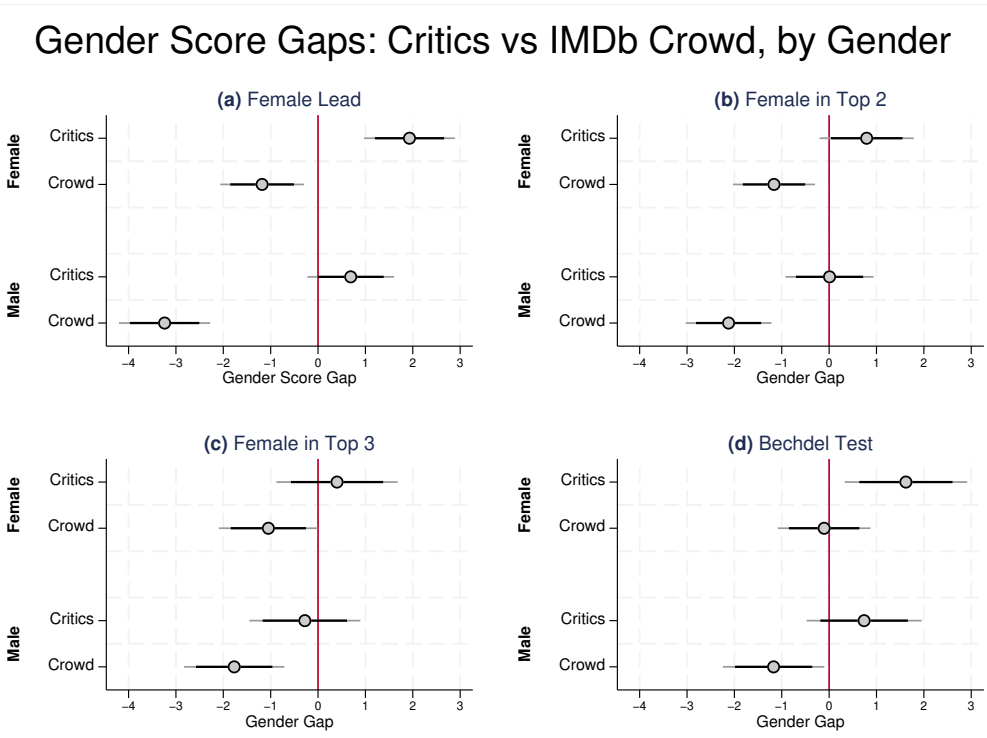
As reported in Section 2.2 of the main text, a striking 79% of a movie's ratings comes from men among the IMDb crowd, on average. While this result is consistent with existing findings relying on similar data (Boyle, 2014; Stroube and Waguespack, 2024; Bayerl et al., 2024), it is interesting to see whether there exist differences in gender share of reviews across male and female movies. When focusing on movies with a male lead actor, the female share of reviews drops further to 18%, while the corresponding share for movies with a female lead

[3] See https://web.archive.org/web/20200626125515/https://www.metacritic.com/about-metascores.

actress is equal to 25.6%. These descriptives suggest that some gender-specific preferences might be at play, but identifying the mechanism driving these inequality in review posting is challenging. As reported by Bayerl et al. (2024), one potential explanation is that men and women might simply differ in their propensity to post online reviews. Another potential mechanism might relate to unequal time constraints between men and women. If women face more restrictions than men on their leisure time, their supply of movie consumption (and, by extension, of movie ratings) could be reduced. Charmes (2019) indeed reports that women participation in the labor market is constrained by a higher burden of participation in unpaid care work, which would naturally affect their participation in leisure-related activities. Given that an important determinant of time dedicated to unpaid care work is motherhood and care of children, one should perhaps observe differences across female reviewers of different age groups if leisure time constraints are driving the lower participation of women in movie-rating activity. The data indeed reveals that the female share of movies ratings declines further for reviewers of higher age groups. Among crowd reviewers less than 29 years old, the female share of reviews reaches 28%, while it declines to 20% and 17% among crowd reviewers aged between 30 and 44, and crowds over 45, respectively. Theses differences also appear when focusing on male and female movies separately. For female movies, the female share of crowd reviews reaches 35% among reviewers less than 29 and drops to 26% and 19% for reviewers aged 30-44 and over 45, respectively. For male movies, the corresponding female shares are 24%, 17%, and 15%. While these figures are consistent with stronger time-constraints put on women, they do not allow to rule out other motives for the gender disparities in movie reviewing.

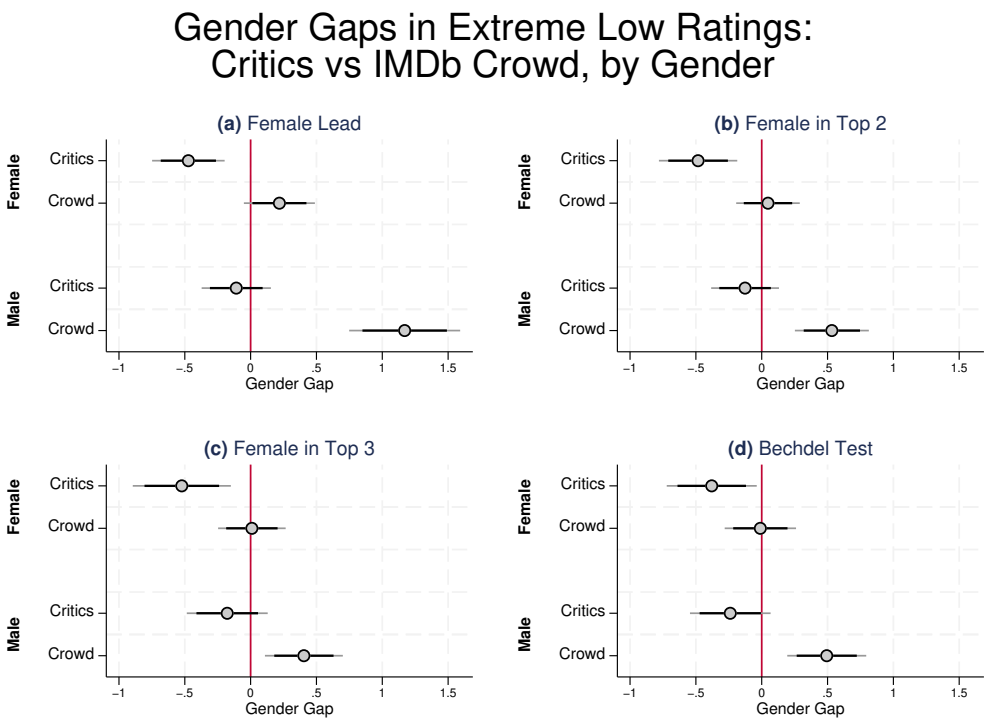
A3 Estimation Results - Alternative Female Movie Definitions

The following reports estimation results based on IMDb and Rotten Tomatoes data relying on various female movie definitions as described in Section 3 in the main text.



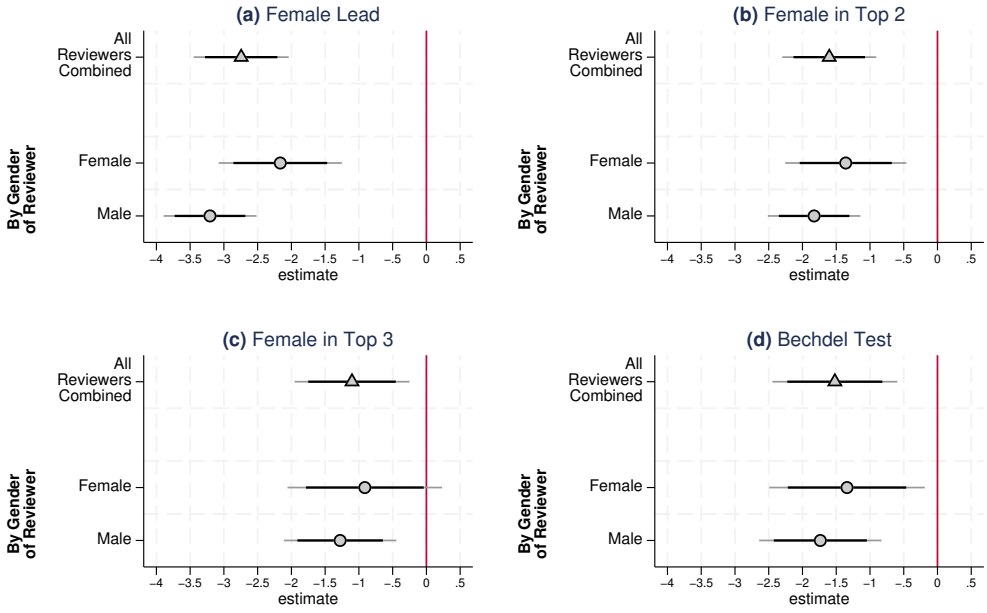
Note: For a given type of reviewer, each coefficient represents the movie gender score gap, which is defined as the difference in score given to female and male movies. Panel **(a)** considers a movie as female if the top lead actress is female. Panels **(b)** and **(c)** consider a movie as female if an actress is present in the Top 2 or Top 3 of the cast, respectively. Panel **(d)** considers a movie as female if it passes the Bechdel test. For each panel, all estimates come from a single regression as described in the text. Standard errors are clustered at the movie level. Horizontal lines depict 99% and 95% confidence intervals.

Figure A1: Gender Score Gap, by Reviewer Type and Gender.



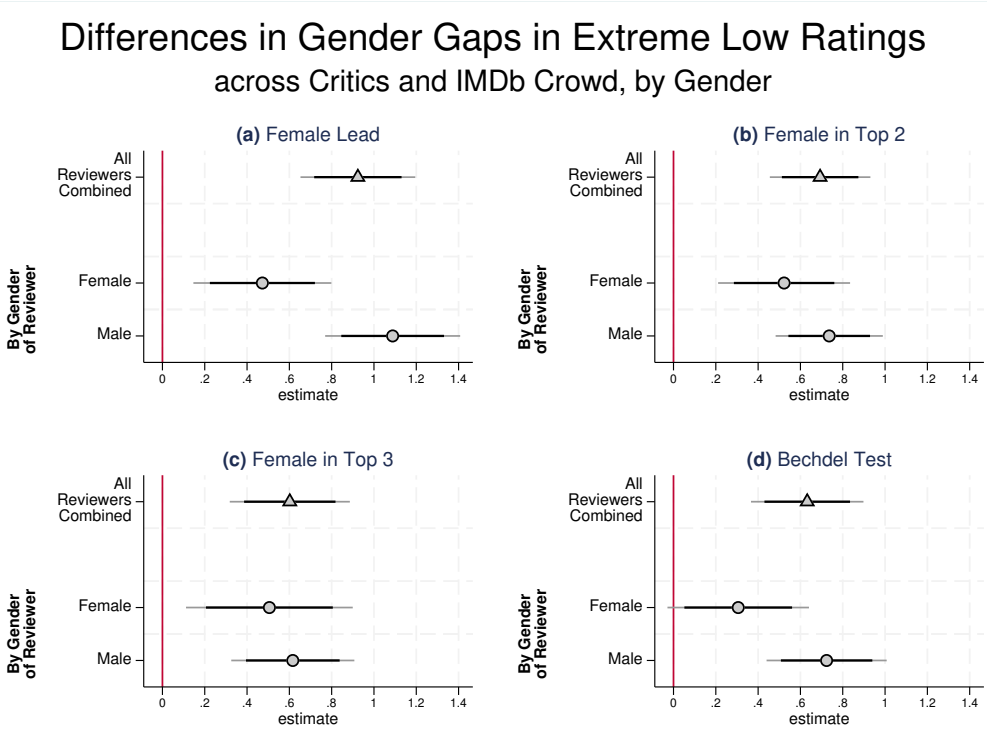
Note: For a given type of reviewer, each coefficient represents the movie gender gap in extreme low ratings, which is defined as the difference in the share of ratings scores of 10 given to female and male movies. Panel **(a)** considers a movie as female if the top lead actress is female. Panels **(b)** and **(c)** consider a movie as female if an actress is present in the Top 2 or Top 3 of the cast, respectively. Panel **(d)** considers a movie as female if it passes the Bechdel test. For each panel, all estimates come from a single regression as described in the text. Standard errors are clustered at the movie level. Horizontal lines depict 99% and 95% confidence intervals.

Figure A2: Gender Score Gap, by Reviewer Type and Gender.



Note: Each coefficient corresponds to a difference-in-differences estimate, which is defined as the difference in the gender score gaps across critics and audience reviewers. Panel **(a)** considers a movie as female if the top lead actress is female. Panels **(b)** and **(c)** consider a movie as female if an actress is present in the Top 2 or Top 3 of the cast, respectively. Panel **(d)** considers a movie as female if it passes the Bechdel test. For each type of reviewer, the gender score gap is itself defined as the difference in score given to female and male movies. For comparison, the difference-in-differences estimate for all reviewers combined is displayed in each panel. Standard errors are clustered at the movie level. Horizontal lines depict 99% and 95% confidence intervals.

Figure A3: Difference-in-Differences Estimates, by Reviewers' Gender.



Note: Each coefficient corresponds to a difference-in-differences estimate, which is defined as the difference in the gender gap in extreme low ratings across critics and audience reviewers. Panel **(a)** considers a movie as female if the top lead actress is female. Panels **(b)** and **(c)** consider a movie as female if an actress is present in the Top 2 or Top 3 of the cast, respectively. Panel **(d)** considers a movie as female if it passes the Bechdel test. For each type of reviewer, the gender gap in extreme low ratings is itself defined as the difference in the share of rating scores of 10 given to female and male movies. For comparison, the difference-in-differences estimate for all reviewers combined is displayed in each panel. Standard errors are clustered at the movie level. Horizontal lines depict 99% and 95% confidence intervals.

Figure A4: Difference-in-Differences Estimates, by Reviewers' Gender.

Table A1: Gender Score Gaps between Critics and IMDb Crowds based on Various Female Movie Definitions. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie	0.986***		0.159		-0.154		0.919*	
	(0.352)		(0.360)		(0.455)		(0.470)	
Female Movie × Crowd Reviewer	-3.537***	-2.744***	-2.004***	-1.603***	-1.412***	-1.102***	-1.794***	-1.520***
	(0.367)	(0.273)	(0.353)	(0.270)	(0.427)	(0.331)	(0.436)	(0.359)
Crowd Reviewer	4.223***	2.362***	4.418***	2.499***	4.286***	2.353***	4.190***	2.456***
	(0.194)	(0.158)	(0.273)	(0.216)	(0.378)	(0.297)	(0.328)	(0.283)
Worldwide Box Office Revenue	0.005***		0.005***		0.005***		0.005***	
	(0.001)		(0.001)		(0.001)		(0.001)	
Movie Released in Theaters	1.688***		1.711***		1.753***		0.844	
	(0.565)		(0.570)		(0.573)		(0.999)	
Number of Release Countries	0.069***		0.070***		0.071***		0.038***	
	(0.010)		(0.010)		(0.010)		(0.012)	
Number of Awards Won	0.073***		0.072***		0.072***		0.071***	
	(0.008)		(0.008)		(0.008)		(0.008)	
Genre Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Production Studio Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
MPAA Rating Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Year of Release Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Origin Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Language Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Movie Fixed Effects	✗	✓	✗	✓	✗	✓	✗	✓
Gender Score Gap from Crowd	-2.550		-1.845		-1.566		-0.875	
p-value	0.000		0.000		0.000		0.028	
R^2	0.091	0.203	0.091	0.203	0.090	0.203	0.082	0.182
Number of Observations	337380397	337380397	337380397	337380397	337380397	337380397	283498730	283498730

† All specifications use the rating score (measured from 10 to 100) as the dependent variable. Box office revenue is measured in million USD. Crowd ratings come from IMDb. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A2: Gender Gaps in Extreme Low Ratings between Critics and IMDb Crowds based on Various Female Movie Definitions. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie	-0.213**		-0.207**		-0.254**		-0.276**	
	(0.097)		(0.097)		(0.117)		(0.115)	
Female Movie × Crowd Reviewer	1.124***	0.924***	0.641***	0.693***	0.584***	0.602***	0.664***	0.632***
	(0.137)	(0.105)	(0.106)	(0.092)	(0.119)	(0.110)	(0.110)	(0.103)
Crowd Reviewer	1.020***	1.546***	0.949***	1.399***	0.884***	1.360***	0.533***	0.887***
	(0.059)	(0.052)	(0.080)	(0.070)	(0.105)	(0.097)	(0.074)	(0.072)
Worldwide Box Office Revenue	0.000		0.000		0.000		0.000	
	(0.000)		(0.000)		(0.000)		(0.000)	
Movie Released in Theaters	-0.889***		-0.901***		-0.910***		-0.425	
	(0.316)		(0.317)		(0.317)		(0.431)	
Number of Release Countries	-0.031***		-0.032***		-0.032***		-0.018***	
	(0.003)		(0.003)		(0.003)		(0.003)	
Number of Awards Won	-0.004***		-0.004**		-0.004**		-0.004**	
	(0.002)		(0.002)		(0.002)		(0.002)	
Genre Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Production Studio Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
MPAA Rating Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Year of Release Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Origin Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Language Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
Movie Fixed Effects	✗	✓	✗	✓	✗	✓	✗	✓
Gender Gap from Crowd	0.912		0.433		0.331		0.388	
p-value	0.000		0.000		0.002		0.000	
R^2	0.011	0.071	0.011	0.071	0.011	0.071	0.007	0.040
Number of Observations	337380397	337380397	337380397	337380397	337380397	337380397	283498730	283498730

† All specifications use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. Box office revenue is measured in million USD. Crowd ratings come from IMDb. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A3: Differences in Gender Gaps between Critics and Rotten Tomatoes Crowds based on Various Female Movie Definitions. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
Dependent Variable:	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd Reviewer	-1.785***	0.906***	-1.248***	0.802***	-0.821***	0.745***	-0.873***	0.393***
	(0.297)	(0.125)	(0.269)	(0.107)	(0.314)	(0.120)	(0.337)	(0.124)
Elapsed Time since First Review	0.711***	0.235**	0.707***	0.237**	0.709***	0.237**	0.831***	0.148
	(0.249)	(0.117)	(0.249)	(0.117)	(0.249)	(0.117)	(0.282)	(0.128)
Review Position within Sequence	-0.052***	0.010***	-0.052***	0.010***	-0.052***	0.010***	-0.052***	0.011***
	(0.010)	(0.002)	(0.010)	(0.002)	(0.010)	(0.002)	(0.010)	(0.002)
Review Year	-0.575***	0.039	-0.573***	0.038	-0.573***	0.038	-0.617***	0.053
	(0.084)	(0.037)	(0.084)	(0.037)	(0.084)	(0.037)	(0.094)	(0.041)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.457	0.262	0.457	0.262	0.457	0.262	0.449	0.265
Number of Observations	15825513	15825513	15825513	15825513	15825513	15825513	13418334	13418334

† For each Female movie definition, the first specification uses the rating score (measured from 10 to 100) as the dependent variable, while the second uses a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A4: Difference in Gender Score Gaps across Critics, Verified, and Non-Verified Rotten Tomatoes Crowds, based on Various Female Movie Definitions. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd	-2.984**		-2.680		-2.299		0.806	
	(1.437)		(1.650)		(2.084)		(2.293)	
Female Movie × Non Verified Crowd		-4.282***		-4.209**		-2.472		0.178
		(1.456)		(1.697)		(2.037)		(2.196)
Female Movie × Verified Crowd		-0.436		0.371		-1.824		1.532
		(1.828)		(1.996)		(2.351)		(2.734)
Non-Verified Crowd		3.036***		4.200***		3.563*		0.592
		(1.066)		(1.532)		(1.878)		(1.905)
Verified Crowd		7.782***		7.526***		9.090***		5.566**
		(1.151)		(1.639)		(2.117)		(2.426)
Elapsed Time since First Review	-10.102***	-5.760***	-10.165***	-6.122***	-10.160***	-5.806***	-10.150***	-5.607***
	(2.128)	(1.801)	(2.127)	(1.806)	(2.126)	(1.790)	(2.469)	(2.063)
Review Position within Sequence	-0.128***	-0.149***	-0.129***	-0.149***	-0.128***	-0.142***	-0.125***	-0.139***
	(0.047)	(0.036)	(0.047)	(0.038)	(0.046)	(0.038)	(0.045)	(0.037)
Review Year	-0.746	-0.521	-0.724	-0.473	-0.736	-0.539	-0.700	-0.560
	(0.480)	(0.433)	(0.481)	(0.434)	(0.482)	(0.448)	(0.565)	(0.520)
Average Rating Assigned to Other Movies	0.392***	0.355***	0.392***	0.355***	0.392***	0.355***	0.383***	0.347***
	(0.022)	(0.019)	(0.022)	(0.019)	(0.022)	(0.019)	(0.025)	(0.022)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✗	✗	✗	✗	✗	✗	✗	✗
R^2	0.264	0.274	0.264	0.274	0.264	0.273	0.252	0.261
Number of Observations	459505	459505	459505	459505	459505	459505	377284	377284

† All specifications use the rating score (measured from 10 to 100) as the dependent variable. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A5: Difference in Gender Gaps in Extreme Low Ratings across Critics, Verified, and Non-Verified Rotten Tomatoes Crowds, based on Various Female Movie Definitions. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd	1.572***		1.555***		1.526*		0.718	
	(0.574)		(0.568)		(0.858)		(0.716)	
Female Movie × Non Verified Crowd		2.590***		2.499***		2.192**		1.576*
		(0.788)		(0.742)		(1.035)		(0.885)
Female Movie × Verified Crowd		-0.388		-0.274		0.276		-0.721
		(0.689)		(0.590)		(0.761)		(0.684)
Non-Verified Crowd		6.018***		5.342***		5.098***		5.673***
		(0.448)		(0.592)		(0.958)		(0.635)
Verified Crowd		4.218***		4.178***		3.894***		4.530***
		(0.340)		(0.431)		(0.708)		(0.577)
Elapsed Time since First Review	2.202*	0.193	2.235**	0.400	2.231**	0.281	2.508*	0.386
	(1.122)	(1.030)	(1.119)	(1.034)	(1.116)	(1.018)	(1.298)	(1.175)
Review Position within Sequence	0.008	0.020***	0.008*	0.019***	0.008	0.015***	0.005	0.013**
	(0.005)	(0.006)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Review Year	-0.246	-0.362*	-0.258	-0.383*	-0.251	-0.354*	-0.160	-0.230
	(0.216)	(0.195)	(0.215)	(0.196)	(0.216)	(0.201)	(0.246)	(0.224)
Average Rating Assigned to Other Movies	-0.176***	-0.159***	-0.176***	-0.159***	-0.177***	-0.159***	-0.175***	-0.158***
	(0.015)	(0.013)	(0.015)	(0.013)	(0.015)	(0.013)	(0.018)	(0.016)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✗	✗	✗	✗	✗	✗	✗	✗
R^2	0.086	0.089	0.086	0.089	0.086	0.089	0.079	0.083
Number of Observations	459505	459505	459505	459505	459505	459505	377284	377284

† All specifications use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A6: Difference in Gender Score Gaps across Critics, Verified, and Non-Verified Rotten Tomatoes Crowd. VADER Sentiment Analysis. †

Dependent Variable:	Rating Score				Extreme Low Rating			
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd	-0.029*	-0.028*			0.361**	0.422***		
	(0.015)	(0.015)			(0.176)	(0.156)		
Female Movie × Non Verified Crowd			-0.045***	-0.046***			0.620***	0.716***
			(0.017)	(0.017)			(0.233)	(0.211)
Female Movie × Verified Crowd			0.002	0.007			-0.124	-0.139
			(0.019)	(0.020)			(0.161)	(0.176)
Non-Verified Crowd				0.181***				1.353***
				(0.010)				(0.091)
Verified Crowd				0.193***				0.713***
				(0.011)				(0.071)
Elapsed Time since First Review	-0.102***	-0.110***	-0.103***	-0.088***	1.235***	1.785***	1.253***	1.118***
	(0.023)	(0.023)	(0.023)	(0.021)	(0.423)	(0.372)	(0.425)	(0.353)
Review Position within Sequence	-0.001***	-0.002***	-0.001***	-0.002***	0.003	0.006***	0.004	0.010***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.002)	(0.004)	(0.004)
Review Year	-0.002	-0.001	-0.001	0.000	-0.190*	0.030	-0.199*	-0.007
	(0.006)	(0.006)	(0.006)	(0.006)	(0.104)	(0.085)	(0.105)	(0.083)
Average Rating Assigned to Other Movies		0.004***		0.004***		-0.027***		-0.021***
		(0.000)		(0.000)		(0.003)		(0.002)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
R^2	0.404	0.127	0.404	0.128	0.274	0.015	0.274	0.017
Number of Observations	396562	444329	396562	444329	396562	444329	396562	444329

† A movie is considered female if the top actress is female. All specifications rely on the VADER classifier to construct reviews' sentiment score. The score ranges from -1 to 1, with a higher score reflecting a more positive sentiment. Columns (1)-(4) use the review sentiment score as the dependent variable. Columns (5)-(8) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 1^{st} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd reviews come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A7: Difference in Gender Score Gaps across Critics, Verified, and Non-Verified Rotten Tomatoes Crowd. TextBlob Sentiment Analysis. †

Dependent Variable:	Rating Score				Extreme Low Rating			
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd	-0.018**	-0.020**			0.170	0.169		
	(0.008)	(0.009)			(0.123)	(0.113)		
Female Movie × Non Verified Crowd			-0.025***	-0.029***			0.222*	0.229*
			(0.009)	(0.009)			(0.133)	(0.124)
Female Movie × Verified Crowd			-0.006	-0.002			0.071	0.057
			(0.011)	(0.013)			(0.126)	(0.122)
Non-Verified Crowd				0.063***				0.552***
				(0.007)				(0.078)
Verified Crowd				0.117***				0.561***
				(0.008)				(0.079)
Elapsed Time since First Review	-0.011	-0.094***	-0.012	-0.048***	0.607*	0.352	0.610*	0.317
	(0.011)	(0.015)	(0.011)	(0.012)	(0.354)	(0.284)	(0.353)	(0.274)
Review Position within Sequence	-0.000	-0.000	-0.000	-0.001***	0.005*	0.006***	0.005**	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.002)	(0.001)
Review Year	-0.005**	-0.006	-0.005*	-0.004	-0.020	-0.149**	-0.022	-0.152**
	(0.003)	(0.004)	(0.003)	(0.003)	(0.078)	(0.063)	(0.078)	(0.063)
Average Rating Assigned to Other Movies		0.003***		0.002***		-0.018***		-0.018***
		(0.000)		(0.000)		(0.002)		(0.002)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
R^2	0.436	0.134	0.436	0.142	0.309	0.009	0.309	0.009
Number of Observations	396562	444329	396562	444329	396562	444329	396562	444329

† A movie is considered female if the top actress is female. All specifications rely on the TextBlob classifier to construct reviews' sentiment score. The score ranges from -1 to 1, with a higher score reflecting a more positive sentiment. Columns (1)-(4) use the review sentiment score as the dependent variable. Columns (5)-(8) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 1^{st} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd reviews come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A8: Difference in Gender Score Gaps across Critics, Verified, and Non-Verified Rotten Tomatoes Crowd. LIWC Sentiment Analysis. †

Dependent Variable:	Rating Score				Extreme Low Rating			
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd	0.054	0.079			0.260**	0.266***		
	(0.130)	(0.126)			(0.111)	(0.098)		
Female Movie × Non Verified Crowd			0.152	0.179			0.306***	0.322***
			(0.137)	(0.135)			(0.118)	(0.105)
Female Movie × Verified Crowd			-0.129	-0.110			0.174	0.168
			(0.141)	(0.144)			(0.125)	(0.130)
Non-Verified Crowd				0.309***				0.707***
				(0.091)				(0.063)
Verified Crowd				0.138*				0.995***
				(0.083)				(0.078)
Elapsed Time since First Review	0.235	0.195	0.241	0.002	0.167	-0.616***	0.170	-0.445*
	(0.178)	(0.169)	(0.178)	(0.152)	(0.286)	(0.235)	(0.286)	(0.234)
Review Position within Sequence	0.004	0.005**	0.005	0.006**	0.004	0.005*	0.004	0.005*
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Review Year	0.091**	0.084*	0.088*	0.073	0.110	0.033	0.108	0.039
	(0.046)	(0.048)	(0.045)	(0.047)	(0.073)	(0.068)	(0.073)	(0.069)
Average Rating Assigned to Other Movies		-0.023***		-0.022***		-0.016***		-0.017***
		(0.002)		(0.002)		(0.002)		(0.002)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✗	✓	✗	✓	✗	✓	✗
R^2	0.361	0.065	0.361	0.066	0.321	0.008	0.321	0.008
Number of Observations	396501	444267	396501	444267	396501	444267	396501	444267

† A movie is considered female if the top actress is female. All specifications rely on the LIWC's negative tone dictionary to construct reviews' sentiment score, ranging from 0 to 100. A higher score reflects a more negative review sentiment. Columns (1)-(4) use the review sentiment score as the dependent variable. Columns (5)-(8) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 99^{th} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd reviews come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

A4 Unobservable Differences in Movie Content

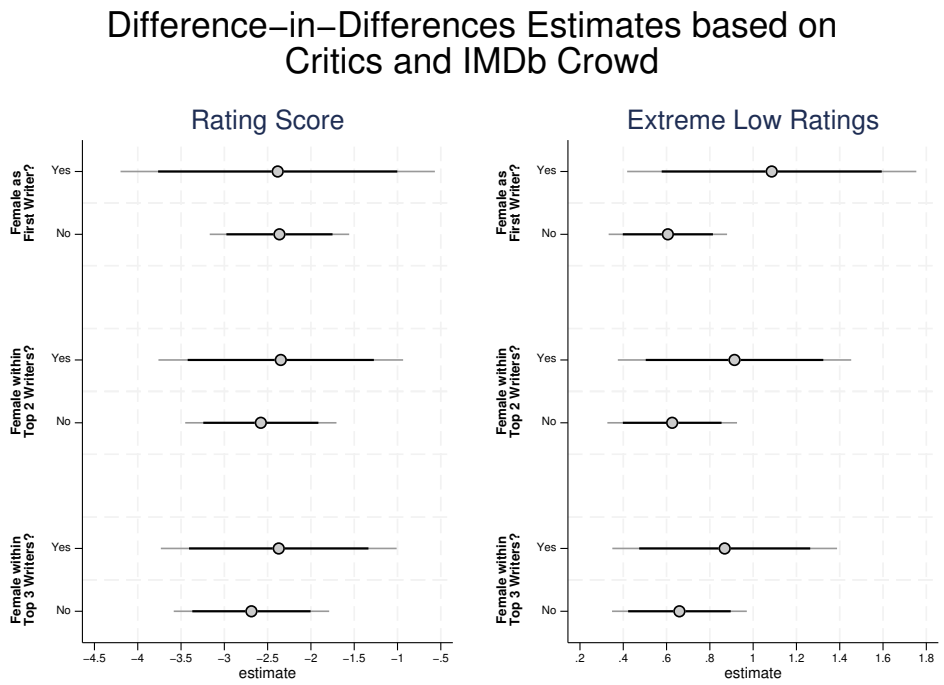
This section reports the results of performing similar estimations as in Sections 4.1 and 4.2 of the main text, but within different subsets of movies. More specifically, I consider the subset of female-written movies and male-written movies as well as the subset of male-produced and female-produced movies separately.

Relying on the IMDb data, I start by estimating equation (1) within the subset of female-written movies and male-written movies separately. In other words, I ask whether the increase in gender gaps between critics and crowds is found among both male-written movies and female-written movies. I perform a similar exercise comparing male produced and female produced movies. IMDb provides the list of each movie's producers as well as screen writers. Screen writers are listed in an order that reflects level of contribution.[4] I can therefore distinguish between movies whose first writer is female, movies that have a female writer among their top 2 writers, and movies that have a female writer among their top 3 writers. For movie producers, IMDb lists them in alphabetical order and does not provide information on their individual level of contribution. I therefore distinguish between movies that have at least 3, at least 2, or at least 1 female producers among their producer team. Figure A5a reports the difference-in-differences coefficients distinguishing between male and female written movies. For instance, the top estimates reported in both graphs rely on the subset of movies with a female top writer (i.e. "female-written movies"), while the second estimates focus on the subset of movies with a male top writer (i.e. "male-written movies"). The remainder of estimates focus on samples constructed with alternative definitions of a female-written movies (based on whether a female is among the top 2 or the top 3 writers). The figure indicates that the difference-in-differences estimates are similar across all subsamples, providing support to the fact that the increase in the gender gaps documented above are driven by differences in female presence among the top cast rather than differences in content. As an alternative test, Figure A5b reports similar estimates focusing on subset of movies based on the gender of their producers. Again, the results indicate that the increase in the

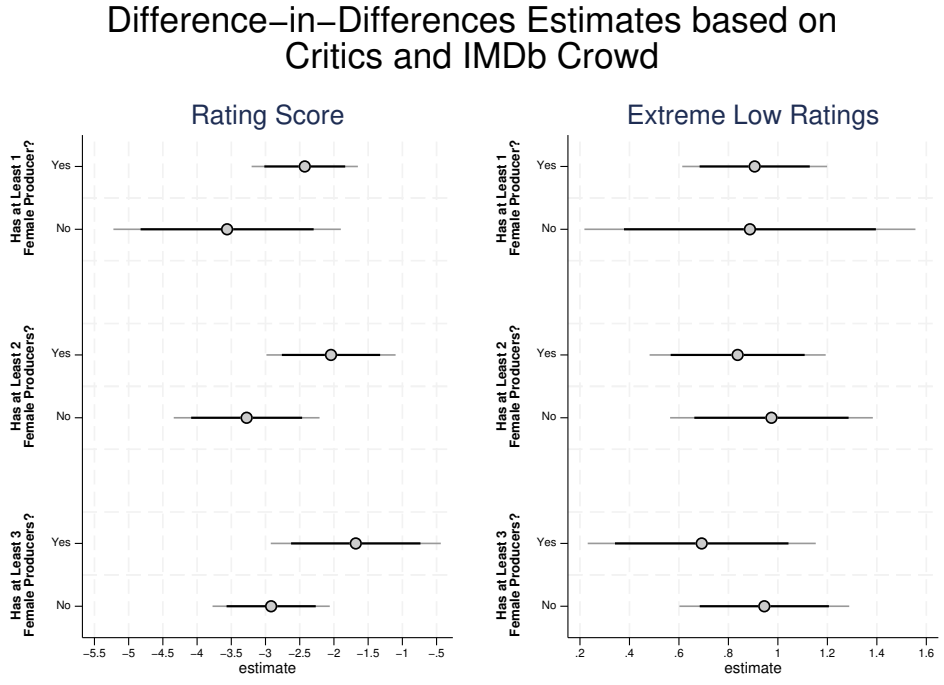
[4]See https://help.imdb.com/article/contribution/filmography-credits/writers/GPLAT3NTCGA67A6R?ref_=helpms_helpart_inline#.

gender gaps are present both within the sets of male and female produced movies. Tables A9 and A10 report the complete results of all regressions.

I perform a similar exercise relying on the Rotten Tomatoes data. Figure A6a reports the corresponding difference-in-differences estimates when considering female and male-written movie separately. Figure A6b performs the same exercise considering male and female-produced movies. In all cases, the resulting estimates are very similar across all subsamples. These results again provide support to the fact that the increase in gender gaps are driven by differences in movies' on-screen female presence rather than differences in movie content. Tables A11 and A12 report the complete estimation results of all the corresponding regressions.

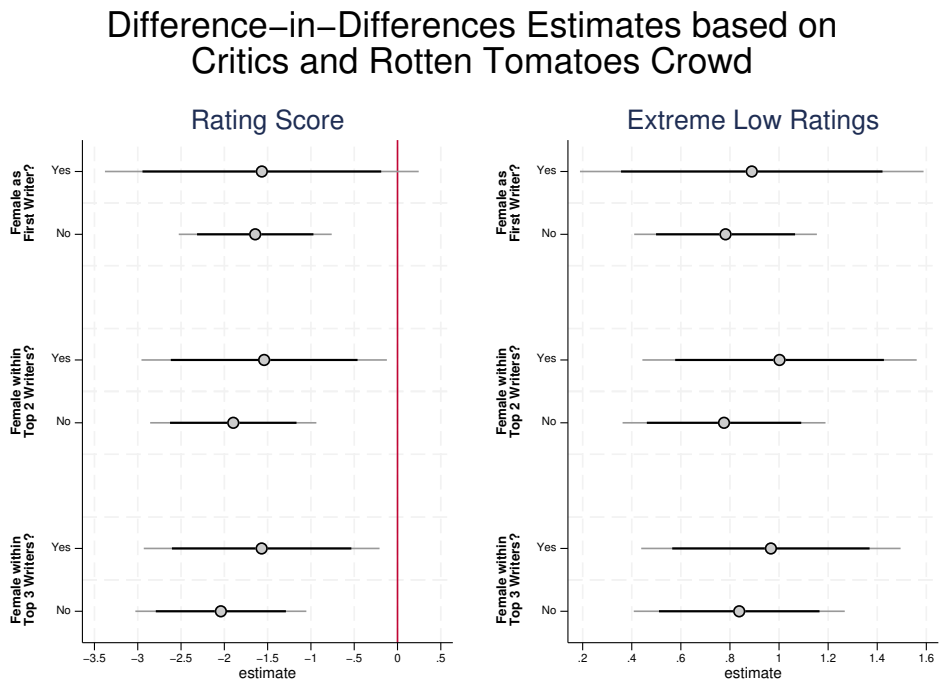


(a) By Gender of Writer.

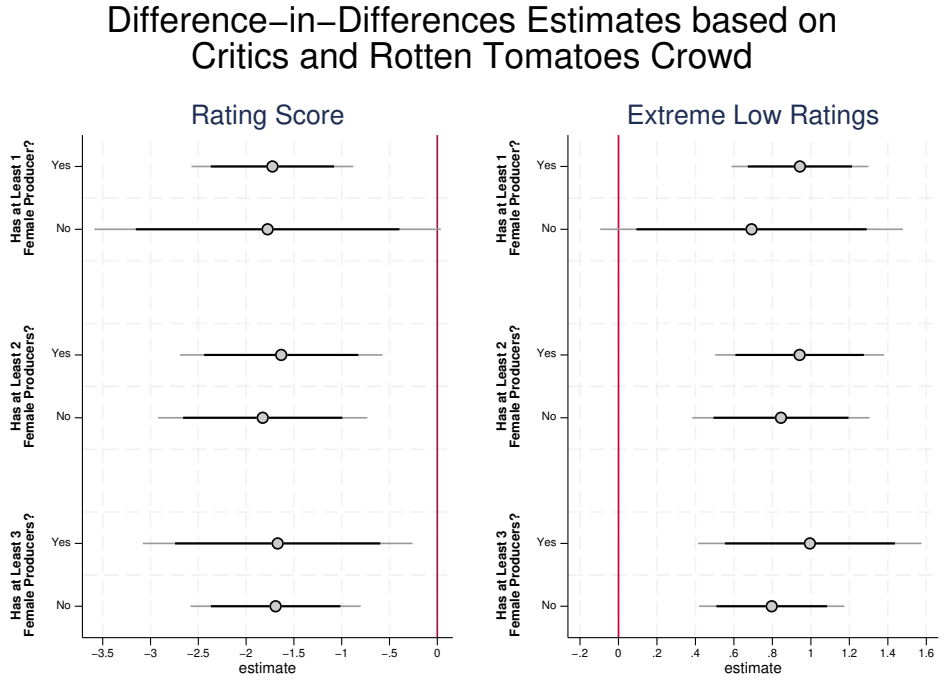


(b) By Gender of Producer.

Figure A5: Difference-in-Differences Estimates based on Critics and IMDb Crowds, by Gender of and Writer and Producer.



(a) By Gender of Writer.



(b) By Gender of Producer.

Figure A6: Difference-in-Differences Estimates based on Critics and Rotten Tomatoes Crowds, by Gender of Writer and Producer.

Table A9: Gender Gaps between Critics and IMDb Crowds, by Gender of Writer. †

Dependent Variable:	Rating Score			Extreme Low Rating		
Female-Written Movie Based on:	Female as Top Writer	Female among Top 2 Writers	Female among Top 3 Writers	Female as Top Writer	Female among Top 2 Writers	Female among Top 3 Writers
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Crowd	2.495***	2.506***	2.480***	1.490***	1.458***	1.454***
	(0.165)	(0.172)	(0.174)	(0.055)	(0.057)	(0.058)
Female Movie × Crowd	-2.363***	-2.578***	-2.687***	0.606***	0.626***	0.659***
	(0.313)	(0.339)	(0.348)	(0.106)	(0.116)	(0.121)
Female Movie × Crowd × Female-Written	-0.020	0.230	0.315	0.480*	0.288	0.209
	(0.771)	(0.645)	(0.633)	(0.280)	(0.239)	(0.235)
Crowd × Female-Written	-1.413**	-0.906**	-0.712	0.634***	0.548***	0.510***
	(0.592)	(0.452)	(0.440)	(0.165)	(0.142)	(0.140)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓
Diff-in-Diff for Female-Written Movies	-2.384	-2.348	-2.372	1.086	0.914	0.869
p-value	0.001	0.000	0.000	0.000	0.000	0.000
R^2	0.203	0.203	0.204	0.071	0.071	0.071
Number of Observations	335626056	333678781	332543004	335626056	333678781	332543004

† A movie is considered female if the top actress is female. Specifications (1)-(3) use the rating score (measured from 10 to 100) as the dependent variable. Specifications (4)-(6) use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. Crowd ratings come from IMDb. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A10: Gender Gaps between Critics and IMDb Crowds, by Gender of Producer. †

Dependent Variable:	Rating Score			Extreme Low Rating		
Female-Produced Movie Based on:	# Female Producers ≥ 1	# Female Producers ≥ 2	# Female Producers ≥ 3	# Female Producers ≥ 1	# Female Producers ≥ 2	# Female Producers ≥ 3
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Crowd	3.924***	3.200***	2.864***	1.453***	1.461***	1.439***
	(0.340)	(0.219)	(0.182)	(0.131)	(0.076)	(0.058)
Female Movie × Crowd	-3.561***	-3.276***	-2.917***	0.887***	0.974***	0.945***
	(0.645)	(0.414)	(0.333)	(0.260)	(0.159)	(0.133)
Female Movie × Crowd × Female-Produced	1.135	1.233**	1.237**	0.019	-0.137	-0.253
	(0.712)	(0.553)	(0.586)	(0.283)	(0.211)	(0.223)
Crowd × Female-Produced	-1.990***	-1.708***	-2.014***	0.117	0.172*	0.430***
	(0.383)	(0.314)	(0.361)	(0.142)	(0.104)	(0.127)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓
Diff-in-Diff for Female-Produced Movies	-2.426	-2.043	-1.681	0.906	0.837	0.692
p-value	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.203	0.203	0.203	0.071	0.071	0.071
Number of Observations	337155930	337155930	337155930	337155930	337155930	337155930

† A movie is considered female if the top actress is female. Specifications (1)-(3) use the rating score (measured from 10 to 100) as the dependent variable. Specifications (4)-(6) use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. Crowd ratings come from IMDb. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A11: Gender Gaps between Critics and Rotten Tomatoes Crowds, by Gender of Writer. †

Dependent Variable:	Rating Score			Extreme Low Rating		
Female-Written Movie Based on:	Female as Top Writer	Female among Top 2 Writers	Female among Top 3 Writers	Female as Top Writer	Female among Top 2 Writers	Female among Top 3 Writers
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Female Movie × Crowd	-1.643***	-1.897***	-2.040***	0.782***	0.776***	0.838***
	(0.343)	(0.373)	(0.383)	(0.144)	(0.160)	(0.167)
Female Movie × Crowd × Female-Written	0.076	0.357	0.471	0.107	0.226	0.129
	(0.781)	(0.662)	(0.649)	(0.307)	(0.269)	(0.262)
Elapsed Time since First Review	0.712***	0.703***	0.709***	0.235**	0.238**	0.237**
	(0.250)	(0.251)	(0.250)	(0.117)	(0.118)	(0.117)
Review Position within Sequence	-0.052***	-0.052***	-0.052***	0.010***	0.010***	0.010***
	(0.010)	(0.010)	(0.010)	(0.002)	(0.002)	(0.002)
Review Year	-0.577***	-0.573***	-0.575***	0.040	0.039	0.039
	(0.084)	(0.085)	(0.084)	(0.037)	(0.037)	(0.037)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓
Diff-in-Diff for Female-Written Movies	-1.567	-1.540	-1.569	0.889	1.002	0.967
p-value	0.026	0.005	0.003	0.001	0.000	0.000
R^2	0.457	0.457	0.457	0.262	0.262	0.262
Number of Observations	15757665	15694821	15663870	15757665	15694821	15663870

† A movie is considered female if the top actress is female. Specifications (1)-(3) use the rating score (measured from 10 to 100) as the dependent variable. Specifications (4)-(6) use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. Crowd ratings come from IMDb. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A12: Gender Gaps between Critics and Rotten Tomatoes Crowds, by Gender of Producer. †

Dependent Variable:	Rating Score			Extreme Low Rating		
Female-Produced Movie Based on:	# Female Producers ≥ 1	# Female Producers ≥ 2	# Female Producers ≥ 3	# Female Producers ≥ 1	# Female Producers ≥ 2	# Female Producers ≥ 3
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.
Female Movie × Crowd	-1.775**	-1.826***	-1.691***	0.692**	0.845***	0.797***
	(0.703)	(0.425)	(0.346)	(0.305)	(0.179)	(0.147)
Female Movie × Crowd × Female-Produced	0.050	0.193	0.022	0.251	0.097	0.198
	(0.777)	(0.589)	(0.643)	(0.336)	(0.245)	(0.266)
Elapsed Time since First Review	0.705***	0.704***	0.704***	0.236**	0.236**	0.238**
	(0.249)	(0.249)	(0.249)	(0.117)	(0.117)	(0.117)
Review Position within Sequence	-0.052***	-0.052***	-0.052***	0.010***	0.010***	0.010***
	(0.010)	(0.010)	(0.010)	(0.002)	(0.002)	(0.002)
Review Year	-0.574***	-0.574***	-0.574***	0.039	0.039	0.039
	(0.084)	(0.084)	(0.084)	(0.037)	(0.037)	(0.037)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓
Diff-in-Diff for Female-Produced Movies	-1.725	-1.633	-1.670	0.943	0.942	0.995
p-value	0.000	0.000	0.002	0.000	0.000	0.000
R^2	0.457	0.457	0.457	0.262	0.262	0.262
Number of Observations	15822004	15822004	15822004	15822004	15822004	15822004

† A movie is considered female if the top actress is female. Specifications (1)-(3) use the rating score (measured from 10 to 100) as the dependent variable. Specifications (4)-(6) use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. Crowd ratings come from IMDb. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

A5 Sentiment Analysis

I rely on three rule-based analyzers to construct measures of each review's sentiment. I first rely on VADER (Valence Aware Dictionary for sEntiment Reasoning) and TextBlob's sentiment analyzer, two popular tools that provide measures of the sentiment of a given text, ranging from -1 (most negative sentiment) to 1 (more positive sentiment). I additionally use the LIWC (Linguistic Inquiry and Word Count) tool to construct additional measures of text sentiment (Pennebaker et al., 2001; Boyd et al., 2022).[5] For a given text, LIWC classifies words in various pre-defined categories (based on linguistic dimensions, psychological processes, etc.) including positive and negative emotions. I use the LIWC's negative tone dictionary to obtain a measure of the percentage of negative-tone words (e.g. "bad," "hate," "wrong" etc.) out of a given review. This measures accordingly ranges from 0 (no words with negative tone) to 100 (all words with negative tone).

Specifications (1)-(3) of Table A13 reports the results of estimating equation (3) using the three sentiment score measures (VADER, TextBlob, and LIWC) as the dependent variable. Each specification includes both movie and reviewer fixed effects and additionally controls for the sequential and temporal dynamics of ratings. The estimates show very similar results as the ones relying on rating scores. Based on the three sentiment score measures, the difference-in-differences estimates indicate that the gender sentiment score gap is significantly larger among crowd reviewers than among critics, and at the expense of female movies.

In order to test whether extreme behavior is also reflected within review's sentiment scores, I further construct measures of extreme low sentiment scores. For the VADER and TextBlob classifiers, I define sentiment scores as extremely low if they fall in the 1^{st} percentile of the respective review sentiment score distribution. For the LIWC classifier, I define sentiment scores as extremely low if they fall in the 99^{th} percentile of the review sentiment score distribution.[6] Specifications (4)-(6) present the results of using these measures as dependent

[5] While LIWC is a very popular text analysis tool, it is well-established that VADER performs better in the context of social media and online text. In particular, "VADER distinguishes itself from LIWC in that it is more sensitive to sentiment expressions in social media contexts while also generalizing more favorably to other domains" (Hutto and Gilbert, 2014).

[6] Recall that higher values of the LIWC sentiment score reflect more negative reviews.

variables. Consistent with the results presented in Table 4, the estimates show that crowds increase the gender gap in extreme low sentiment scores relative to critics. Overall, the estimates presented in Table A13 provide robustness to our main results by showing that crowd reviewers increase gender gaps in movie reviews' sentiment relative to critics.[7]

Table A13: Difference in Gender Sentiment Gaps across Critics and Rotten Tomatoes Crowds. †

	Sentiment Score			Extreme Sentiment Score		
Sentiment Classifier:	VADER	TextBlob	LIWC	VADER	TextBlob	LIWC
	(1)	(2)	(3)	(4)	(5)	(6)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Female Movie × Crowd Reviewer	-0.028***	-0.016***	0.136***	0.223***	0.126***	0.142***
	(0.004)	(0.002)	(0.030)	(0.056)	(0.035)	(0.035)
Elapsed Time since First Review	-0.009**	-0.003	0.107***	0.103**	0.177***	0.217***
	(0.004)	(0.002)	(0.028)	(0.043)	(0.041)	(0.043)
Review Position within Sequence	-0.001***	-0.000***	0.003***	0.001	0.004***	0.004***
	(0.000)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)
Review Year	-0.005***	-0.003***	0.008	0.014	-0.013	-0.011
	(0.001)	(0.001)	(0.009)	(0.014)	(0.014)	(0.015)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓
R^2	0.262	0.281	0.213	0.138	0.165	0.165
Number of Observations	14703460	14703460	14701429	14703460	14703460	14701429

† A movie is considered female if the top actress is female. Specifications (1)-(3) use the review sentiment score from the corresponding classifier as the dependent variable. Both the VADER and TextBlob score range from -1 to 1, and a larger score reflects a more positive sentiment. Specifications (3) and (6) rely on LIWC's negative tone dictionary to construct reviews' sentiment score, ranging from 0 to 100. A higher score reflects a more negative review sentiment. Specifications (4)-(6) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. For the VADER and TextBlob classifiers, extreme scores are defined as scores that fall in the 1^{st} percentile of the respective review sentiment score distribution. For the LIWC classifier, extreme scores are defined as scores that fall in the 99^{th} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd scores come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

[7] Tables A14, A15, and A16 further provide the results of estimating similar specifications using alternative definitions of female movies based on the VADER, TextBlob, and LIWC sentiment scores, respectively.

Table A14: Differences in Gender Gaps between Critics and Rotten Tomatoes Crowd based on Various Female Movie Definitions. VADER Sentiment Analysis. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
Dependent Variable:	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd Reviewer	-0.028***	0.223***	-0.023***	0.150***	-0.027***	0.098*	-0.016***	0.096*
	(0.004)	(0.056)	(0.004)	(0.048)	(0.004)	(0.059)	(0.004)	(0.057)
Elapsed Time since First Review	-0.009**	0.103**	-0.009**	0.103**	-0.009**	0.103**	-0.007	0.093**
	(0.004)	(0.043)	(0.004)	(0.043)	(0.004)	(0.043)	(0.005)	(0.046)
Review Position within Sequence	-0.001***	0.001	-0.001***	0.001	-0.001***	0.001	-0.001***	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Review Year	-0.005***	0.014	-0.005***	0.013	-0.005***	0.013	-0.006***	0.016
	(0.001)	(0.014)	(0.001)	(0.014)	(0.001)	(0.014)	(0.001)	(0.015)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.262	0.138	0.262	0.138	0.262	0.138	0.266	0.140
Number of Observations	14703460	14703460	14703460	14703460	14703460	14703460	12469125	12469125

† All specifications rely on the VADER classifier to construct reviews' sentiment score. The score ranges from -1 to 1, with a higher score reflecting a more positive sentiment. For each Female movie definition, the first specification uses the review sentiment score as the dependent variable, while the second uses a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 1^{st} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A15: Differences in Gender Gaps between Critics and Rotten Tomatoes Crowd based on Various Female Movie Definitions. TextBlob Sentiment Analysis. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
Dependent Variable:	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd Reviewer	-0.016***	0.126***	-0.015***	0.133***	-0.014***	0.104***	-0.008***	0.039
	(0.002)	(0.035)	(0.002)	(0.033)	(0.002)	(0.040)	(0.002)	(0.039)
Elapsed Time since First Review	-0.003	0.177***	-0.003	0.177***	-0.003	0.177***	-0.001	0.144***
	(0.002)	(0.041)	(0.002)	(0.041)	(0.002)	(0.041)	(0.002)	(0.045)
Review Position within Sequence	-0.000***	0.004***	-0.000***	0.004***	-0.000***	0.004***	-0.000***	0.004***
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Review Year	-0.003***	-0.013	-0.003***	-0.014	-0.003***	-0.014	-0.004***	-0.004
	(0.001)	(0.014)	(0.001)	(0.014)	(0.001)	(0.014)	(0.001)	(0.015)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.281	0.165	0.281	0.165	0.281	0.165	0.285	0.176
Number of Observations	14703460	14703460	14703460	14703460	14703460	14703460	12469125	12469125

† All specifications rely on the TextBlob classifier to construct reviews' sentiment score. The score ranges from -1 to 1, with a higher score reflecting a more positive sentiment. For each Female movie definition, the first specification uses the review sentiment score as the dependent variable, while the second uses a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 1^{st} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A16: Differences in Gender Gaps between Critics and Rotten Tomatoes Crowd based on Various Female Movie Definitions. LIWC Sentiment Analysis. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
Dependent Variable:	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score	Sentiment Score	Extreme Sentiment Score
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd Reviewer	0.136***	0.142***	0.142***	0.148***	0.164***	0.142***	0.093***	0.076**
	(0.030)	(0.035)	(0.029)	(0.033)	(0.034)	(0.039)	(0.033)	(0.038)
Elapsed Time since First Review	0.107***	0.217***	0.107***	0.218***	0.107***	0.218***	0.079**	0.197***
	(0.028)	(0.043)	(0.028)	(0.043)	(0.028)	(0.043)	(0.032)	(0.048)
Review Position within Sequence	0.003***	0.004***	0.003***	0.004***	0.003***	0.004***	0.003***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Review Year	0.008	-0.011	0.008	-0.012	0.008	-0.012	0.015	-0.004
	(0.009)	(0.015)	(0.009)	(0.015)	(0.009)	(0.015)	(0.010)	(0.016)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.213	0.165	0.213	0.165	0.213	0.165	0.219	0.175
Number of Observations	14701429	14701429	14701429	14701429	14701429	14701429	12467392	12467392

† All specifications rely on the LIWC's negative tone dictionary to construct reviews' sentiment score, ranging from 0 to 100. A higher score reflects a more negative review sentiment. For each Female movie definition, the first specification uses the review sentiment score as the dependent variable, while the second uses a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 99^{th} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

A6 Dynamics

To test whether reviewers who post a rating are reacting to the prevailing review environment, I follow Moe and Trusov (2011) and Godes and Silva (2012) and start by including two additional variables in equation (3): the mean and the standard deviation of all previous ratings of a given movie. The estimation results are presented in Table A17. In line with Godes and Silva (2012) – and consistent with herding behavior among reviewers – the estimates show that higher prevailing mean ratings are associated with higher subsequent ratings. On the other hand, more noisy prevailing ratings (i.e. a higher standard deviation of prevailing ratings) are associated with lower subsequent ratings. Note that the difference-in-differences estimates remain significant even controlling for these additional variables.[8]

To further account for the possibility of herding behavior among reviewers, I now explore whether the differences in the gender gaps across critics and crowds vary along a movie's life (Godes and Mayzlin, 2004). I do so by distinguishing between reviews posted in the early and late period of a movie's life and by relying on various cutoffs to define these two periods. More specifically, I define the following dummy variable:

$$Early_{it} = \begin{cases} 1 & \text{if } t \leq \tau \\ 0 & \text{if } t > \tau, \end{cases}$$

where τ corresponds to the cutoff (in days) determining a movie's early life relative to its release date. For instance, a value of $\tau = 7$ would consider a movie's early life as the first seven days after its release date. I then expand equation (3) by interacting all the coefficients with the dummy variables $Early_{it}$ and $Late_{it} = 1 - Early_{it}$ in order to obtain distinct coefficients for the early and late periods of a movie's life.[9]

[8] Note that given the lack of appropriate instruments for these lagged variables, there is a possibility that including them would bias the difference-in-differences coefficient of interest if these new variables are correlated with the error term. I therefore refrain from including these controls in the main model (3). The results presented in Table A17 nevertheless provide a good test for the robustness of the main results.

[9] In particular, I also interact the movie and reviewer fixed effects with the dummy variable for the early and late periods. Note that such empirical specification is equivalent to splitting the sample between the early and late periods. However, relying on the full sample has the advantage of allowing to test whether the coefficients of interest are statistically different across periods.

Table A18 presents the results of estimating various of these specifications across the range of τ values $\tau = \{7, 14, 30, 180\}$. Specifications (1) to (4) use the rating score as the dependent variable and all control for movie and reviewer fixed effects as well as sequential and temporal dynamics. Across all cutoff values, the results show that the difference in the gender score gap across crowds and critics realizes both in the early and late life of movies. As indicated in the bottom row, a t-test on the equality of both the early and late difference-in-differences estimates reveals no statistical difference between the two. Specifications (4)-(8) reveal a similar pattern when focusing on extreme low ratings. Overall, these results are again consistent with herding behavior among reviewers and suggest that reviews posted later in a movie's life contain limited additional information.[10]

Estimation Relying on Cross-Movie Variation Given the large number of crowd reviews posted for each movie and the results presented above, one alternative empirical approach is to construct, for each movie, a measure of the average rating score and conduct a cross-sectional analysis at the movie level rather than at the review level. Beyond providing a robustness check of the results presented above, such approach has several advantages. First, it does not need to address dynamic concerns. Second, it also reduces concerns about the clustering of standard errors given that all the data is collapsed at the movie level. Finally, such approach can also be implemented using the IMDb crowd data and therefore offers another way of comparing the results across the two platforms. I therefore estimate specifications of the following form:

$$\Delta_m = \quad \alpha + \delta FemaleMovie_m + \beta X_m + \varepsilon_m, \tag{1.a}$$

where $\Delta_m \equiv \overline{score}_m^{crowd} - \overline{score}_m^{critics}$ is the difference in the average score received by crowd

[10]Tables A19, A20, and A21 present the results of performing similar estimations relying on the VADER, TextBlob, and LIWC sentiment scores as dependent variables, respectively. All show similar results as the ones presented in Table A18.

Table A17: Gender Gaps between Critics and Rotten Tomatoes Crowds. †

Dependent Variable:	Rating Score		Extreme Low Rating	
	(1)	(2)	(3)	(4)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Female Movie × Crowd Reviewer	-1.785***	-1.401***	0.906***	0.637***
	(0.297)	(0.251)	(0.125)	(0.108)
Elapsed Time since First Review	0.711***	0.983***	0.235**	-0.007
	(0.249)	(0.209)	(0.117)	(0.092)
Review Position within Sequence	-0.052***	-0.024***	0.010***	0.006***
	(0.010)	(0.009)	(0.002)	(0.001)
Review Year	-0.575***	-0.349***	0.039	0.001
	(0.084)	(0.068)	(0.037)	(0.030)
Lag Mean Rating Score		0.719***		
		(0.027)		
Lag S.D. of Rating Score		-0.549***		
		(0.038)		
Lag Mean Extreme Low Ratings				1.047***
				(0.043)
Lag S.D. Extreme Low Ratings				-0.017
				(0.012)
Movie Fixed Effects	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓
R^2	0.457	0.461	0.262	0.266
Number of Observations	15825513	15815566	15825513	15815566

† A movie is considered female if the top actress is female. Specifications (1)-(4) use the rating score (measured from 10 to 100) as the dependent variable. Specifications (5)-(8) use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. Box office revenue is measured in million USD. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A18: Differences in Gender Score Gaps between Critics and Rotten Tomatoes Crowds. Early vs Late Periods. †

Dependent Variable:	Rating Score				Extreme Low Ratings			
Early/Late Period Cut-off (in days):	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Female Movie × Crowd Reviewer × Early	-1.244***	-1.340***	-1.403***	-1.503***	0.765***	0.828***	0.860***	0.876***
	(0.388)	(0.389)	(0.384)	(0.339)	(0.168)	(0.167)	(0.163)	(0.141)
Female Movie × Crowd Reviewer × Late	-1.760***	-1.656***	-1.653***	-1.760***	0.803***	0.745***	0.709***	0.638***
	(0.276)	(0.270)	(0.267)	(0.294)	(0.120)	(0.119)	(0.118)	(0.148)
Review Position within Sequence × Early	-0.050***	-0.058***	-0.054***	-0.040***	0.006***	0.006***	0.006***	0.006***
	(0.013)	(0.011)	(0.009)	(0.008)	(0.002)	(0.001)	(0.001)	(0.001)
Review Position within Sequence × Late	-0.031***	-0.032**	-0.036*	-0.043***	0.010***	0.012***	0.012***	-0.008*
	(0.010)	(0.013)	(0.020)	(0.008)	(0.002)	(0.003)	(0.003)	(0.004)
Elapsed Time since First Review × Early	-5.899***	-6.504***	-7.429***	-12.510***	1.048***	1.262***	1.379***	3.091***
	(0.717)	(0.701)	(0.749)	(0.870)	(0.286)	(0.279)	(0.263)	(0.318)
Elapsed Time since First Review × Late	0.402*	0.195	0.173	-0.149	0.056	0.002	-0.096	-0.211*
	(0.240)	(0.242)	(0.230)	(0.171)	(0.108)	(0.107)	(0.112)	(0.125)
Review Year × Early	0.093	0.000	-0.228	-0.581***	-0.088	-0.108	-0.062	-0.015
	(0.182)	(0.163)	(0.146)	(0.100)	(0.079)	(0.071)	(0.054)	(0.043)
Review Year × Late	-0.366***	-0.281***	-0.262***	-0.154***	0.049	0.050	0.059	0.043
	(0.085)	(0.085)	(0.076)	(0.060)	(0.037)	(0.037)	(0.039)	(0.044)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
t-test: Diff. between Early and Late periods	2.56	0.94	0.58	0.64	0.06	0.28	0.96	2.07
p-value	0.110	0.332	0.444	0.423	0.806	0.594	0.328	0.150
R^2	0.473	0.477	0.480	0.479	0.276	0.280	0.285	0.287
Number of Observations	15051250	14932159	14940994	15178547	15051250	14932159	14940994	15178547

† A movie is considered female if the top actress is female. Specifications (1)-(4) use the rating score (measured from 10 to 100) as the dependent variable. Specifications (5)-(8) use a dummy equal to 1 (and multiplied by 100) for ratings equal to 10 as the dependent variable. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Box office revenue is measured in million USD. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A19: Differences in Gender Sentiment Gaps between Critics and Rotten Tomatoes Crowds. Early vs Late Periods. VADER Sentiment Analysis. †

Dependent Variable:	Sentiment Score				Extreme Sentiment Score			
Early/Late Period Cut-off (in days):	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Female Movie × Crowd Reviewer × Early	-0.024***	-0.026***	-0.025***	-0.026***	0.227***	0.219***	0.213***	0.237***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.077)	(0.072)	(0.068)	(0.060)
Female Movie × Crowd Reviewer × Late	-0.024***	-0.022***	-0.022***	-0.026***	0.221***	0.210***	0.225***	0.138
	(0.004)	(0.004)	(0.004)	(0.005)	(0.066)	(0.069)	(0.074)	(0.098)
Review Position within Sequence × Early	-0.000**	-0.001***	-0.001***	-0.001***	0.002	0.001	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Review Position within Sequence × Late	-0.001***	-0.001***	-0.000***	0.000	0.001	0.001	0.002	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)
Elapsed Time since First Review × Early	-0.031***	-0.033***	-0.044***	-0.163***	0.433**	0.418***	0.362***	0.301***
	(0.010)	(0.010)	(0.010)	(0.013)	(0.173)	(0.152)	(0.122)	(0.082)
Elapsed Time since First Review × Late	-0.012***	-0.015***	-0.014***	-0.011***	0.161***	0.157***	0.116**	0.180**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.051)	(0.054)	(0.058)	(0.088)
Review Year × Early	-0.001	-0.003	-0.006***	-0.004***	-0.055	-0.031	-0.016	0.007
	(0.003)	(0.002)	(0.002)	(0.002)	(0.050)	(0.040)	(0.029)	(0.017)
Review Year × Late	-0.002	-0.000	-0.000	-0.000	0.004	0.007	0.022	0.019
	(0.001)	(0.001)	(0.001)	(0.001)	(0.018)	(0.019)	(0.020)	(0.031)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
t-test: Diff. between Early and Late periods	0.00	0.61	0.48	0.00	0.01	0.01	0.02	1.10
p-value	0.991	0.435	0.490	0.974	0.930	0.907	0.876	0.295
R^2	0.279	0.282	0.284	0.283	0.152	0.154	0.156	0.156
Number of Observations	13974232	13864631	13875259	14099914	13974232	13864631	13875259	14099914

† A movie is considered female if the top actress is female. All specifications rely on the VADER classifier to construct reviews' sentiment score. The score ranges from -1 to 1, with a higher score reflecting a more positive sentiment. Columns (1)-(4) use the review sentiment score as the dependent variable. Columns (5)-(8) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 1^{st} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A20: Differences in Gender Sentiment Gaps between Critics and Rotten Tomatoes Crowds. Early vs Late Periods. TextBlob Sentiment Analysis. †

Dependent Variable:	Sentiment Score				Extreme Sentiment Score			
Early/Late Period Cut-off (in days):	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd Reviewer × Early	-0.016***	-0.016***	-0.016***	-0.016***	0.173***	0.135***	0.141***	0.131***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.051)	(0.048)	(0.046)	(0.040)
Female Movie × Crowd Reviewer × Late	-0.013***	-0.012***	-0.012***	-0.014***	0.094**	0.100**	0.071	0.034
	(0.002)	(0.002)	(0.002)	(0.003)	(0.046)	(0.048)	(0.051)	(0.070)
Review Position within Sequence × Early	-0.000***	-0.000***	-0.000***	-0.000***	0.000	0.002*	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Review Position within Sequence × Late	-0.000***	-0.000***	-0.000***	-0.001***	0.004***	0.005***	0.006***	0.004**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
Elapsed Time since First Review × Early	-0.024***	-0.024***	-0.028***	-0.062***	0.204**	0.238***	0.323***	0.965***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.097)	(0.087)	(0.081)	(0.087)
Elapsed Time since First Review × Late	-0.004**	-0.005**	-0.004**	-0.004*	0.161***	0.181***	0.169***	0.015
	(0.002)	(0.002)	(0.002)	(0.002)	(0.044)	(0.045)	(0.049)	(0.066)
Review Year × Early	0.001	-0.000	-0.001	-0.003***	-0.021	-0.015	-0.018	-0.022
	(0.001)	(0.001)	(0.001)	(0.001)	(0.031)	(0.026)	(0.022)	(0.016)
Review Year × Late	-0.002***	-0.002**	-0.002***	-0.002**	-0.023	-0.036**	-0.037**	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.015)	(0.016)	(0.017)	(0.024)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
t-test: Diff. between Early and Late periods	1.08	2.53	2.16	0.38	1.48	0.30	1.17	1.59
p-value	0.300	0.112	0.141	0.536	0.224	0.586	0.279	0.208
R^2	0.297	0.299	0.303	0.303	0.181	0.184	0.188	0.192
Number of Observations	13974232	13864631	13875259	14099914	13974232	13864631	13875259	14099914

† A movie is considered female if the top actress is female. All specifications rely on the TextBlob classifier to construct reviews' sentiment score. The score ranges from -1 to 1, with a higher score reflecting a more positive sentiment. Columns (1)-(4) use the review sentiment score as the dependent variable. Columns (5)-(8) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 1^{st} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A21: Differences in Gender Sentiment Gaps between Critics and Rotten Tomatoes Crowds. Early vs Late Periods. LIWC Sentiment Analysis. †

Dependent Variable:	Sentiment Score				Extreme Sentiment Score			
Early/Late Period Cut-off (in days):	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$	$\tau = 7$	$\tau = 14$	$\tau = 30$	$\tau = 180$
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie × Crowd Reviewer × Early	0.098**	0.107***	0.093**	0.119***	0.156***	0.157***	0.144***	0.149***
	(0.042)	(0.041)	(0.039)	(0.034)	(0.051)	(0.049)	(0.046)	(0.040)
Female Movie × Crowd Reviewer × Late	0.134***	0.125***	0.141***	0.128***	0.134***	0.126**	0.134**	0.103
	(0.033)	(0.034)	(0.036)	(0.047)	(0.048)	(0.051)	(0.055)	(0.075)
Review Position within Sequence × Early	0.002***	0.003***	0.003***	0.002***	0.001	0.003***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Review Position within Sequence × Late	0.002**	0.002**	0.003	0.004**	0.002**	0.003	0.002	0.007***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.004)	(0.002)
Elapsed Time since First Review × Early	0.375***	0.387***	0.463***	1.038***	0.331***	0.314***	0.476***	1.213***
	(0.069)	(0.063)	(0.061)	(0.069)	(0.116)	(0.105)	(0.095)	(0.099)
Elapsed Time since First Review × Late	0.111***	0.136***	0.130***	0.061	0.241***	0.270***	0.278***	0.144*
	(0.028)	(0.030)	(0.033)	(0.039)	(0.050)	(0.054)	(0.062)	(0.077)
Review Year × Early	-0.025	-0.013	-0.004	0.001	-0.022	-0.010	-0.033	-0.020
	(0.021)	(0.018)	(0.016)	(0.012)	(0.035)	(0.030)	(0.024)	(0.018)
Review Year × Late	-0.004	-0.016	-0.017	-0.006	-0.025	-0.039**	-0.045**	-0.027
	(0.010)	(0.010)	(0.011)	(0.014)	(0.018)	(0.019)	(0.021)	(0.027)
Movie Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Reviewer Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
t-test: Diff. between Early and Late periods	0.65	0.14	1.10	0.03	0.10	0.21	0.02	0.30
p-value	0.421	0.705	0.295	0.852	0.750	0.643	0.885	0.584
R^2	0.229	0.232	0.235	0.237	0.182	0.184	0.188	0.191
Number of Observations	13972288	13862698	13873308	14097982	13972288	13862698	13873308	14097982

† A movie is considered female if the top actress is female. All specifications rely on the LIWC's negative tone dictionary to construct reviews' sentiment score, ranging from 0 to 100. A higher score reflects a more negative review sentiment. Columns (1)-(4) use the review sentiment score as the dependent variable. Columns (5)-(8) use a dummy equal to 1 (and multiplied by 100) if the review's sentiment score is extreme as the dependent variable. Extreme scores are defined as scores that fall in the 99^{th} percentile of the review sentiment score distribution. The variables corresponding to the Elapsed Time since First Review and the Review Position within Sequence are divided by 1,000 to facilitate the readability of the corresponding coefficients. Crowd ratings come from Rotten Tomatoes. Standard errors are clustered at the movie level and reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

reviewers and critics for movie m. The coefficient δ therefore directly corresponds to the difference-in-difference estimate of interest.

Column (1) of Table A22 report the results of estimating equation (1.a) using the Rotten Tomatoes crowd data. Consistent with the results presented above, the difference-in-differences estimates is equal to -1.9, almost identical to our preferred estimate in specification (4) in Table 4. Using the IMDb crowds data – in specification (3) – leads to a similar estimate. Specification (2) and (4) perform similar exercises using the difference in the share of extreme low ratings received by crowd reviewers and critics by each movie. Perhaps unsurprisingly, the estimates are again consistent with the results presented above and show that the gender gap in extreme low ratings is significantly larger among crowds than critics, and even more so on Rotten Tomatoes than IMDb.[11] Table A23 reports the results of estimating similar specifications using sentiment scores as dependent variables. All show similar results.

[11]Tables A24 and A25 present the corresponding estimation results when relying on less restrictive definitions of female movies using the Rotten Tomatoes crowd data and IMDb crowd data, respectively. The results are once again robust to such different definitions.

Table A22: Gender Gaps between Critics and Crowd using Cross-Section of Movies. †

Dependent Variable Based On:	Rotten Tomatoes Crowds		IMDb Crowds	
	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.
Female Movie	-1.890***	0.853***	-1.550***	0.419**
	(0.295)	(0.258)	(0.225)	(0.195)
Worldwide Box Office Revenue	0.003***	-0.001**	-0.000	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Movie Released in Theaters	0.800*	-0.979**	0.883***	-1.443***
	(0.415)	(0.397)	(0.318)	(0.297)
Number of Release Countries	-0.010	-0.022***	0.028***	-0.058***
	(0.010)	(0.007)	(0.008)	(0.005)
Number of Awards Won	-0.133***	-0.002	-0.138***	-0.001
	(0.009)	(0.005)	(0.009)	(0.003)
Genre Fixed Effects	✓	✓	✓	✓
Production Studio Fixed Effects	✓	✓	✓	✓
MPAA Rating Fixed Effects	✓	✓	✓	✓
Year of Release Fixed Effects	✓	✓	✓	✓
Origin Fixed Effects	✓	✓	✓	✓
Language Fixed Effects	✓	✓	✓	✓
R^2	0.110	0.086	0.158	0.084
Number of Observations	12657	12657	12657	12657

† A movie is considered female if the top actress is female. Each observation in the sample corresponds to a movie. In specifications (1) and (3), the dependent variable is the difference in the average rating score between crowd and critics. In specifications (2) and (4), the dependent variable is the difference in the share of extreme low ratings between crowd and critics. Box office revenue is measured in million USD. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A23: Gender Sentiment Score Gaps between Critics and Crowd using Cross-Section of Movies. †

	Sentiment Score			Extreme Sentiment Score		
Sentiment Classifier:	VADER	TextBlob	LIWC	VADER	TextBlob	LIWC
	(1)	(2)	(3)	(4)	(5)	(6)
	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.	Coef./s.e.
Female Movie	-0.035***	0.182	-0.016***	0.172**	0.224***	0.271**
	(0.006)	(0.116)	(0.003)	(0.081)	(0.063)	(0.132)
Worldwide Box Office Revenue	-0.000	-0.000	0.000***	-0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Movie Released in Theaters	0.014	-0.392**	0.012**	-0.010	0.054	0.109
	(0.009)	(0.181)	(0.005)	(0.137)	(0.101)	(0.218)
Number of Release Countries	0.001***	-0.014***	0.000***	0.002	0.002	0.007*
	(0.000)	(0.002)	(0.000)	(0.002)	(0.002)	(0.004)
Number of Awards Won	-0.000	-0.003	-0.000***	-0.003**	0.001	0.004**
	(0.000)	(0.002)	(0.000)	(0.001)	(0.001)	(0.002)
Genre Fixed Effects	✓	✓	✓	✓	✓	✓
Production Studio Fixed Effects	✓	✓	✓	✓	✓	✓
MPAA Rating Fixed Effects	✓	✓	✓	✓	✓	✓
Year of Release Fixed Effects	✓	✓	✓	✓	✓	✓
Origin Fixed Effects	✓	✓	✓	✓	✓	✓
Language Fixed Effects	✓	✓	✓	✓	✓	✓
R^2	0.041	0.057	0.067	0.023	0.021	0.019
Number of Observations	12566	12657	12566	12657	12566	12657

† A movie is considered female if the top actress is female. Each observation in the sample corresponds to a movie. In specifications (1) and (3), the dependent variable is the difference in the average sentiment score between crowd and critics. In specifications (2) and (4), the dependent variable is the difference in the share of extreme sentiment score between crowd and critics. Box office revenue is measured in million USD. Crowd review scores come from Rotten Tomatoes. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A24: Gender Gaps between Critics and Rotten Tomatoes Crowd using Cross-Section of Movies. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
Dependent Variable:	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie	-1.890***	0.853***	-1.386***	0.820***	-1.202***	0.679**	0.336	0.188
	(0.295)	(0.258)	(0.292)	(0.259)	(0.351)	(0.320)	(0.392)	(0.230)
Worldwide Box Office Revenue	0.003***	-0.001**	0.003***	-0.001**	0.003***	-0.001**	0.002**	-0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Movie Released in Theaters	0.800*	-0.979**	0.831**	-0.992**	0.834**	-0.994**	-0.207	-0.042
	(0.415)	(0.397)	(0.416)	(0.397)	(0.416)	(0.397)	(0.969)	(0.701)
Number of Release Countries	-0.010	-0.022***	-0.011	-0.022***	-0.011	-0.022***	0.023*	-0.010
	(0.010)	(0.007)	(0.010)	(0.007)	(0.010)	(0.007)	(0.013)	(0.008)
Number of Awards Won	-0.133***	-0.002	-0.133***	-0.002	-0.133***	-0.002	-0.120***	0.005
	(0.009)	(0.005)	(0.009)	(0.005)	(0.009)	(0.005)	(0.008)	(0.004)
Genre Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Production Studio Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
MPAA Rating Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year of Release Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Origin Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Language Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.110	0.086	0.108	0.086	0.108	0.085	0.224	0.193
Number of Observations	12657	12657	12657	12657	12657	12657	3571	3571

† Each observation in the sample corresponds to a movie. For each Female movie definition, the first specification uses the difference in the average rating score between crowd and critics as the dependent variable, while the second uses the difference in the share of extreme low ratings between crowd and critics as the dependent variable. Crowd ratings come from Rotten Tomatoes. Box office revenue is measured in million USD. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table A25: Gender Gaps between Critics and IMDb Crowd using Cross-Section of Movies. †

Female Movie Definition Based On:	Female Lead		Female in Top 2		Female in Top 3		Bechdel Test	
Dependent Variable:	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating	Rating Score	Extreme Low Rating
	(1) Coef./s.e.	(2) Coef./s.e.	(3) Coef./s.e.	(4) Coef./s.e.	(5) Coef./s.e.	(6) Coef./s.e.	(7) Coef./s.e.	(8) Coef./s.e.
Female Movie	-1.550***	0.419**	-1.037***	0.405*	-0.762***	0.233	-0.194	0.465***
	(0.225)	(0.195)	(0.228)	(0.208)	(0.281)	(0.267)	(0.322)	(0.175)
Worldwide Box Office Revenue	-0.000	0.002***	-0.000	0.002***	-0.000	0.002***	-0.000	0.001***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
Movie Released in Theaters	0.883***	-1.443***	0.908***	-1.450***	0.910***	-1.450***	0.550	-0.580
	(0.318)	(0.297)	(0.319)	(0.297)	(0.319)	(0.297)	(0.784)	(0.573)
Number of Release Countries	0.028***	-0.058***	0.027***	-0.058***	0.026***	-0.058***	0.042***	-0.034***
	(0.008)	(0.005)	(0.008)	(0.005)	(0.008)	(0.005)	(0.010)	(0.005)
Number of Awards Won	-0.138***	-0.001	-0.139***	-0.001	-0.139***	-0.001	-0.115***	0.001
	(0.009)	(0.003)	(0.009)	(0.003)	(0.009)	(0.003)	(0.008)	(0.002)
Genre Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Production Studio Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
MPAA Rating Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Year of Release Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Origin Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Language Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.158	0.084	0.157	0.084	0.156	0.084	0.270	0.125
Number of Observations	12657	12657	12657	12657	12657	12657	3571	3571

† Each observation in the sample corresponds to a movie. For each Female movie definition, the first specification uses the difference in the average rating score between crowd and critics as the dependent variable, while the second uses the difference in the share of extreme low ratings between crowd and critics as the dependent variable. Crowd ratings come from Rotten Tomatoes. Box office revenue is measured in million USD. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

References

Bayerl, A., Y. Dover, H. Riemer, and D. Shapira (2024): "Gender rating gap in online reviews," *Nature Human Behavior*.

Boyd, R. L., A. Ashokkumar, S. Seraj, and J. W. Pennebaker (2022): "The development and psychometric properties of LIWC-22," *Austin, TX: University of Texas at Austin*, 1–47.

Boyle, K. (2014): "Gender, comedy and reviewing culture on the Internet Movie Database," *Participations*, 11, 31–49.

Charmes, J. (2019): "The Unpaid Care Work and the Labour Market. An analysis of time use data based on the latest World Compilation of Time-use Surveys," *International Labour Office*.

Godes, D. and D. Mayzlin (2004): "Using online conversations to study word-of-mouth communication," *Marketing science*, 23, 545–560.

Godes, D. and J. C. Silva (2012): "Sequential and temporal dynamics of online opinion," *Marketing Science*, 31, 448–473.

Hutto, C. and E. Gilbert (2014): "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, vol. 8, 216–225.

Moe, W. W. and M. Trusov (2011): "The value of social dynamics in online product ratings forums," *Journal of marketing research*, 48, 444–456.

Pennebaker, J. W., M. E. Francis, and R. J. Booth (2001): "Linguistic inquiry and word count: LIWC 2001," *Mahway: Lawrence Erlbaum Associates*, 71, 2001.

Stroube, B. K. and D. M. Waguespack (2024): "Status and consensus: Heterogeneity in audience evaluations of female-versus male-lead films," *Strategic Management Journal*.